A closer look at pronoun comprehension: comparing different methods

Christina Bergmann, Markus Paulus, Paula Fikkert
Radboud University Nijmegen, The Netherlands

1. Introduction

External input is necessary to acquire language. Consequently, the comprehension of various constituents of language, such as lexical items or syntactic and semantic structures should emerge at the same time as or even precede their production. However, in the case of pronouns this general assumption does not seem to hold. On the contrary, while children at the age of four use pronouns and reflexives appropriately during production (de Villiers, et al. 2006), a number of comprehension studies across different languages found chance performance in pronoun trials up to the age of seven, which co-occurs with a high level of accuracy in reflexive trials (for an overview see e.g. Conroy, et al. 2009; Elbourne 2005).

From a theoretical perspective, only syntactic rules are assumed to affect the interpretation of reflexives, which receive their interpretation from a previously mentioned antecedent within the same sentence. Pronouns, on the other hand, require semantic and pragmatic considerations to identify the correct antecedent and their interpretation is affected by phenomena such as focus and prominence (e.g. Chomsky 1981; Hendriks and Spenader 2006; Reinhart 1983). This implies more complex resolution processes to find the correct interpretation of pronouns, which in turn poses a higher demand during language comprehension than reflexives. This has previously been pointed out by for example Bloom and colleagues (1994) and Grodzinsky and Reinhart (1993).

Can this additional demand during pronoun comprehension alone explain the year-long asymmetry in production and comprehension and the strong advantage of reflexives? As noted by Matthews and colleagues (2009), no linguistic account so far can satisfyingly capture children's behaviour in studies regarding pronoun acquisition. An alternative or additional explanation of previous results would be an impact of the experimental setting itself on children's performances. The methods usually employed to investigate child language comprehension, such as act-out-tasks, generally require the generation of an action in response to a previously comprehended sentence (for an overview of tasks used and the variety of outcomes see Conroy, et al. 2009). This naturally poses additional demands on

top of mere language comprehension. A recently emerging and alternative method to tap into language comprehension facilities is eye-tracking, which does not require a secondary task and allows insight into language comprehension as sentences unfold. This method has been applied successfully in both adults (e.g. Tanenhaus, et al. 1995) and children (e.g. Trueswell, et al. 2004).

In our study, we closely examined a picture selection task, which is commonly applied in studies investigating the acquisition of pronoun comprehension (e.g. Deutsch, et al. 1986), and contrasted the results with eye-tracking to further assess the factors influencing the acquisition of pronouns and to disentangle methodological concerns from actual linguistic knowledge.[1]

2. Experiments

We tested children in three age groups (age three, four and five) with two methods. Across tasks, we used the same linguistic and visual stimuli to allow for a comparison of performance within participants.

2.1 Participants

For age group three, 22 children with a mean age of 3;0.04 (14 male, 2;11 to 3;1 years old) were tested. Sixteen children with a mean age of 3;10.04 (eight male, 3;9 to 3;11 years old) constituted the age group four. For the oldest age group of five-year-olds, 16 children with a mean age of 4;11;26 (seven male, 4;10 to 5;1 years old) participated. Six additionally tested children had to be excluded due to fussiness and uncooperative behaviour. All parents gave informed consent before participation. As a reward, the parents could choose between a book and monetary compensation. All participants were monolingual speakers of Dutch, recruited via the database of the Baby Research Centre Nijmegen, The Netherlands.

2.2 Materials

In the pronoun and the reflexive condition, identical visual and linguistic stimuli were used across age groups and tasks. For the linguistic stimuli, this meant varying the referential expression in otherwise identical sentences. Sentence (1) gives an example.

(1)   Meneer Beer is (hem | zich) aan het aankleden.
            (*Mister Bear is dressing (him | himself).*)

---

1 For a thorough description and discussion of the stimulus material, the procedure and data analysis, please see Bergmann, Paulus and Fikkert (submitted).

Accordingly, the visual stimuli showed either a transitive or a reflexive action with the subject of the sentence as agent and the second figure either as a patient or as unrelated distractor.

