UNBALANCED ADULT PRODUCTION AND PERCEPTION IN PROSODY

Tomas O. Lentz and Aoju Chen

Utrecht Institute of Linguistics OTS, Utrecht University
t.o.lentz@uu.nl, aoju.chen@uu.nl

ABSTRACT

Previous work on the production and comprehension of focus-to-prosody mapping make different claims on the existence of a production-preceding-comprehension asymmetry in acquisition, whereby children pass through a stage of producing the right prosodic pattern but not using the focus-to-prosody mapping to the fullest extent in their comprehension, before reaching adult-like behaviour in both. The question that arises is whether there is an asymmetry in adults' use of prosody to mark and interpret focus as well. We found that adults' production correlated negatively with their comprehension. This puzzling result can be explained by exploring the data: many participants performed at the expected 'adult' level in one of the two skills, but not in both. We have therefore shown for the first time that adults are not balanced in their production and comprehension of prosody, which has important implications for acquisition.

Keywords: prosody, ultimate attainment, production-comprehension asymmetry, individual differences, focus

1. INTRODUCTION

Earlier studies on the acquisition of prosody as a means of marking (pragmatic) focus have suggested that pre-schoolers have an asymmetric pattern with regard to production and comprehension (perception). Adult-like production of prosody precedes comprehension (in contrast to the more common pattern where perception precedes production; e.g., see [4, 5] for an overview). Recent studies report that children's comprehension is at par with their production, suggesting that the supposed asymmetry could be due to the comprehension task being too difficult, leading to a score that underestimates comprehensive capacity of children [4].

The acquisition asymmetry literature is based on a comparison between children and adults on a group-level, with the adult-like level defined as the mean score of adult control groups. Hence, there is an implicit assumption that adult comprehension and production are balanced, i.e., perform similarly on both tasks (probably at ceiling). To verify this assumption, this paper presents a preliminary analysis of individual differences between adults' use of prosody in comprehension and production, initially gathered to make comparison between adults and children possible.

We analysed individual differences in two ways. First, we tested the hypothesis that adults production and comprehension capacities do not correlate positively (the null hypothesis being therefore that they do). The idea behind this null hypothesis is that prosody/focus mappings are one construct that emerges in both the use of prosody in production to mark focus and in the sensitivity to prosody in sentence comprehension. If this were the case, adults' comprehension and production should be closely correlated (possibly spuriously, when both were at ceiling, which would mean it cannot be confirmed statistically as all the data would be in one point, not on a line, even though every line can be fitted through the point). On the other hand, if adults' comprehension and production vary considerably and independently, it is possible that they are better considered separate skills [14]. Second, we investigate the size of individual differences and the variation in ultimate attainment, which would even be informative if the general correlation was positive, as individuals might still differ from a general (group-level) tendency.

We used a subset of data obtained from a battery experiments with Dutch native speakers. These experiments had the form of a game to facilitate comparison with experiments with children; one experiment measured production and one measured comprehension in the form of sentence processing speed. In both experiments, narrow focus in Dutch was tested; these production and comprehension scores were used for a correlation analysis. A focused constituent is (here) defined as one containing information that is new to the hearer, following [13] (cf. [12]). In Dutch, such constituents are supposed to be marked prosodically with an appropriate accent (for phonetic and phonological details, see e.g.,[11, 7] and for English [6]). As an example, consider the sentence 'John ate an apple'; if the previous discourse contained John and the eating event,

but not the apple, the speaker might accent the NP 'an apple' to convey its focus status to the listener, leading to the utterance 'John ate AN APPLE'. Small caps indicate focus here and in the rest of the paper.

If there are individual differences in prosody production, we expect phonetic and phonological marking of focus to differ between participants. As there is no consensus on the exact parameters of correct focus marking and their relative importance, the accuracy of focus marking in production was rated by three trained professionals. Comprehension scores were directly based on participants' reaction time; as correct focus marking is expected to facilitate processing, participants are expected to speed up when they have to process correctly marked focus as compared with incorrectly marked focus [5, 3]. Individual differences between participants in comprehension were operationalised as the reaction time reduction caused by correct prosody.

2. METHOD

2.1. Participants

18 participants (age: range 18–35, mean 21, s.d. 2.3; 7 females) participated in the experiments. These participants were recruited from the Utrecht University participant pool and received a modest monetary compensation. Two participants had to be excluded from the analysis, as they did not complete both experiments correctly.

2.2. Production data: procedure and materials

A picture-matching task was used to elicit production data that was later rated for appropriateness of the prosody. Experimenter and participant engaged in a guided conversation that was primarily composed of short question-answer dialogues, in which the participant had access to pictures of a complete event with three elements (subject, action and object), while the experimenter, who had incomplete pictures with two of the three elements, asked for the other one. To make the use of definite noun phrases (e.g., 'the girl' and 'the carrot') likely, participants named these objects and persons in a picture naming task before the matching task.

