

Production-Comprehension (A)symmetry: Individual Differences in the Acquisition of Prosodic Focus-marking

Aoju Chen

Utrecht University & Max Planck Institute for Psycholinguistics
 aoju.chen@uu.nl

Abstract

Previous work based on different groups of children has shown that four- to five-year-old children are similar to adults in both producing and comprehending the focus-to-accentuation mapping in Dutch, contra the alleged production-precedes-comprehension asymmetry in earlier studies. In the current study, we addressed the question of whether there are individual differences in the production-comprehension (a)symmetry. To this end, we examined the use of prosody in focus marking in production and the processing of focus-related prosody in online language comprehension in the same group of 4- to 5-year-olds. We have found that the relationship between comprehension and production can be rather diverse at an individual level. This result suggests some degree of independence in learning to use prosody to mark focus in production and learning to process focus-related prosodic information in online language comprehension, and implies influences of other linguistic and non-linguistic factors on the production-comprehension (a)symmetry.

Index Terms: prosody, focus, comprehension-production asymmetry, individual differences, first language acquisition

1. Introduction

A common pattern in language acquisition is that comprehension precedes production [1, 2]. There are however aspects of language which are characterised by production-precedes-comprehension asymmetries [2]. The acquisition of sentence-level prosody (or: intonation) is a case in point. A most widely discussed phenomenon is the mapping between accentuation and focus. Focus refers to the predication on a topic and typically contains new information to the hearer [3, 4]. In many languages, there is a strong association between focus and accentuation. Specifically, speakers tend to accent the focused constituent and deaccent the unfocused constituents, especially if they are post-focal [5]. Listeners take this into account in online language comprehension such that appropriate focus-to-accentuation mapping speeds up comprehension and inappropriate focus-to-accentuation mapping slows down comprehension [6]. Further, listeners use this mapping to anticipate the upcoming referent in online reference resolution [7]. Adult-like competence in prosodic focus-marking thus entails that children not only can place accentuation to encode focus but also efficiently exploit the focus-to-accentuation mapping in language comprehension. The literature on the acquisition of the focus-to-accentuation mapping over the past two decades has been dominated by the claim that children can use accentuation to mark focus before they can interpret or efficiently use the focus-to-accentuation in comprehension [8].

Recently [9] has pointed out that the alleged asymmetry in the acquisition of the focus-to-accentuation mapping could arise from asymmetries in the test materials used in the production and comprehension studies. More specifically, the

test materials in the comprehension studies were usually syntactically more complex and semantically more demanding than the materials used in the production studies. Consequently, what children were supposed to comprehend went beyond the simple focus-to-accentuation mapping in SVO or SV sentences, e.g. the use of accentuation to disambiguate pronouns, accentuation as a cue to contrastive focus in sentences with the focus particle ‘only’. [9] examined 4- to 5-year-olds’ and adults’ production and online comprehension of the focus-to-accentuation mapping in Dutch SVO sentences and found that the 4- to 5-year-olds were similar to the adults in both production and comprehension and the differences between them were of a gradient nature (i.e. more frequent use of accentuation in post-focus constituents and slower reaction times in the children). This is the first evidence suggesting that there is no asymmetry in the acquisition of prosodic focus-marking when the syntactic complexity is controlled for in production and comprehension.

However, prosodic focus-marking is more than the use of accentuation. Accent type also plays a role in focus-marking [10]. For example, in Dutch the preferred accent type to mark focus is H*L in both contrastive and non-contrastive focus [11]. When the same accent type occurs, speakers vary the phonetic realisation of the accent for the purpose of focus marking [12, 13]. Four- to five-year-old Dutch-speaking children are adult-like in choice of accent type in sentence-initial position but not in sentence-final position in SVO sentences [11]. Further, they cannot use phonetic realisation for focus-marking purposes until the age of 8 [12]. Together with the differences in the frequency of accentuation and deaccentuation between children and adults, these differences may have perceptual consequences. It can thus be very useful to examine children’s use of prosody in focus-marking through adults’ evaluation of children’s production.