To account for a modulation of fixation behaviour by the visual stimuli alone, which would be picked up during eye-tracking, a baseline was constructed. In the baseline, the visual stimuli matched those seen in the same session paired with pronoun or reflexive sentences; the acoustic stimuli, however, only consisted of an exclamation and a general positive statements, such as given in (2).

(2)   Kijk! Wat mooi!
            (*Look! How nice!*)

Onset times were paired across conditions, with the positive statement in the baseline condition starting at the same time as the pronoun or reflexive. This allowed for comparing the fixation behaviour when encountering a reflexive with a pronoun while at the same time accounting for effects of the visual stimuli.

2.3 Procedure

Each experiment started with a play session. During that initial warm-up phase, the children were familiarised with the toys and their associated names used throughout the experiment. Both during eye-tracking and in the picture selection task, 12 pronoun and 12 reflexive trials were conducted.

2.3.1 Eye-tracking

During eye-tracking, the children either sat on their parent's lap (age groups three and four, 11 of the children in age group five) or on their own with their parent on a separate chair behind them (five children in age group five). The parents heard a mixture of music and spoken language throughout this part of the experiment. The children were asked to watch a short movie, without any additional information or task.

2.3.2 Picture Selection Task

Children sat in front of a wooden display and saw two cards at a time. Both cards showed the same pair of referents, with the subject of the sentence either performing a transitive or reflexive action. The second toy figure was the same in both pictures. Children were asked to point out the card that best fit the sentence they heard after the pictures were put on the display. After each trial, the child was praised, irrespective of his or her actual performance.

3. Results and Discussion

In the picture selection task, as predicted from previous experimental and theoretical findings, the children performed on chance level across age groups when confronted with a pronoun. Reflexives, however, see an increase in mean performance, which is significantly different from chance level from the age of four onwards (see figure 1).
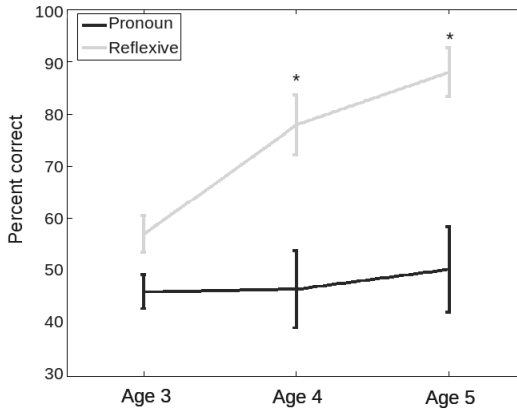

Figure 1: Accuracy in pronoun and reflexive trials across age groups. The bars indicate 1 SE, significant difference from chance performance is indicated.

Eye-tracking, however, presents a different behaviour. The fixation behaviour following the encounter of a pronoun or reflexive can be inspected in figure 2. The depicted measure shows the mean increase in fixation probability on the correct referent in a time-slot of 200 to 800 ms after onset of the referential expression. Any effect of mapping the referential expression onto a referent on-line should be expected to occur within that time-frame (e.g. Tanenhaus, et al. 1995; Trueswell, et al. 2004). The fixation behaviour, or more precisely the change, shows, that during eye-tracking, children as young as four years old already directed their gaze towards the correct referent of a pronoun. This behaviour was significantly different from the baseline (not depicted). In the reflexive trials, all age groups fixated the correct referent significantly more than in the baseline trials.
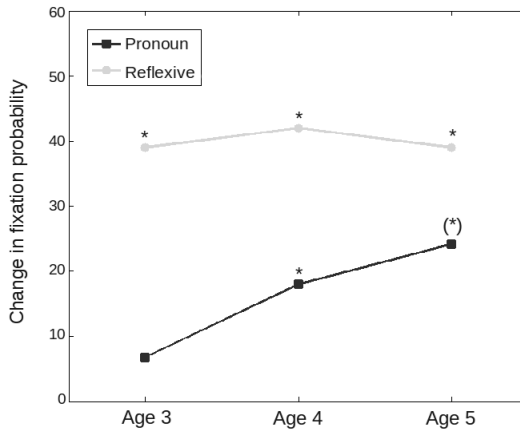
Figure 2: Increase of fixations on the correct referent upon hearing a pronoun during eye-tracking. (The baseline condition did not show a systematic increase and is omitted for clarity. Significant testing employed direct comparison to the baseline and is indicated.)[2]

