Comprehension and Production of Definite and Indefinite Noun Phrases in English Preschoolers

Angeliek van Hout, Kaitlyn Harrigan, and Jill de Villiers
University of Groningen, Smith College and Smith College

1. Introduction

It is well known that children use indefinite and definite noun phrases in a non-target-like way. Children up to the age of 5 or 6 use definite NPs as well as indefinite NPs when they refer to a new discourse entity, where the target grammar requires an indefinite. This so-called *the*-overuse is well-established and has received a variety of explanations. There are many fewer studies that examine children's interpretation of definite and indefinite articles. Ours is the first study that systematically compares comprehension and production of definite and indefinite articles within the same set of children. By adding comprehension data to production data our comparison puts the old production findings, which we indeed replicate, in a new light. We find a mirror image of the overuse of *the* in production in the over-allowance of associating *the* with a new discourse referent in comprehension. We look for a generalized explanation for both phenomena.

Our explanation of children's incomplete knowledge, framed in Bidirectional Optimality Theory, posits that children take an exclusive hearer perspective when they face a comprehension task, and an exclusive speaker perspective when they face a production task (de Hoop & Krämer, 2005/6). This is in contrast to adults who integrate both perspectives when they are speaking and understanding, which triggers strong preferences for certain form-meaning mappings and blocks others. Children lack the distinction between preferred versus blocked-out mappings because they fail to take into account the other perspective. In addition we believe that children have a more liberal mechanism for establishing anaphoric relations than adults, in particular, for children the visual context can establish an antecedent for a discourse anaphor (e.g., a definite NP), whereas adults require a linguistic antecedent in the verbal context. We furthermore argue that various constraints interact in different ways for children and adults.

After reviewing some acquisition literature in section 2, we formulate our research question. We propose a constraint-based analysis for definite and indefinite articles framed in a bi-directional Optimality Theory model in section 3 and formulate predictions under the hypothesis that children lack bidirectional reasoning. We then describe our experiments and the results in section 4. Section 5 presents the discussion of the results and the conclusions for the OT analysis. Section 7 concludes the paper with suggestions for follow-up research steps.

2. Acquisition studies on definite and indefinite noun phrases

In Karmiloff-Smith's (1979) "playroom" task, there were many objects in front of the subject. The experimenter touched either a singleton object or one of several identical objects and says, "Ask the doll to lend you that,", the singleton object targeting *the*, and one of several identical objects targeting *a*. In the "girl/boy acting" task, a doll knocks over an object, either a singleton or one of three identical objects and experimenter asks child, "What did the doll do?". In this task (and various others) preschoolers overused *the* in the *a*-targeting condition by 39-63% with one of several identical objects, while correctly producing *the* 100% in the *the*-eliciting condition with a singleton. According to Karmiloff-Smith *the*-overuse happens because children use definites deictically, instead of anaphorically taking into account the listener's knowledge of the "current context". We note, however, that in this particular set-up, in which the experimenter or a doll touches a certain object, the use of *the* (or *that*) is a natural response, even in the indefinite condition, as a specific item is singled out visually.

 Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 2008), ed. Jean Crawford et al., 76-87. Somerville, MA: Cascadilla Proceedings Project.

Using a purely verbal task without any visual support Maratsos (1974) examined the production of definite or an indefinite noun phrases in children between 3 and 5 with stories such as (1). He also found an overuse of *the* in the *a*-targeting condition of 42%. Note that *the*-overuse occurs even when the referent is not present in the visual context. Maratsos attributes these mistakes to egocentrism; children do not care whether or not the listener knows the referent.

(1) Once there was someone who wanted to have an animal. He went out to the pond and …
   a. He saw two bunches of animals, lots of frogs and lots of turtles. He went up with his box, and he put one of them into his box. What did he put in? Target: a frog
   b. He saw two animals, a frog and a turtle. He went up with his box, and he put one of them in the box. What did he put in? Target: the frog

Schaeffer and Matthewson (2005) further examined children's (ages 2;1-3;10) production of articles in relation to the common ground of speaker and hearer. In one condition (their condition B) only the speaker knows the referent, so it is not part of the common ground and requires the indefinite article *a*, see (2). In another (their condition C) neither speaker nor hearer know the referent, so it is not part of the common ground and requires the indefinite *a*, see (2).

(2) a. Experimenter shows picture of Mickey Mouse who just finished drawing a house and asks: What did Mickey Mouse just do? Target: He drew a house.
   b. Big Bird: Oh, I'm so bored. I don't know what to do. Oh, you know what, I'm going to the forest, and I'm gonna draw something there.
   Elmo: What do you think Big Bird is gonna do in the forest? Target: draw a tree

Schaeffer and Matthewson found *the*-overuse in condition B (about 25%), but importantly not in condition C. They claim that children lack a pragmatic principle which states that speaker and hearer assumptions are always independent, and thus in principle different. If children do not make a proper distinction between speaker and hearer assumptions, establishing a correct common ground is problematic and mistakes are expected in condition B, where speaker and hearer knowledge states differ. We note that in this set-up the referent is present in the visual context, and hence in the visual common ground, which may have triggered the use of definites.