More importantly, the children in [9]’s production experiment were not the same children as the ones in her comprehension experiment, although they were similar in age and recruited from the same schools. Individual differences in children’s intonational skills have been reported for both production and comprehension tasks in earlier work [9, 14]. Can we reduplicate [9]’s results if we examine both production and comprehension in the same group of children? Further, if at the group level, children’s production is similar to their comprehension, relative to adults’ production and comprehension, does it then mean that production and comprehension go in tandem in every child? In other words, are there individual differences in the production-comprehension symmetry?

To address the aforementioned questions, we investigated the use of prosody to mark focus in production and the processing of focus-related prosodic information in a single group of Dutch 4- to 5-year-olds. Production data were obtained from the children in a semi-spontaneous setting. Sentence produced by the children were subsequently

evaluated for the appropriateness of the prosody in the corresponding context by trained raters. Comprehension was examined using the same method as in [9]. The focus conditions at issue were narrow focus on the subject NP and narrow focus on the object NP.

2. The production study

The production study consisted of a production experiment and an evaluation experiment. In the production experiment, SVO sentences were elicited in different focus conditions. In the evaluation experiment, trained raters evaluated the appropriateness of the prosody in the participants' sentences.

2.1. The production experiment

2.1.1. Data elicitation

The picture-matching game used in [9] was adapted for the current purpose. The game was played in experimenter-participant dyads. The participant's task was to help the experimenter to find the matching picture for each of her pictures by answering her questions. The conversation between the experimenter and the participant was primarily composed of short question-answer dialogues. The participant had direct access to the information that the experimenter needed and could respond directly to the experimenter's queries. This was achieved by providing the participant with his own set of pictures, each of which depicted a complete event including an agent, a patient and an action. Prior to the game, each participant completed a picture-naming task, in which he named each animal, personage, object and action present in the game and got corrected if he misnamed an entity or used a non-target form. The entities in the pictures were thus referentially given to the participant at the start of the game, rendering the use of a definite article in reference in the game appropriate.

In the game, the experimenter showed the participant one picture at a time, drew the participant's attention to the picture, briefly described it (e.g. Look! The girl. There is also the pan. It seems that the girl cooks something.), and then asked the participant a question about the picture (e.g. What does the girl cook?). The participant took a picture from his own set of pictures, which were pre-arranged in an order corresponding to the order of the experimenter's pictures, and tried to identify the information requested by the experimenter. When the participant looked away from his picture, the experimenter repeated her question. The participant then answered the question in an SVO sentence (e.g. The girl cooks the carrot.).

Fifteen question-answer dialogues were embedded in the game to elicit fifteen SVO sentences in five focus conditions: narrow focus in sentence-initial position (NF-i), responding to who-questions and narrow focus in sentence-final position (NF-f), responding to what-questions, in addition to narrow focus in sentence-medial position (NF-m), responding to what-does-X-do-with-Y questions, contrastive focus in sentence-medial position (CF-m), correcting the experimenter's statement about the action, broad focus over the whole sentence (BF), responding to what-happens questions. The target SVO sentences were unique combinations of 3 verbs, 3 object-nouns and 6 subject-nouns. All words were highly familiar words to Dutch 4-year-olds. Each verb and object noun occurred once in each focus condition but never appeared with the same subject noun twice in the game.

2.1.2. Participants

Seventy-five 4- and 5-year-olds (age range: 4;1 to 5;11) participated in the production study. The children were all from monolingual Dutch-speaking families and were recruited from four primary schools in Utrecht Province. Nine adult female native speakers of Dutch took part in the experiment as controls. All participants had normal hearing and speaking ability. Three children did not finish the game. Each

session was audio-recorded at a sampling rate of 44.1 kHz with 16 bits resolution and video-recorded.

2.1.3. Data annotation

For each participant, the recording was first orthographically annotated in Praat [15]. Second, full-sentence responses were selected as usable responses if they were not plagued by any of the following factors: self-correction, use of pronouns, use of non-target words, detectable hesitation-induced silences, responding to a non-target question, elided responses, overlap with the experimenter's speech, and poor recording quality. Third, the usable full-sentence responses and the corresponding questions or statements were selected and extracted as individual .wav files.