The participant answered with an SVO sentence (e.g., 'What does the girl cook? The girl cooks THE CARROT', vs. 'Who is cooking the carrot? THE GIRL is cooking the carrot'; small caps indicate focus). The audio was recorded and at a sampling rate of 44.1 kHz with 16 bits resolution. Correctly elicited full-sentence responses and corresponding questions were combined into dialogues (with a 1000 ms interval between question and response). Three trained native speakers of Dutch listened to these dialogues and evaluated the appropriateness of the prosody in each response on a five-point scale from 'does not fit' to 'fits perfectly'. The ratings were averaged per speaker and per focus condition, yielding two scores per speaker, one for initial focus (i.e., on the subject) and one for final focus (i.e., on the direct object).

2.3. Comprehension data: procedure and materials

The comprehension experiment also had a game-like form. Participants were told to accept or reject answers given to a boy by one of his three pets. Participants were shown pictures about events that contained the information the boy was asking for, while hearing one of the pets answer a question by the boy. For instance, participants could see that a girl was cooking a carrot and hear the boy ask 'Who is cooking the carrot?'. A pet answered with a semantically correct sentence, but either with appropriate prosody ('THE GIRL is cooking the carrot.') or inappropriate prosody ('The girl is cooking THE CARROT.')

Twenty-four answer sentences were used to create 24 different dialogues. Different lists were created, such that each answer sentence was presented only once to each participant, but overall (between participants), each sentence was presented with both initial (subject) and final (object) focus, and both when this focus was incorrect and when it was correct Twenty filler trials in which the pets gave incorrect answers were added to avoid a yes-bias.

Each trial proceeded as follows: the target picture (and a picture of the boy and a pet) appeared, acoustically accompanied by the attention getter 'look' said by the boy. 800 ms later, the boy named an entity in the picture. 1200 ms later, the boy asked a question, making either the subject or the object of the unknown (and hence new information in the answer, thus forcing focus). 2200 ms after the question, the pet answered the question, after which the picture of a push button appeared on screen. Participants could then answer by pressing a key with a green label for a correct answer or a red label for an incorrect answer. The reaction times of correct answers, measured from the end of each sentence, were log-transformed and then further analysed.

2.4. Correlation analysis

To see if production and comprehension performance correlate, linear models were created for the Production and Comprehension scores to obtain random effects per Participant (indicating the deviance

from group behaviour [9, 1]), generally following the approach of [8] (see also the online explanation, www.danmirman.org/gca). The models were built using the lme4 package [2] in R 3.1.2 [10]. The Production model only contained an intercept as fixed effect, as Focus Position (initial vs. final) did not significantly improve the model fit and was left out to avoid overfitting. The dependent variable was the average rating.

The Comprehension model contained the Prosody condition (inappropriate or inappropriate) as a fixed effect and a random effect for Participant × Prosody. The dependent variable was the log-transformed reaction time. Focus Position again did not significantly improve the model and was left out. Note that the fixed effect of Prosody was also not significant overall ($B = -0.003$, $t = -0.06$, $p = 0,95$), which does not mean that no participant was sensitive to prosody, but that there was no evidence for such an effect at the group level, possibly because of different effects of Prosody for different Participants. The fixed effect of Prosody matters for the random effect estimates and therefore had to be left in the model.

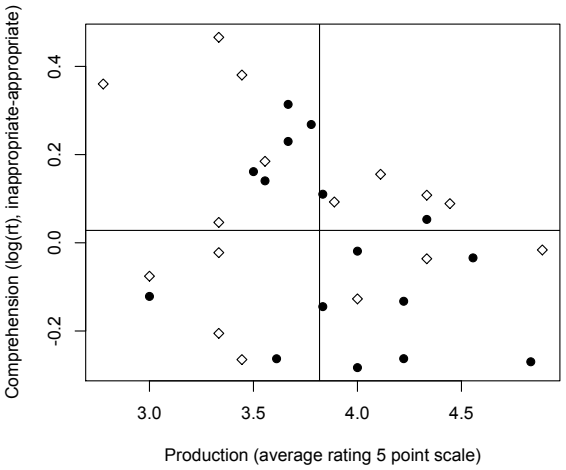
As the linear models assign individual variation at the participant level to the random effects for participant, these random effects were extracted from both models. For the comprehension score, the random effect per Participant was calculated as the difference between the random effect for inappropriate prosody minus the random effect of inappropriate prosody. A negative value indicates that less sensitivity to prosody than the group mean, because the participant is either faster for inappropriate prosody (below the mean) than (s)he is for appropriate prosody, indicating, or further above the mean for appropriate accentuation than for inappropriate.

Shapiro-Wilk normality tests did not suggest that production and comprehension random effects for participants were not normally distributed ($W = 0,98$, $p = 0.93$; $W = 0.97$, $p = 0.40$); neither did visual inspection of quantile plots and histograms. Hence, the data could be submitted to a correlation analysis. Production and comprehension scores were negatively correlated (Pearson's correlation: $r = -0.54, t(14) = -2.40, p = 0.016$* (one-tailed)). Less sensitivity to prosody in comprehension is thus associated with better ratings for production.