Schafer and de Villiers (2000) looked at production of articles in children aged 3;6-5;5. Like Maratsos, they used a purely verbal context with no pictures or props. One condition involved specific indefinites, (3). In another, called the multipac condition illustrated in (4), several entities of the same kind are introduced in the preceding discourse context, similar to Maratsos' task in (1).

(3) I bet you have something hanging on your wall at home. What is it? Target: a picture

(4) Three ducks and two dogs were walking over a bridge. One of the animals fell off and said, "Quack!" What was it? Target: a duck

Schafer and de Villiers found *the*-overuse only in the multipac condition (between 49-64%), but importantly, no overgeneralization of *the* in the specific-*a* condition. The latter finding goes against Maratosos' theory of egocentricity: If children's misuse of articles reflected egocentricity, they would also overuse *the* in the specific-*a* condition, since they themselves are familiar with the item.

Wexler (2003) presents a different explanation for *the* overuse, arguing that children initially misconstrue the semantics of definite articles. For an adult *the* requires a unique referent, typically a referent previously introduced in the domain of discourse, and familiar to speaker and hearer, i.e., part of the common ground. For children—Wexler claims—definite articles carry an existence presupposition, but lack the uniqueness presupposition. In Maratsos' animal stories, children infer there exists an animal in the box, and so they answer *the frog (turtle)*. However, Wexler's explanation fails to explain the Schafer and de Villiers finding in the specific-*a* condition in (3). Wexler predicts *the*-overuse because there exists an object on the child's bedroom wall, but children uniformly produce *a*.

There are several recurring results across the different experimental set-ups reviewed above. In production we see consistent overuse of *the* in certain *a*-eliciting circumstances: when there is visual support and when a set of entities is mentioned in the previous discourse, though without a unique referent singled out. There is no overuse of *the* across the board, however. Children correctly use *a* when they introduce a new referent into the discourse (Schaeffer and Matthewson's condition C, Schafer and de Villiers's specific *a*-condition). Note also that nobody found an exclusive use of *the* at the expense of *a* in the indefinite conditions; the percentages vary between 17-58% across studies. The explanations that have been proposed miss this generalization. Any explanation that claims a cognitive problem (Maratsos' egocentrism), lack of a pragmatic principle (Schaeffer and Matthewson), or deviating semantics (purely deictic reference for Karmiloff-Smith, no uniqueness for Wexler), predicts not just overuse of *the*, but use of *the* to the exclusion of *a*, contrary to fact. A crucial characteristic of this stage of acquisition is optionality of *the* and *a*.

In one of few comprehension studies Van Hout, de Ree and de Ree (2008) tested the interpretation of definite and indefinite articles in Dutch children aged 4;1-5;4, using a truth-value judgment task accompanied by pictures. The pictures varied the referent of the second clause, whether it was the same referent mentioned in the first clause or a different one, as illustrated in (5).

(5) a. A mouse is kicking a ball and the mouse has a long tail.
True if referent is same, false if different
b. A cook is stirring the soup and a cook has a hat on.
True if referent is different, true if the same (but infelicitous)

Van Hout et al. found that children correctly associate definites with a mentioned referent, but do the same for indefinites. They also found that children correctly accept a new referent for indefinites, but also (to some extent) for definites. Van Hout et al. "blame" the visual context for over-acceptance of definites with new referents: for children the picture suffices to establish a new referent as "known" and so they allow a definite to refer to it. This explanation also explains *the*-overuse in production in tasks with a visual context (Karmiloff-Smith; Schaeffer and Matthewson), but does not carry over to the purely verbal tasks which also triggered *the*-overuse (Maratsos; Schafer and de Villiers).

In our study, we want to see if production problems are accompanied by comprehension problems, and if they are, if this holds for indefinites as well as definites. Alternatively we may also find an asymmetry between production and comprehension, for both or one of the articles. We combined production and comprehension tasks in one test battery, measuring the different modalities in parallel contexts. We thus contribute crucial new comprehension data to the old topic of article production.

3. Definite and indefinite noun phrases in bidirectional Optimality Theory

We couch our analysis of the adult use of definite and indefinite noun phrases in the framework of bidirectional Optimality Theory (bi-OT) by formulating two constraints about form-meaning associations. The analysis relies crucially on the requirement to take into account both speaker and hearer perspectives at the same time. We then argue that children's deviation from the adult grammar is caused by a lack of the ability to reason with both perspectives, following de Hoop and Krämer (2005/6) and Hendriks and Spenader (2005/6). Children furthermore let an additional constraint take priority.

For Heim (1982) and Kamp (1981) the difference between definite and indefinite noun phrases is semantic: indefinites introduce new referents in the domain of discourse, while definites pick out existing referents. Definite NPs carry the presuppositions of existence and uniqueness. They typically refer to referents previously introduced in the domain of discourse, and familiar to speaker and hearer. De Swart (in press) summarizes this classic view as an OT constraint: "A definite article corresponds with a discourse referent with determined reference". A discourse referent has determined reference if it is unique and/or has anaphoric reference (Farkas 2002). We employ De Swart's constraint and call it DETERMINED REFERENCE.