2.2. The evaluation experiment

The usable full-sentence responses and corresponding questions or statements were combined into context-response dialogues with a 1000-ms interval between the question and the response in each dialogue. Subsequently, three intonationally-trained native speakers of Dutch listened to the dialogues and evaluated each response on how well its prosody fitted in the context on a five-point Equal Appearing Interval scale, with 1 standing for 'does not fit' and 5 standing for 'fits perfectly'. In total, 105 dialogues from 42 children and 25 dialogues from 9 adults in the NF-i condition and 92 dialogues from 36 children and 25 dialogues from 9 adults in the NF-f condition were subjected to evaluation, together with the dialogues from the other focus conditions. The production of 32 of the children was evaluated in both the NF-i and NF-f conditions. To minimise variation in the scores due to comparisons between speakers, the dialogues were presented to the raters per speaker and the experiment was conducted in four 20- to 30-minute sessions. The raters could listen to each dialogue maximally three times before finalising the score. Inter-rater reliability analysis showed that there was a high inter-rater agreement (Cronbach's Alpha = .793; Interclass correlation coefficient = .774)

2.3. Results

Mixed modelling was used to examine the difference in the scores between the children and the adults in each focus condition. The fixed factor was 'age-group' (children vs. adults). The random factors included 'sentence' and 'participant'. The dependent variable was the mean score of the raters. In each analysis on the effect of 'age-group', two models were built, one with only the random factors, and one with both the random factors and the fixed factor. The two models were then compared to each other in an ANOVA test. A statistically significant difference between these two models indicated a main effect of the fixed factor. The p-values reported here were the p-values of the ANOVA tests. Our models showed that the fixed factor 'age-group' had a main effect on the scores in the NF-i condition but not in the NF-f condition. The difference between the children (mean: 2.56) and the adults (mean: 3.65) was thus statistically significant in the NF-f condition ($p < .001$), indicating that the children were not adult-like in their use of prosody in sentences with a focal subject. The differences between the children (mean score: 3.45) and the adults (mean score: 3.22) in the NF-i condition was not statistically significant, indicating that the children could use prosody as appropriately as the adults did in sentences with a focal object.

Taking results from [11, 12] into account, these results suggest that the adult listeners found accenting the post-focus object NPs, acceptable, as found by [16, 17]. But they found the children's failure to exploit phonetic realisation to

distinguish focus from non-focus in the NF-i condition less acceptable.

3. The comprehension study

The reaction-time experiment used in [9] was conducted on the four- to five-year-olds who took part in the production study about two weeks after the production experiment. We only describe the most essential details of the RT experiment here and refer the reader to [9] for further information.

3.1. The RT experiment

The experiment was presented to the children as a game. The children were told that in the game a boy was going to look at a number of pictures with his three pets, a parrot, a chicken, and a duck. The boy wanted to know whether his pets knew the pictures well and which of the pets knew the pictures best. To find this out, the boy showed one picture at a time to one of his pets and asked the pet a question about the picture. The children could hear the conversations between the boy and the pets via a headphone set and see each picture on a computer screen together with the boy and his pets. The children were asked to listen to the dialogues between the boy and his pets carefully and judge whether the pets have given correct answers to the boy's questions or not. If they thought a pet gave a correct answer, they should press the green button of the pushbutton box. If they thought that the pet's answer was incorrect for some reason, they should press the red button.

Four lists of 24 experimental dialogues were created from 24 source answer-sentences together with the accompanying pictures. The answer sentences were lexically identical but appeared in different focus conditions (NF-i and NF-f) and prosody conditions in different lists. The prosody of the answer sentences was appropriate in half of the dialogues and inappropriate in the other half of the dialogues. In addition, 20 fillers were included. The answers in the experimental dialogues were all semantically correct but the answers in the fillers contained either a lexical error or a pronunciation error. The questions of the dialogues were recorded by a male speaker of Dutch and the answers by a female speaker of Dutch at a sampling rate of 44.1 kHz with 16 bits resolution in the recording studio of the Max Planck Institute for Psycholinguistics. Prosodic analysis on the answer sentences confirmed that the prosody was as intended in both the appropriate and inappropriate prosody conditions. The answer sentences were similar in length across focus conditions and prosody conditions.