These results of the two tasks on pronoun comprehension are seemingly conflicting, as eye-tracking indicates knowledge of the correct referent emerging at the age of four, whereas the picture selection task points to a confusion when encountering pronouns until at least the age of five. This conflict in the results across tasks suggests an impact of the experimental method on the performances of the children. Two conclusions can be drawn from the discrepancy between both tasks: First, there is a difference between tasks. Eye-tracking is sensitive to automatic responses and measures comprehension on-line, whereas the picture selection task results only account for an integrated, late result of the comprehension process. This result is in line with previous findings, such as for example reported by Sekerina and colleagues (2004): In their study, eye-tracking showed short-lived considerations and decisions in the ongoing process of language comprehension, which were not visible in a parallel pointing task. This finding shows, that the results of tasks can be modulated by non-linguistic effects and hence do not only reflect linguistic competence.

The second conclusion is, that pronoun comprehension is more susceptible to task effects than reflexives and can be assumed to be less stable in the age groups

---

2   For the significance tests and detailed eye-tracking results, please see Bergmann, Paulus, and Fikkert (2009, 2010).

investigated. This interpretation is supported by a number of theoretical accounts, as discussed previously, where reflexives are usually associated with a relative fast and automatic resolution, whereas pronouns require more complex processing, as the referent is usually not found within the same clause as the pronoun. Rather, it is provided by the previous discourse or indicated deictically.

To further examine the picture selection task and its effect on the children's behaviour and accuracy, individual performances were inspected. This led to the discovery of strategic behaviour[3] across age groups, which manifested itself in two distinct phenomena: At the age of three, 55% of the participants (12 of 22 children) showed a preference for choosing one side across trials and independent of the condition of the sentence heard or depicted. This behaviour implies no effect of either the acoustic or the visual input during the picture selection task. This result, however, stands in contrast with the eye-tracking data, which show a modulation of fixation behaviour in the three-year-olds when encountering a reflexive.

In the older age groups, namely with the four and five-year-olds, a different strategy was observed. Eighteen of the 32 participants (56%) in those combined age groups pointed to the picture depicting the reflexive action in most trials, irrespective of the actual sentence heard. This behaviour masked the performance of 73 % correct choices in the pronoun trials in the remaining 14 children. This accuracy is significantly different from chance level. These results show that children do not perform on chance level; they rather fall into two distinct groups significantly different from chance performance, since the strategically behaving children perform significantly below chance level in the pronoun trials. This finding is in contrast with previous reports and accounts of pronoun comprehension, which would have predicted chance performance for each child and not only in the group averages.

Dynamic systems theory allows for an integrated interpretation of this puzzling behaviour. As Smith and Thelen (2003) suggested, a child's overt behaviour is the result of a number of mutually influencing components, that change across different time lines (from the fast decay of action potentials in neurons up to the establishment of habits over months and years). These components and their interaction result in a non-linear dynamic system, that fits the child's behaviour and its multiple causes better than a unidimensional approach.

In the data at hand, this reasoning implies that children's behaviour is not only guided by their linguistic knowledge, or a lack thereof, during the picture selection

---

3   The systematicity of strategic behaviour was tested separately for each subject based on the number of successfully completed trials (at least 20) using a binomial test. The average cutoff was 70% of the trials or more.

task. Consequently, the results do not reflect linguistic competence. Rather, the responses merely imply that constantly preferring one side or one type of picture is a strong attractor at the time point when the picture selection task is conducted or in the context of this task in general. At the age of three and in the case of preferring one side, the attractor might be based on the preferred hand of the children. This effect is enhanced over time, as a motor memory of this action is build and subsequently serves as the strongest attractor in the different test trials (e.g., Diedrich, et al. 2000). On-line, in contrast, as seen during eye-tracking, there is a linguistically influenced attractor towards the correct interpretation of reflexives at the age of three and towards the correct interpretation of both referential expressions from the age of four onwards. The attractor that guides behaviour during eye-tracking either decays too fast to be picked up by the picture selection task or is changed by the additional factors present. The non-strategically behaving children in the age group of three-year-olds are at chance level and seemingly chose a picture randomly. This suggests no strong attractor in those children during the picture selection task.