The choice between definite or indefinite articles thus depends on whether or not a discourse entity is already present in the discourse context, and if so, if it is uniquely identifiable. Gundel, Hedberg and

Zacharski (1993) order different kinds of NPs on a hierarchy from little to much informativity: Pronouns >> Definite NP >> Indefinite NP, and argue that speakers select the NP form on the scale that has just the right level of informativity given the previous discourse. Typically, when a certain discourse entity is given in the immediate discourse, a pronoun is used; when a new entity is introduced into the discourse, an indefinite NP is used; and when the entity has been mentioned in the discourse but is not the main focus of attention, a definite NP is used.

In an OT framework with soft, violable constraints, Gundel et al.'s Givenness Hierarchy has been reinterpreted as a hierarchy of markedness constraints (Spenader and Hendriks, 2005/06; Wubs, Hendriks, Hoeks and Koster, this volume) called Referential Economy (Burzio, 1998): Avoid indefinites >> Avoid definites >> Avoid pronouns. This hierarchy is the driving force for speakers to prefer pronouns over definite NPs, and definite NPs over indefinite NPs, because pronouns are the least costly NP forms. If all else is equal speakers would always speak in pronouns and never use definite and indefinite NPs. Of course pronouns are not always appropriate, nor are definite NPs, given their semantics and their discourse use, and this is where the interaction of the two constraints plays out. Since our study focuses on the choice between definite and indefinite NPs, we employ the first constraint of this family: AVOID INDEFINITES. Note that it only applies to production, unlike DETERMINED REFERENCE which applies to comprehension and production.

The tableaux illustrate this interaction for production (Tableaux 1 and 2 for the mapping from meaning to form) and comprehension (Tableaux 3 and 4 with the mapping from form to meaning).

Meaning input Determined reference	DETERMINED REFERENCE	AVOID INDEFINITES
a		*!
☞ *the*		

Tableau 1: The optimal form expressing a determined reference meaning

Meaning input Non-determined reference	DETERMINED REFERENCE	AVOID INDEFINITES
☞ *a*		*
the	*!	

Tableau 2: The optimal form expressing a non-determined reference meaning

In both production Tableaux 1 and 2, *a* violates AVOID INDEFINITES equally since indefinites are to be avoided. In Tableau 1 this is the only violation and so *the* is the optimal candidate. In Tableau 2 DETERMINED REFERENCE penalizes *the* for non-determined reference, which is the fatal violation because this constraint is higher ranked, and therefore *a* is the winner. Each tableau thus yields exactly one optimal candidate: <*the*, Determined reference> and <*a*, Non-determined reference>.

Form input *the*	DETERMINED REFERENCE	AVOID INDEFINITES
☞ Determined reference		
Non-determined reference	*!	

Tableau 3: The optimal meaning for interpreting *the*

Form input *a*	DETERMINED REFERENCE	AVOID INDEFINITES
☞ Determined reference		
☞ Non-determined reference		

Tableau 4: The optimal meaning for interpreting *a*

In Tableau 3 for *the*-comprehension, determined reference is the optimal meaning, because the other meaning violates DETERMINED REFERENCE. In Tableau 4 for *a*-comprehension DETERMINED REFERENCE does not apply because it only cares about *the*, and AVOID INDEFINITES does not apply

either because it is a production constraint. As a result both meanings are tied, and so this unidirectional tableau gives two output meanings for indefinites.

This is where adults in the interpretation process engage not only the hearer perspective, but also double-check the speaker perspective. Had the speaker intended the determined reference meaning, she would have used *the*, because that is the most optimal form for that meaning. Since she didn't use *the* but *a*, the speaker must have intended the other meaning: non-determined reference. This reasoning about comprehension-via-production thus blocks one of the two meanings for *a*. This is modeled in bi-directional OT, where a form-meaning pair that is optimal in both speaker and hearer perspectives—here <*the*, Determined reference> marked with ☝—blocks any association of *the* with another meaning as well as the association of the determined reference meaning with any other form. Blocking is indicated with strike-outs in the bi-directional Tableau 5. As a result of his optimization process two pairs win: <*the*, Determined reference> and <*a*, Non-determined reference>.

<Form, Meaning>	DETERMINED REFERENCE	AVOID INDEFINITES
☝ <*the*, Determined reference>		
~~<*the*, Non-determined reference>~~	*	
~~<*a*, Determined reference>~~		*
✌ <*a*, Non-determined reference>		*

Tableau 5: The optimal form-meaning pairs for production and comprehension

De Hoop and Krämer (2005/6) put forward the hypothesis that children cannot simultaneously entertain and compare speaker and hearer perspectives; in bi-OT terms, they do not reason bidirectionally (see also de Villiers, Cahillane and Altreuter, 2006; Hendriks, 2008; Hendriks and Spenader, 2005/6; van Hout, 2007; van Hout et al., 2008). Instead they rely on their unidirectional tableaux. This gives a straightforward prediction for the difference between the child's and the adult's use of the grammar of articles. In comprehension, adults block one meaning for indefinite articles and accept only non-determined reference, whereas children are tied between the two meanings, which should lead to guessing behavior. Lack of bidirectional reasoning does not yield any differences between children and adults for production because the unidirectional production tableaux each give one winner.

Our comparison of production and comprehension tasks tests these predictions. We already note here that this analysis cannot be the whole story in view of children's *the*-overuse, which we also find in our data, and so we will add another twist to this bi-directional explanation in section 6.