The children did the experiment individually in a quite room at their schools. The experiment was conducted by means of the Zep Experimental Control Application [18]. An approximately equal number of children were assigned to each of the list. The exact stimulus list a child got was however randomly decided. Each session lasted about 20 minutes starting with a practice session. In the practice session, the children were familiarised with the task and trained to use the push-button box properly.

The timeline of a trial was as follows: A target picture together with the picture of the boy and one of his pets appeared on the screen. At the same time, the boy said *Kijk* 'look' as an attention getter. 800 ms later, he named an entity in the picture. The 800-ms delay was built in to allow the participants to take a proper look at the picture. 1200 ms after the naming, the boy asked the question. 2200 ms after the end of the question, the pet answered the question. At the end of the answer sentence, a timer with 1 ms accuracy was activated and a picture of the push-button box appeared on the screen.

The RT was recorded from the end of the answer sentence until a button was pressed and the correct–incorrect judgment was automatically recorded. The children were instructed to press the

button as quickly as they could, but not before the end of the answer sentence. A timeout device was set at 4 s after the end of a sentence.

Seventy-one out of the 75 children completed the experiment. Two measures were taken from each child, i.e. the 'correct–incorrect' judgement and the RT.

3.2. Statistical analysis and results

The children judged the answers on the experimental trials to be correct in nearly all cases, as expected. The reaction times obtained from the children whose production was evaluated in the NF-i condition and the reaction times obtained from the children whose production was evaluated in the NF-f condition were included for further analysis. The reaction times were log-transformed. Mixed-effect modelling was used to assess the effect of two fixed factors, 'prosody' (appropriate vs. inappropriate) and 'focus condition' (NF-i vs. NF-f) and their interaction on the log-transformed reaction times. Two random factors were included, 'participant' and 'list'. The models were built as described in section 2.2. Our models revealed a main effect the fixed factor 'focus' ($p < .0001$) and a significant interaction between 'focus' and intonation' ($p < .0001$). The main effect of 'focus condition' was such that object focus triggered a longer mean RT than subject focus. This was related to the fact that one had to wait till the end of the sentence in the NF-f condition to form his judgement [9]. The effect of the interaction between 'prosody' and 'focus condition' was such that inappropriate prosody triggered a longer mean RT than appropriate prosody in the NF-f condition but a shorter mean RT in the NF-i condition. This showed that the children were similar to adults in their processing of the focus-prosody interface only when focus was sentence-final.

4. Production and Comprehension

The discussion above centred on the general patterns that have emerged in the children as a group in production on the one hand and in comprehension on the other hand. The results revealed a production-comprehension symmetry whereby the children's production went in tandem with their comprehension. In this section, we address the question of whether there are individual differences in the production-comprehension symmetry. We focus on the data from the children who were both evaluated for their production and completed the reaction experiment (40 in the NF-i condition and 36 in the NF-f condition; 32 of the children appeared in both conditions; mean age: 5;3).

To quantify individual differences in the production-comprehension interface, two kinds of scores were obtained for each child: 'production scores' that could reflect a child's ability to use prosody in different focus conditions in production; 'comprehension scores' that could reflect how a child processed inappropriate prosody compared to appropriate prosody in different focus conditions. Regarding the production scores, a score was computed for each child in each focus condition by first calculating the mean score of the raters for each response and then averaging the mean scores of all available responses in each focus condition. Regarding the comprehension scores, a score was computed for each child by calculating the ratio between the mean log-transformed reaction time in the inappropriate-prosody condition and that in the appropriate-prosody condition in each focus condition. If the child responded faster in the appropriate-prosody condition than in the inappropriate-prosody condition, the child should have a comprehension score bigger than 1.

We examined the correlation between the production scores and comprehension scores by conducting a Spearman's

correlation coefficient test in each focus condition. The Spearman's correlation coefficient was .113 in the NF-i condition and .034 in the NF-f condition. The significance of the correlation coefficient was .481 and .845 respectively. The results suggested that there was no significant relationship between production and comprehension.

