Pronoun co-referencing errors: Challenges for generativist and usage-based accounts

DANIELLE MATTHEWS,* ELENA LIEVEN, ANNA THEAKSTON AND MICHAEL TOMASELLO

Abstract

This study tests accounts of co-reference errors whereby children allow ''Mama Bear'' and ''her'' to co-refer in sentences like ''Mama Bear is washing her'' (Chien and Wexler 1990). 63 children aged 4;6, 5;6 and 6;6 participated in a truth-value judgment task augmented with a sentence production component. There were three major finding: 1) contrary to predictions of most generativist accounts, children accepted co-reference even in cases of bound anaphora e.g., ''Every girl is washing her'' 2) contrary to Thornton and Wexler (1999), errors did not appear to occur because children understood referring expressions to be denoting the same person in different guises 3) contrary to usage-based accounts, errors were less likely in sentences that contained lower as opposed to higher frequency verbs. Error rates also differed significantly according to pronoun type (''him'', ''her'', ''them''). These challenging results are discussed in terms of possible processing explanations.

Keywords: *language acquisition; anaphora; binding; pronouns; truth-value judgment task; frequency; quantifiers.*

1. Introduction

On all theoretical accounts of language acquisition, we would expect English-speaking children to have a fairly robust understanding of sentences like ''*Mama bear is washing her*'' by the age of three. It is therefore

* Correspondence address: Danielle Matthews, Max Planck Child Study Centre, School of Psychological Sciences, University of Manchester, M13 9PL. UK. E-mail: ⟨danielle.matthews@manchester.ac.uk⟩. Acknowledgements: We thank Siu-lin Rawlinson, who created the stimuli illustrations, Anna Roby for help with data collection, Jarrad Lum for advice on accuracy statistics and Oughtrington Primary School for participating.

Cognitive Linguistics 20–3 (2009), 599–626
DOI 10.1515/COGL.2009.026
0936–5907/09/0020–0599

surprising that such a sentence can cause children as old as seven serious comprehension problems. Yet years of language acquisition research has consistently revealed a comprehension error whereby children accept this sentence as a description of a picture of a mother bear washing her*self*. The need to explain this pervasive and persistent comprehension difficulty has led to several accounts of how children develop the ability to assign referents to anaphoric pronouns. All of these accounts have taken a generativist approach to language acquisition and have assumed innate knowledge of syntactic binding constraints. This paper presents an alternative, constructivist proposal whereby children learn anaphora constraints in terms of noun phrase accessibility, contrastive forms and conventionalized sentential contexts. We test predictions from both generativist and constructivist accounts and attempt to investigate whether the latter can "scale up" to explain complex areas of syntax that have typically only been investigated within the generativist framework.

1.1. *Generativist approaches to anaphora: Priniciple B and Rule I*

Co-referencing errors of the type described above originally came to light in studies that investigated children's understanding of syntactic binding principles such as Principle B in the Government and Binding framework (Chomsky 1981), which can be stated as follows[1]:

Principle B: A pronoun must be free (not co-indexed with a c-commanding NP) in its governing category.

Chien and Wexler (1990) noted that Principle B only strictly rules out co-reference in cases of bound variable uses of pronouns, as in the sentence "*Every bear is washing her*". In such cases, the pronoun cannot be co-referential with the quantified antecedent, since quantifiers do not refer. Instead of co-reference, then, the relationship between the pronoun and the antecedent is one of binding. In this case Principle B directly rules out the incorrect interpretation without reference to the pragmatic context (c.f. Avrutin 1999 for a discussion of the difference between binding and co-reference). In other sentences, such as "*Mama bear is washing her*", the pronoun is not a bound variable and Principle B does not dic-

1. There has been much debate about the formulation of binding principles (Jackendoff 2002; Lust et al. 1994; Reinhart and Reuland 1993; Runner et al. 2003) and the given definition has changed considerably in the Minimalist framework (Baauw and Cuetos 2003; Chomsky 1995;). The details of this debate will not affect the thrust of the arguments presented here.

tate whether the two expressions co-refer. Instead, co-reference is decided by applying pragmatic principles, such as Reinhart's Rule I (Grodzinsky and Reinhart 1993; Reinhart 1983):

Reinhart's Rule I (Intrasentential Co-reference): NP A cannot co-refer with NP B if replacing A with C, C a variable A-bound by B, yields an indistinguishable interpretation.

Essentially, Rule I (referred to as Principle P by Chien and Wexler 1990) means that if by saying "*Mama bear is washing her*" we mean *Mama bear* and "*her*" to refer to the same person, then we should replace the pronoun "*her*" with a bound variable, "*herself*", as long as this doesn't change the intended meaning of the sentence. For some pragmatically peculiar cases, such as, "*That must be John. At least he looks like him*", the meaning of the sentence would be changed if such a replacement occurred. In this example, John is referred to under two different *guises* (on the one hand "John who we know" and on the other "the person we are looking at") and the use of the pronoun "*him*" marks this difference in senses (see, chapter 2 of Avrutin 1999 for further introduction and Grodzinsky in prep for further discussion of distinctness of interpretation). Application of Rule I thus requires sensitivity to context (see Foster-Cohen 1994 for further explanation of Rule I applied to child language).