4. Comprehension and production experiments

4.1. Participants

Twenty five children were tested in the age range of 3;1-5;8 with a mean age of 4;0. The children were all from the Fort Hill Center for Early Childhood Education in Northampton, MA. In addition a small control group of six adults was tested (mean age 29 years, range 16-34). Each participant was tested individually in a quiet room, on a variety of production and comprehension tasks.[1] The responses were recorded by one researcher scoring as the tests were being administered by another researcher. The experiments were also being videotaped for later checking.

4.2. Production task

Like Schafer and de Villiers we elicited NPs with a purely verbal task with no visual support. We copied Schafer and de Villiers' multipac condition, varying the unique vs. non-unique nature of the target NP by introducing a referent into the discourse with the first sentence which was either a singleton entity or a set of two or three entities of the same kind. The second sentence either refers to

[1] In addition to the two experiments described below we also tested production and comprehension of pronouns and bridging constructions, and elicited noun phrases in narratives. These results are not included here.

that singleton entity (unique reference), targeting *the* as in (6), or one of the entities from the set (non-unique reference), targeting *a* as in (6). We had six unique items and five non-unique items (as one of our items failed).

(6) a. Unique reference in discourse
A dog and a pig were walking over a bridge. Something fell in the water and said "oink!" What was it? Target: the pig

b. Non-unique reference in discourse
A cat and two birds were sitting in a tree. Something flew out. What was it?
Target: a bird

4.3. Results production task

It was no surprise that we also found an overuse of *the* in the condition with a non-unique referent for which the target was *a*: the children produced 22% *a* and 46% *the*. Moreover, there was some article drop (22%) and some of the answers were irrelevant (10%). Four participants gave exclusively zero-article answers or almost all zero-articles; these children were all perfectly able to use articles in other contexts. We believe that their article drop in this task is due to a different interpretation of the game which triggered kind readings, for example, *Which kind of animal says "oink"?* Answer: *pig*.

Zooming in on the answers with articles and leaving out the answers with no articles, Figures 1a and 1b compare child and adult production in the unique and non-unique referent conditions.

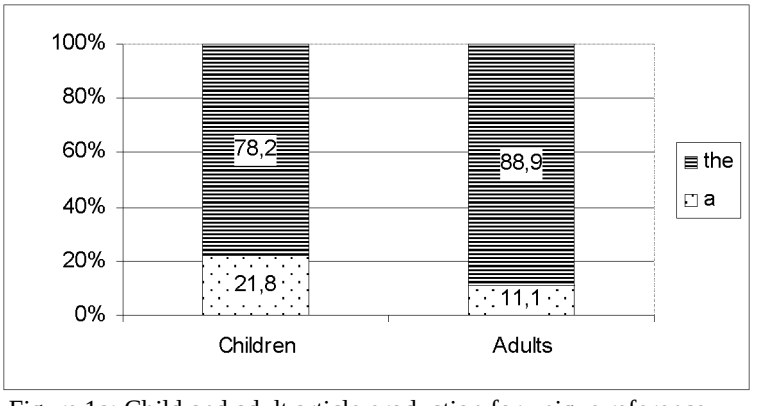


Figure 1a: Child and adult article production for unique reference

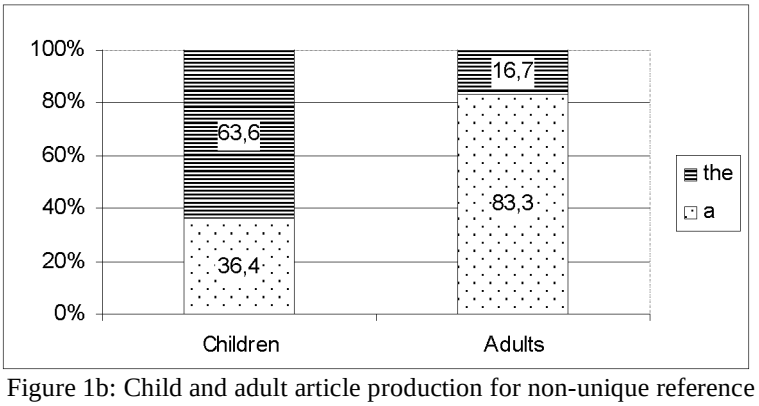


Figure 1b: Child and adult article production for non-unique reference

In the condition with a unique discourse referent—where the target answer was *the*—the children correctly used *the* 78% and used *a* 22% of the time. After taking out the four article-droppers, most children (n=15) correctly produced *the* for all six items or all except one item. The non-target use of *a* mostly came from three children who exclusively produced *a*. Another three children produced a mix

of *a* and *the*. The adults produced 89% *the*; there was one outlier, who produced the unexpected *a*-answers (11%).

In the condition with a non-unique referent adults produced the target answer *a* 83%; there was a slight overuse of *the* 17% (half of the adults produced *the* once or twice our of 5 times). The children showed a *the*-overuse of 64%. Looking at the individual answer patterns we found that *the*-overuse came in two kinds. Eight children consistently produced *the* in all five items or all except one. Nine children gave mixed answers, varying *a* and *the*. In addition four children consistently produced *a*. This included the same three children who also gave *a* in the unique referent condition, revealing an *a*-strategy across both conditions and no target discrimination of *a* and *the*. The fourth child was completely on target in both conditions: consistently *a* in the non-unique condition and *the* in the unique condition. The subject analyses are illustrated in Figures 2a and 2b.