The behaviour of the older children, who preferred the reflexive pictures during the picture selection task, can be captured by a similar line of reasoning. Here, reflexives constituted a strong attractor state during the picture selection task. This suggestion is supported by the fixation behaviour, as previously discussed with respect to the three-year-olds, which indicates a change over time or a stronger strategic attractor state when a behavioural response is required. The comparatively stronger attraction of reflexives across linguistic input can be accounted for by a faster resolution process, as laid out previously.

For the non-strategically behaving children, the strongest attractor in both eye-tracking and the picture selection task is seemingly the correct interpretation, which indicates a more robust knowledge concerning the correct interpretation of pronouns and reflexives.

It has to be noted, that in both strategies, the praise after each trial might have encouraged this behaviour by reinforcement. Further investigation is required to examine this additional factor.

In summary, we found a strong task effect, that included a variety of strategies. This task effect masked individual proficiency in the older age groups and thus helped draw a misleading picture of children's linguistic competences. A multicausal dynamic approach helped explain seemingly contradicting data.

References

Bergmann, C., Paulus, M., and Fikkert, P. 2009. *Task effects in pronoun interpretation: Methods under investigation*. Manuscript submitted for publication.

Bergmann, C., Paulus, M., and Fikkert, P. 2010. *Time course of pronoun comprehension: The case of gender*. Manuscript in preparation.

Bloom, P., Barss, A., Nicol, J., and Conway, L. 1994. Children's knowledge of binding and coreference: Evidence from spontaneous speech. *Language* 70(1): 53-71.

Chomsky, N. 1981. *Lectures on government and binding: the Pisa lectures*. Dordrecht, The Netherlands: Foris Publications.

Conroy, A., Takahashi, E., Lidz, J., and Phillips, C. 2009. Equal treatment for all antecedents: How children succeed with Principle B. *Linguistic Inquiry* 40 (3): 446-486.

De Villiers, J., Cahillane, J., and Altreuter, E. 2006. What can production reveal about Principle B? In K. U. Deen, J. Nomura, B. Schulz, & B. D. Schwartz (Eds.), *Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition–North America* (pp. 89–100). University of Connecticut: Occasional Papers in Linguistics 4.

Deutsch, W., Koster, C., and Koster, J. 1986. What can we learn from children's errors in understanding anaphora? *Linguistics* 24: 203–225.

Diedrich, F. J., Thelen, E., Smith, L. B., and Corbetta, D. 2000. Motor memory is a factor in infant perseverative errors. *Developmental Science* 3: 479-494.

Elbourne, P. 2005. On the Acquisition of Principle B. *Linguistic Inquiry* 36(3): 333-365.

Grodzinsky, Y. and Reinhart, T. 1993. The innateness of binding and coreference. *Linguistic Inquiry* 24(1): 69–101.

Hendriks, P. and Spenader, J. 2006. When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition* 13(4): 319-348.

Mattews, D., Lieven, E., Theakston, A., and Tomasello, M. 2009. Pronoun co-referencing errors: Challenges for generativist and usage-based accounts. *Cognitive Linguistics* 20(3): 599-626.

Reinhart, T. 1983. Coreference and bound anaphora: A restatement of the anaphora questions. *Linguistics and Philosophy* 6(1): 47-88.

Sekerina, I., Stromswold, K., and Hestvik, A. 2004. How do adults and children process referentially ambiguous pronouns? *Journal of Child Language* 31(1): 123-152.

Smith, L. B. and Thelen, E. 2003. Development as a dynamic system. *Trends in Cognitive Sciences* 7: 343–348.

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., and Sedivy, J. 1995. Integration of visual and linguistic information during spoken language comprehension. *Science* 268(5217): 1632–1634.

Trueswell, J. and Gleitman, L.R. 2004. Children's eye-movements during listening: Evidence for a constraint-based theory of parsing and word learning. In J. M. Henderson & F. Ferreira (Eds.). *Interface of Language, Vision, and Action: Eye-movements and the Visual World*. New York: Psychology Press.