Interestingly, examining the production and comprehension scores in each child, we observed notable individual differences among the children. Assuming that a comprehension score higher than 1 meant successful processing of the focus-prosody interface and a production score higher than 3 indicated relatively accurate use of prosody in focus marking, there appeared to be four sub-groups in the children regardless of the focus condition: (1) poor production, poor comprehension; (2) good production, good comprehension; (3) poor production, good comprehension; (4) good production, poor comprehension, as shown in Figure 1. Such notable individual differences account for the insignificant correlation coefficients in the correlation tests.

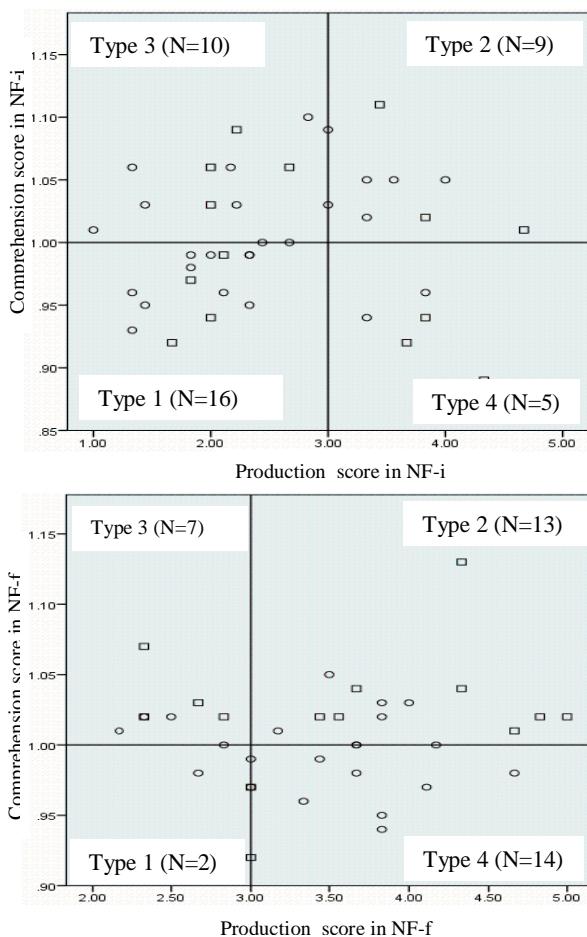


Figure 1. Children's production and comprehension scores (40 in the NF-i condition, 36 in the NF-f condition). Circle: female; square: male

5. Discussion and conclusions

In the current study, we have taken into account the use of multiple prosodic cues in focus marking and examined production and comprehension in a single group of Dutch 4- to 5-year-olds. The children were perceived to be able to mark

focus prosodically like adults when focus was in sentence-final position, but not when focus was in sentence-initial position. This asymmetry in the children's production ability was in line with the results based on phonological and phonetic analysis on children's prosody [11, 12] and the tendency to accent given information in adult Dutch [15,16]. By evaluating the overall use of prosody in focus marking, we have obtained a more accurate picture of children's prosodic focus-marking in production. Our results however differed from the results reported in [9] regarding comprehension. The 4- to 5-year-olds in this study differed from their age-matched peers and adults in [9], who responded faster in the appropriate-prosody condition regardless of focus conditions. At first sight, these results may look perplexing. However, these results can be well explained by the children's own production. The children were adult-like in how they processed the focus-prosody interface in the NF-f condition but not in the NF-i condition. Interestingly, their use of prosody was judged to be as appropriate as the adults' use of prosody in the NF-f condition but not so in the NF-i condition (section 2.3). Thus, the children had difficulty with the focus-prosody interface in the NF-i condition in both production and comprehension. Our results thus show at a more fine-grained level that production and comprehension are symmetrical when we treat children as a homogenous group.