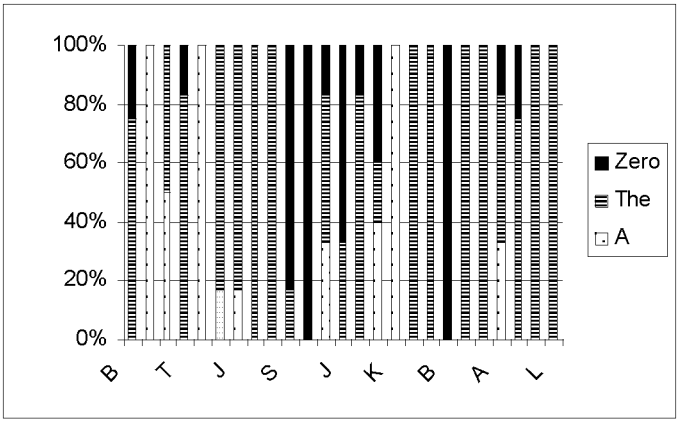


Figure 2a: Subject analysis article production for unique reference

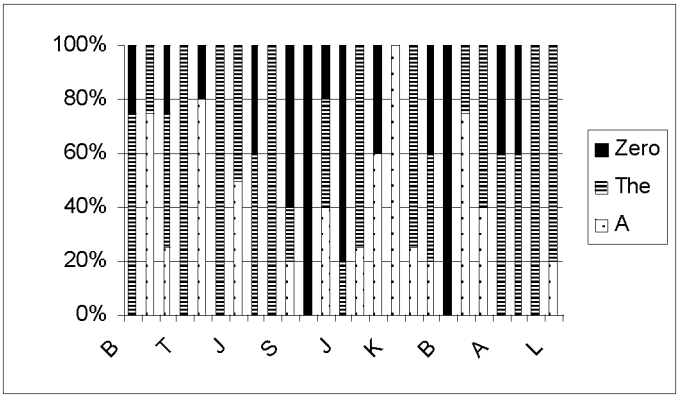


Figure 2b: Subject analysis article production for non-unique reference

Summarizing the production results (and leaving aside the few children who consistently did not produce any articles), most children were on target for determined referents and correctly produce *the* for a referent introduced in the previous discourse. However, all children except one overused *the* for non-determined referents; some of them produced *the* all the time, while others optionally gave *a* and *the*. Only one child out of a total of twenty-five preschoolers was fully on target in both conditions.

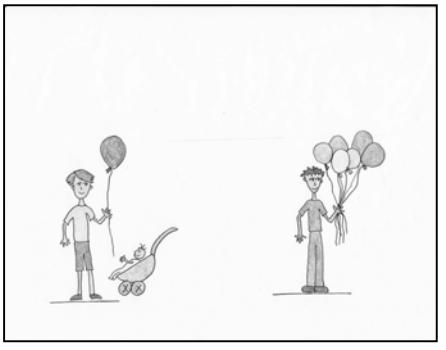
4.4. Comprehension task

Comprehension was tested with a truth-value judgment task using picture sequences. Participants were shown the first picture and given a description which singled out a certain discourse entity as the topic. In the next picture something happens to another, non-mentioned referent, Figure 3. We asked yes/no-questions about the second picture, varying definite and indefinite NPs, (8). The answers reveal whether participants allowed the NP to refer to the new referent (if they said yes) or to the mentioned

referent (if they said no). The target was “no” for *the*, since *the* must refer to a determined referent, and the new referent was not introduced before, and “yes” for *a*, because *a* introduces new referents.

(7) The dad gave the baby a balloon.

a. Did a balloon fly away? Target: Yes
b. Did the balloon fly away? Target: No



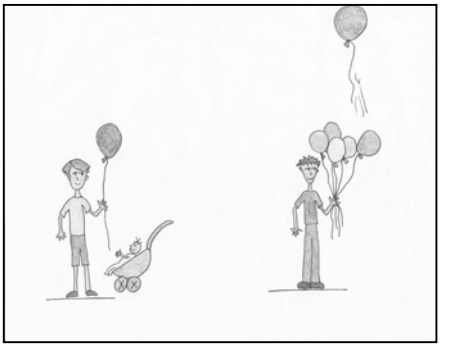
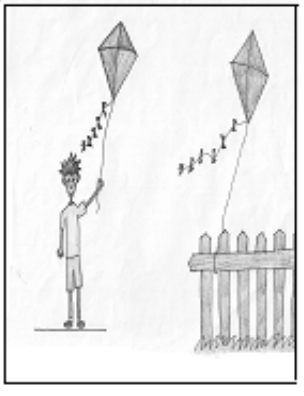


Figure 3: Test items about new referent in second picture

We also included control items where something happens to the singled-out referent, illustrated in Figure 4 and (8). Here a yes-answer indicates reference to the mentioned referent and a no-answer indicates reference to the new referent. The control items targeted “yes” for *the* because it is the correct article for the referent already introduced in the first picture. Since *a* typically introduces new referents, it is not the most felicitous article in this context, and so people might reject it. On the other hand, a sentence with *a* is not a false description of the picture, so people may simply accept it.