2.5. Further data exploration

As the null hypothesis that Production and Comprehension are positively correlated has to be rejected, we looked for patterns that could possibly explain the imbalance. Note that the current part of the present paper should be interpreted as an exploration of the data, rather than hypothesis testing. Figure 1 shows how the data is distributed. The measured values, rather than the random effects, are plotted, as these are easier to interpret visually. Note that each participant is represented by two data point in the plot (but not in the analysis presented above, as these data points are not independent). Mean values are indicated by solid lines.

Figure 1: Comprehension and Production scores per participant per focus position (black circles: initial, white squares: final). The solid lines indicate the means for Comprehension and Production.



The scatterplot in Figure 1 suggests a scarcity of data points (hence participants) that performed (much) better than average on both comprehension and production. The negative correlation is in accordance with this suggestion, but could theoretically also be present with fewer participants performing below average on both tasks. To see if more observations fell in the different quadrants of the plot, a cross-tabulation was made. Data points were simplified to two binary values: above or below-mean performance, for comprehension and for production. Six data points were below average on both scores and another six above average on both. On the other hand, ten data points were above average on Production but below average on Comprehension, and another ten the other way around. However, this distribution was not significantly different from random according to a chi-square test (Pearson's $\chi^2(1) = 2$ with Yates' continuity correction, $p = 0,29$). Another interpretation of the data is that twelve data points were as expected if production and comprehension were positively corre-

lated, while twenty were not. However, this difference in frequencies is also not significant (Pearson's $\chi^2(1) = 2$, $p = 0.157$). Both frequency differences suffer from a limited number of data points; the first test would be significant with three times as many data points, the last with two times as many, under the tentative assumption that the distribution of the observations represents the underlying distribution faithfully.

3. DISCUSSION AND CONCLUSION

The rejection of positive correlation means that adults are not generally balanced on their production and comprehension of prosodic focus marking. Participants that attained the best scores on either comprehension or production did not tend to perform well on the other task. Few participants scored above average on both production and comprehension. Hence, it is problematic to assume that group means represent some ceiling level of ultimate attainment of the focus/prosody mapping. To quantify this interpretation, a crude categorisation of the data as below or above average was made; although this grouping did not turn out to be significant, significance might be obtained with more data points.

If adults are not balanced in their production and comprehension of prosody, the interpretation of acquisition data should be re-evaluated; children that are found to be at a so-called adult level in production but below a so-called adult level in comprehension might be at (their own) adult level, as adults with such imbalances exist. Hence, more data is needed to really shed light on the possibility of imbalance between prosodic comprehension and production tasks in adults.

4. ACKNOWLEDGEMENTS

The authors wish to thank Joe Rodd, Saskia Verstegen and Martine Veenendaal for executing the experiments and Zenghui Liu, Anna Sara Romøren and Anqi Yang for valuable feedback. This research was made possible by VIDI grant 276-89-001 from NWO (Netherlands Organisation for Scientific Research) and an Utrecht University Dynamics of Youth Talent Grant, both to the second author.

5. REFERENCES

[1] Baayen, R. H., Davidson, D. J., Bates, D. M. 2008. Mixed effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4), 390–412.

[2] Bates, D., Maechler, M., Bolker, B. M., Walker, S. 2014. lme4: Linear mixed-effects models using eigen and s4. ArXiv e-print; submitted to *Journal of Statistical Software*.

[3] Birch, S., Clifton, C., Jr. 1995. Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech* 38(4), 365–391.

[4] Chen, A. 2010. Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua* 120, 1926–1939.

[5] Cutler, A., Swinney, D. A. 1987. Prosody and the development of comprehension. *Journal of Child Language* 14(1), 145–167.

[6] Dahan, D., Tanenhaus, M. K., Chambers, C. G. 2002. Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 47(2), 292–314.

[7] Hanssen, J., Peters, J., Gussenhoven, C. 2008. Prosodic effects of focus in dutch declaratives. Barbosa, P. A., Madureira, S., Reis, C., (eds), *Proceedings of fourth International Conference on Speech Prosody in Campinas, Brazil* Campinas. Editora RG/CNPq 609–612.

[8] Mirman, D., Yee, E., Blumstein, S. E., Magnuson, J. S. 2011. Theories of spoken word recognition deficits in aphasia: Evidence from eye-tracking and computational modeling. *Brain & Language* 117, 53–68.

[9] Quené, H., van den Bergh, H. 2008. Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language* 59(4), 413–425.

[10] R Development Core Team, 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna.

[11] Swerts, M., Krahmer, E., Avesani, C. 2002. Prosodic marking of information status in Dutch and Italian: a comparative analysis. *Journal of Phonetics* 30, 629–654.

[12] Vallduví, E. 1993. *Information packaging: A survey*. Edinburgh: Edinburgh Centre for Cognitive Science.

[13] Vallduví, E., Engdahl, E. 1996. The linguistic realization of information packaging. *Linguistics* 34(3), 459–519.

[14] Williams, B., Myerson, J., Hale, S. 2008. Individual differences, intelligence, and behavior analysis. *Journal of the Experimental Analysis of Behavior* 90(2), 219–231.