The picture is quite different regarding the relationship between production and comprehension in each child. There was not just one type of relationship between production and comprehension present in the data. Zooming in on the production and comprehension scores of each child, we have identified four kinds of relationships between production and comprehension: (1) poor production, poor comprehension; (2) good production, good comprehension; (3) poor production, good production; (4) good production, poor comprehension. These results suggest some degree of independence in learning to use prosody to mark focus in production and learning to process focus-related prosodic information in online language comprehension. Having the representation of the focus-prosody interface does not guarantee success in the actual production at the age of 4 or 5. Conversely, being able to use prosody accurately in focus marking does not entail the ability to integrate focus-related prosodic information into online language comprehension at this age. Possibly, the children who did well in both production and comprehension differed from those who did poor in both or did well in either production or comprehension in other aspects of language development and/or in other areas developmentally. Future research is needed to have a better understanding of the causes for the striking individual differences in the production-comprehension (a)symmetry.

6. Acknowledgements

This study is supported by a VIDI grant (276-89-001) from the NWO (Netherlands Organisation for Scientific Research). A big thank-you goes to the children and teaching staff from Houten Montessori Primary School, De Ontdekkingstreis Primary School (Doorn), Soest Montessori Primary School, De Wegwijzer Primary School (Soest) for their indispensable cooperation in this research. We thank Paula Cox, Martine Veenendaal, Saskia Versteegen for administering the tests. We also thank Paula Cox for drawing the pictures, and Frank Bijlsma, Alex Manus, Sjef Pieters and Theo Veenker for technical support, Tom Lenz and Mattis van den Bergh, Huub van den Bergh for statistical advice, Ao Chen, Nivja de Jong, Xiaoli Dong, Zenghui Liu, René Kager, Anna Sara Romøren, Anqi Yang and Wim Zonneveld for their feedback.

7. References

- [1] Clark, E. (1993). *The Lexicon in acquisition*. Cambridge: CUP.
- [2] Hendriks, P., and Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua*, 120 1887-1897.
- [3] Lambrecht, K. (1994). *Information structure and sentence form: Topics, focus, and the representations of discourse referents*. Cambridge: CUP.
- [4] Vallduví, E., and Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics*, 34(3): 459-520.
- [5] Ladd, D. R. (2006). *Intonational Phonology*. Cambridge: CUP.
- [6] Birch, S., Clifton, C.J., (1995). Focus, accent, and argument structure: effects on language comprehension. *Language and Speech*, 38 (4): 365-391.
- [7] Dahan, D., Tanenhaus, M. K., and Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, 47: 292-314.
- [8] Cutler, A., and Swinney, D. A. (1987). Prosody and the development of comprehension. *Journal of Child Language* 14: 145-67.
- [9] Chen, A. (2010). Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping. *Lingua*, 120: 1926-1939.
- [10] Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: CUP.
- [11] Chen, A. (2011a). Tuning information packaging: intonational realization of topic and focus in child Dutch. *Journal of Child Language*, 38: 1055-1083.
- [12] Chen A. (2009). The phonetics of sentence-initial topic and focus in adult and child Dutch. In M. Vigário, S. Frota and M. J. Freitas (eds.), *Phonetics and Phonology: Interactions and interrelations* (pp. 91-106). Amsterdam: Benjamins.
- [13] Hanssen, J., Peters, J., and Gussenhoven, C., (2008). Prosodic effects of focus in Dutch declaratives. In: Barbosa, P.A., Madureira, S., Reis, C. (Eds.), Proceedings of the 4th International Conferences on Speech Prosody, Editora RG/CNPq, Campinas, pp. 609-612.
- [14] Wells, B., Peppé, S., and Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language*, 31: 749-78.
- [15] Boersma, P., and Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.60, retrieved from <http://www.praat.org/>
- [16] Nootboom, S. G., and Kruyt, J. G. (1987). Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *Journal of Acoustical Society of America*, 82: 1512-1524.
- [17] Terken, J., and Nootboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and cognitive processes*, 2: 145-163.
- [18] Veenker, T.J.G. (2013). The Zep Experiment Control Application (Version 2.0) [Computer software]. Beexy Behavioral Experiment Software. Available from <http://www.beexy.org/zep/>