(8) A boy was flying his kite.

a. Did a kite fly away? Target: Yes?
b. Did the kite fly away? Target: Yes



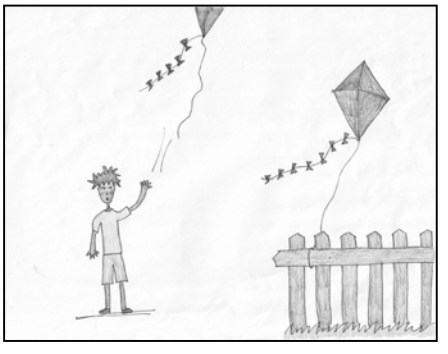


Figure 4: Control items about same referent in second picture

For each article there were four test items and two control items, making a total of 12 items.

4.5. Results comprehension task

In the test conditions which showed something happening to a new referent in the second picture the adults differentiated the two articles as expected: they accepted *a* for the new referent (92%), and mostly rejected *the* (only 21% yes). The children however did not make this distinction, equally accepting descriptions with *a* (92%) and *the* (80%). This is illustrated in Figure 5. In the control conditions children and adults accepted the description of the second picture with something happening to the previously singled-out referent for both articles.

Summarizing the comprehension results, both children and adults correctly accept *a* for a new referent in the discourse. However, children overaccept *the* for reference to a new referent in the discourse, while adults reject this. Putting production and comprehension together, we have found that children's *the*-overuse for non-determined referents in the production task is mirrored in their overacceptance of *the* for non-determined referents in the comprehension task.

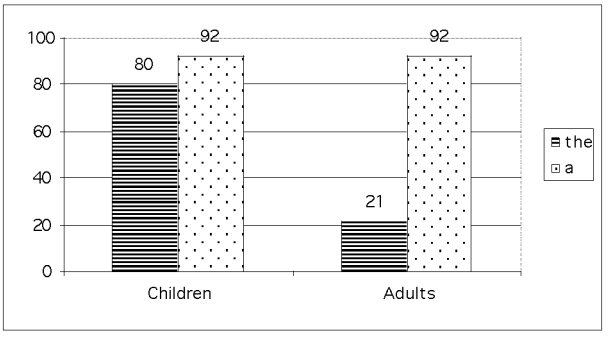


Figure 5: Acceptance of new referents for *the* versus *a* by children and adults

5. Discussion

We set out to see if deviances in article production are accompanied by deviances in comprehension. We replicated the well-known finding from the literature that children produce definites in contexts where an indefinite is required, specifically, where the referent is not uniquely determined. In our production study the referent was introduced in the previous discourse as one of a group, none of which was singled out, and so an indefinite would be appropriate. Our comprehension results indeed revealed a similar deviance from the target system as the production results. Children accepted definites for a non-determined referent, specifically, a referent that was present in the visual context but that had not been mentioned in the verbal context. Overuse of *the* for non-determined reference in production was thus coupled with overacceptance of *the* for non-determined reference in comprehension. We also note that children appropriately used definites in contexts where they were called for, specifically, when a unique referent was introduced in the previous discourse. Similarly they accepted definites for uniquely determined referents. So it seems that children associate two meanings with definite NPs: <*the*, determined reference> and <*the*, non-determined reference>. There was no asymmetry between comprehension and production of definites.

As for indefinite NPs, the children only produced an indefinite in appropriate contexts, specifically, when the referent was non-determined; they never produced an indefinite for determined reference. This shows that they distinguish the two articles, albeit not in a target-like way. In production indefinites are exclusively associated with non-determined reference.

In comprehension the children accepted both determined and non-determined referents for indefinites. But so did the adults. This was unexpected given that the target grammar associates indefinites with non-determined reference (see section 3). However, our method of asking yes/no-questions to probe truth value judgments was probably not sensitive enough to establish a difference. When a referent is determined because it is mentioned and singled out in the preceding discourse, a description of that same referent using an indefinite NP is at most infelicitous, because *the* is a better form to use here according to the Givenness Hierarchy, but it is not a false description, and so all participants accepted the combination of *a* with determined reference.

On the hypothesis that children lack bi-directional reasoning our bi-directional OT analysis in section 3 predicted target behavior in production, contrary to fact. Moreover, our analysis predicted that children would be on target for the comprehension, again contrary to fact. We furthermore predicted that children, but not adults, would allow two meanings for indefinites. This is also contrary to fact, given that adults accepted both meanings. The hypothesis of no bi-directional reasoning with the analysis with the two constraints presented above does not explain our results. Something is missing.

Our findings in the production task are in line with previous findings of *the*-overuse (Karmiloff-Smith, 1979; Maratsos, 1974; Schaeffer and Matthewson, 2005; Schafer and de Villiers, 2000). Our

findings in the comprehension task are similar to Van Hout et al.'s (2007) results with Dutch preschoolers who also accepted definites for non-determined reference. Van Hout el al. explained the latter result with the visual support: the referent that had not been mentioned in previous discourse is present in the picture, and so its reference is visually determined. So for children a definite NP qualifies as a proper description because the visual context suffices to establish its reference, whereas adults know that a definite can only be used to refer to a discourse entity that has been introduced verbally.

This explanation, even though it is straightforward and appealing and indeed explains the present comprehension results, does not explain our own and previous production results where children produced *the* in a purely verbal task without visual support. There is another crucial element that needs to enter into the explanation, and that is the finding that children overuse *the* only when the referent has been established as a non-unique referent in the verbal discourse (as one of a group), but not when the referent is a completely new element in the discourse context, see (2) from Schaeffer and Matthewson and (3) from Schafer and de Villiers. The fact that children produce an indefinite in these contexts shows that they correctly know that *the* cannot be used to introduce a completely new referent into the discourse.

The missing link in the explanation seems to relate to the discourse notion of coherence: you cannot start talking about a new referent out of the blue, because that would violate the coherence of an ongoing story in which there is an established discourse topic. Beaver (2004) formulated this as an OT constraint COHERE: "Don't change the discourse topic". If COHERE were the only force in a speaker's story, one would never switch to another discourse topic and could not talk about a new referent. Obviously COHERE must sometimes be violated in order to introduce new referents, else stories would be infinitely boring. Yet, COHERE requires a speaker to stick to a discourse topic and thus forces her to use the best NP form for topics, which under the Givenness Hierarchy is a pronoun or a definite NP. Suppose children are "overly coherent", in other words, suppose COHERE is ranked above DETERMINED REFERENCE. Then children will produce a pronoun or definite NP in order to satisfy COHERE, even if this violates DETERMINED REFERENCE when the referent is non-determined. This accounts for *the*-overuse in those cases where the non-determined referent is a member of a set of referents that has been introduced in the discourse and functions as the present discourse topic, as in our production paradigm in (6).[2] In the production paradigms in (2)and (3), on the other hand, the intended referent is not (part of) the discourse topic, and so COHERE does not apply. Adding COHERE to our production Tableau 2, this effect is illustrated in Tableau 6.

Meaning input Non-determined reference	COHERE	DET REF	AVOID INDEFINITES
a	*!		*
☞ *the*		*	

Tableau 6: The optimal form expressing non-determined reference for a member of discourse topic set

In our review of the literature we focused on the finding that *the*-overuse is an optional phenomenon in the sense that children produce both *the* and *a* for non-determined reference, which is also what we found in our data. In fact, the subject analysis showed that half of the children produced definites across the board and the other half produced both articles. Our explanation in Tableau 6 does not cover this latter result; with COHERE outranking DETERMINED REFERENCE we predict massive overuse of *the* and no use of *a* anymore. One way to achieve the optionality of *the* and *a* is to have these two constraints unranked, which gives as the effect that both forms are winners. And so the prediction is that both are used more or less at chance.

The difference between children and adults involves the ranking of COHERE and DETERMINED REFERENCE. For adults COHERE ranks below DETERMINED REFERENCE, so that the latter effect takes

[2]In our complete test battery we also included a condition with pronouns with a completely similar paradigm to the comprehension task with definite NPs presented here. One item read: *See the girl with the book? Is she drinking juice?* where the picture showed a different girl drinking juice, and the target answer was no. We found 40% acceptance in this condition, which is lower than the 80% finding with definite NPs, but still much too high. Our explanation with COHERE ranked highest also explains to this result.

priority and definites really must have determined reference. For children the two constraints are ranked in the opposite order or alternatively they are unranked. Either way, children need to acquire their appropriate ranking.

6. Conclusion

We tested twenty five English preschoolers and a small group of adult controls on their use and interpretation of definite and indefinite NPs using parallel production and comprehension tasks. Our results add comprehension data to the existing production literature. Our results indicate that overuse of *the* for non-determined reference has its mirror image in comprehension: children did not restrict their interpretation of definites to determined reference, but also accepted new referents. Our explanation for *the*-overuse makes use of the notion of coherence and posits that coherence overrules the semantics of *the*, so that *the* can even be used for a non-unique referent as long as it is part of the discourse topic. Our explanation of the overly liberal interpretation of *the* relies on the effect of the visual context which exactly determines the referents because they are present in the picture, even if they have not been mentioned in the discourse. This leads to a different application of the semantic rule for definites than adults have; for adults the visual context does not suffice to establish an antecedent for a definite NP.

Our comprehension task turned out to be insensitive for adults to the difference between definites and indefinites in the *a*-comprehension conditions. We think in hindsight that a truth value judgment task like ours cannot fulfill this task and so we want to find a way that reveals the typical form-meaning mappings of definite and indefinite articles, at least for adults: <*the*, determined reference> and <*a*, non-determined reference>. A different method in which participants need to choose which interpretation is best for *a* may yield the target pattern of <*a*, non-determined reference>. Harrigan (in preparation) designed a picture selection task in which participants must "fix" a picture book page to an accompanying story by picking up a referent in the picture and sticking it in the right spot. Picking up a referent involves choosing among a set of possible referents, either one that has been mentioned just before or one that has not been referred to yet. A definite or indefinite prompt should lead to different choices: for a definite the mentioned referent is best, while for an indefinite the non-mentioned referent is best. Preliminary results with ten adult participants are very promising; they make the intended distinctions in this task and associate *a* with a new referent by 90%. Children (n=6 so far) in contrast choose either referent more or less at chance. This pattern is predicted by our bi-directional analysis. On this account adults, but not children, block the mapping between *a* and determined reference, because *the* would have been the better form to express this meaning, and—so they reason—the speaker must have intended a non-determined referent. On the hypothesis that children only have unidirectional reasoning, they do not block any interpretations and allow both non-determined as well as determined reference for indefinites.

We investigated whether production problems with definite and indefinite NPs co-occur with comprehension problems, and we found that they do. However, we have not been able so far to find one unified explanation to explain both effects. More research into the role of coherence in children's discourse and the interaction with semantic and other discourse structure rules will hopefully bring the effects of overuse and overacceptance together as two sides of the same coin.

References

Beaver, David (2004) "The optimization of discourse anaphora," *Linguistics and Philosophy* 27, 3-56.

Burzio, Luigi (1998) "Anaphora and soft constraints," in P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky (Eds) *Is the best good enough? Optimality and competition in syntax*. Cambridge, MA: MIT Press.

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

Farkas, Donca (2002) "Varieties of indefinites", in *Proceedings of SALT 12*.

Gundel, Jeannette, Nancy Hedberg and Ron Zacharski (1993) "Cognitive status and the form of referring expressions in discourse," *Language* 69, 274-307.

Harrigan, Kate (2009) *Children's production and comprehension of definite and indefinite noun phrases in English: An Optimality Theory analysis*. Unpublished Honors Thesis, Smith College.

Heim, Irene (1982) *The Semantics of Definite and Indefinite NPs*. Ph.D. Dissertation, University of Massachusetts. Amherst: GLSA.

Hendriks, Petra and Jennifer Spenader (2005/6) "When production precedes comprehension: An pptimization approach to the acquisition of pronouns," *Language Acquisition* 13, 319-348.

Hendriks, Petra (2008) "A unified explanation of production-comprehension asymmetries," in A. Gavarró and M. João Freitas (Eds.) *Proceedings of GALA 2007*. Newcastle upon Tyne: Cambridge Scholars Publishing, pp. 240-251.

de Hoop, Helen and Irene Krämer (2005/6) "Children's optimal interpretations of indefinite subjects and objects," *Language Acquisition* 13, 103-123.

van Hout, Angeliek (2007) "Optimal and Non-Optimal Interpretations in the Acquisition of Dutch Past Tenses," in A. Belikova, L. Meroni and M. Umeda (Eds) *Proceedings of the 2nd GALANA*. Somerville: Cascadilla Proceedings Project, pp 159-170.

van Hout, Angeliek, Lianne de Ree, and Marlies de Ree (2008) "Dutch children's interpretation of definites and indefinites in discourse," in A. Gavarró and M. João Freitas (Eds.) *Proceedings of GALA 2007*. Newcastle upon Tyne: Cambridge Scholars Publishing, pp 279-289.

Kamp, Hans (1981) "A theory of truth and discourse representation," in J. Groenendijk, T. Jansen and M. Stokhof (Eds.), *Formal Methods in the Study of Language*. Amsterdam: Mathematical Centre tracts 135.

Karmiloff-Smith, Annette (1979) *A Functional Approach to Child Language: A study of determiners and reference*. Cambridge: Cambridge University Press.

Maratsos, Michael (1974) "Preschool children's use of definite and indefinite articles," *Child Development*, 45(2), 446-455.

Schafer, Robin and Jill de Villiers (2000) "Imagining articles: What *a* and *the* can tell us about the emergence of DP," in C. Howell, S. Fish, and T. Keith-Lucas (Eds.) *Proceedings of the Boston University Language Development conference*, Somerville: Cascadilla Press: 609-620.

Schaeffer, Jeannette, and Lisa Matthewson (2005) "Grammar and Pragmatics in the Acquisition of Article Systems," *Natural Language and Linguistic Theory* 23(1): 53-101.

de Swart, Henriëtte (in press) "Nominals with and without an article: Distribution, interpretation and variation," in P. Hendriks, H. de Hoop, I. Krämer, H. de Swart, and J. Zwarts *Conflicts in Interpretation*. London: Equinox Publishing.

Wexler, Ken (2003) "Cues don't Explain Learning: Maximal Trouble in the Determiner System," in T. Gibson and N. Pearlmutter (Eds.) *Proceedings of the 2003 CUNY Sentence Processing Conference*.

Wubs, Ellis, Petra Hendriks, John Hoeks and Charlotte Koster (this volume) "Tell me a story! Children's capacity for topic shift, " in *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 3)* Somerville: Cascadilla Press.

Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 2008)

edited by Jean Crawford,
Koichi Otaki, and Masahiko Takahashi

Cascadilla Proceedings Project Somerville, MA 2009

ISBN 978-1-57473-436-2 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

van Hout, Angeliek, Kaitlyn Harrigan, and Jill de Villiers. 2009. Comprehension and Production of Definite and Indefinite Noun Phrases in English Preschoolers. In *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 2008)*, ed. Jean Crawford, Koichi Otaki, and Masahiko Takahashi, 76-87. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2308.