Researchers adopting a generativist approach to acquisition (e.g., Chien and Wexler 1990; Grodzinsky and Reinhart 1993; Thornton and Wexler 1999) argue that children know Principle B (or a reformulated equivalent) from the start and only have problems with the pragmatic conditions, such as Rule I, that dictate when one should or should not assume co-reference in cases where binding theory is not decisive. The acid test of this hypothesis has been to look at children's comprehension of what we will refer to as "quantifier sentences", such as 2 below, where principle B applies free of any pragmatic considerations. Following this logic, Chien and Wexler (1990) employed a truth-value judgment task, to test children on pictures that either matched or mismatched sentences such as 1 and 2.

1. Mama Bear is touching her.
2. Every bear is touching her.

They found that, once children demonstrated knowledge of quantifiers with proper names at age 5 (i.e., they responded correctly to control sentences such as "*Every bear is touching Goldilocks*"), they successfully rejected cases of illicit co-reference for sentences such as 2 (84 percent

correct rejections), whilst they remained less capable of doing so for sentences such as 1 (60 percent correct rejections). Chien and Wexler (1990) thus concluded that the children demonstrated knowledge of Principle B. Whilst many aspects of Chien and Wexler's account have been challenged (Avrutin 1994; Grimshaw and Rosen 1990; Grodzinsky and Reinhart 1993; McDaniel and Maxfield 1992; McKee 1992;), the general assumption that children will reject co-reference in sentences such as 2 due to knowledge of principle B has generally been accepted (see, for example, Baauw and Cuetos 2003; Hestvik and Philip 1999/2000; Thornton and Wexler 1999).

In contrast to the consensus on Principle B, there has been much debate as to the appropriate explanation for the difficulties that lead to high error rates with sentence 1 above. Grodzinsky and Reinhart (1993) attributed these difficulties to children's limited memory and processing capacity: applying Rule I requires holding many propositions in memory to compare them, a process which is liable to break down. In a more recent account, Thornton and Wexler (1999) propose an *extended guise creation* hypothesis to explain errors. The central idea here is that children need to learn the contexts in which a speaker can intend a local co-reference interpretation. As mentioned above, situations in which adults would allow local co-reference for pronouns are characterized as involving the creation of two separate guises for the referent (in the case of "*That must be John. At least he looks like him*", John and the man being viewed are seen in different guises even though they may be the same person).

Thornton and Wexler attribute children's co-referencing errors to their over-assignment of different guises in contexts where adults would require extra contextual support. Children are thus proposed to accept sentences such as "*Mama bear is washing her*" to describe a bear washing herself because, unlike adults, children find it unexpected that Mama Bear should wash herself and thus assume the speaker intended to draw attention to this event as the surprising climax of the story. They therefore argue that children actually represent Mama Bear with two guises in mind such that the sentence would read as "*Mama bear washed the individual that washed somebody*". Why children should be more prone to doing this than adults is not clear. Nonetheless, children are proposed to recover from extended guise creation by observing that every time co-reference is used in adult speech it is accompanied by a specific context in which an alternative guise is intended. Thornton and Wexler (1999: 105) state, "the problem of learnability is circumvented by the accrual of real-world knowledge in combination with innate pragmatic principles that govern the assignment of interpretation to sentences in conversational contexts."

Importantly, extended guise creation is not permitted for bound variables, which are not referential and therefore cannot bear guises. So, children should not allow extended guise creation with test sentences like "*Every bear washes her*" because such an interpretation would require violating Principle B.

One problem with Thornton and Wexler's (1999) approach is that it does not predict recent findings which show that children's knowledge of co-reference is significantly affected by additional syntactic and lexical factors. For example, Philip and Coopmans (1996) found that Dutch children of 7 years of age made significantly more co-reference errors when:

a) the pronoun and the antecedent were not co-arguments (e.g., "*The girl sees her blow bubbles*").
b) sentences contained the third person, feminine pronoun "*haar*" (her) rather than third person, masculine pronoun "*hem*" (him).
c) sentences contained highly reflexive verbs such as *wash* as compared to more transitive verbs such as *point at*.

Although there have been attempts to explain some of these lexical effects (see for example Hestvik and Philip, 1999/2000) there is to date no theory that directly predicts them all. These findings are, however, consistent with usage-based accounts of language acquisition. On such accounts, significant and varied lexical effects on children's ability to deploy a linguistic principle are taken to provide evidence that this principle is being gradually learnt (Tomasello 2003). The question is whether such a theory could extend to explaining children's errors in pronoun co-referencing. This issue has not so far been addressed in the child language literature although a cognitive grammar account of anaphora has been put forward. We briefly review this account before considering how it might explain language acquisition and what predictions it would make with respect to children's errors.

1.2. *A Cognitive Grammar account of anaphora (van Hoek 1997)*

The most complete alternative to generativist accounts of anaphora is provided by van Hoek (1997). Working within the Cognitive Grammar framework, van Hoek argues that the facts explained by the structural notion of c-command can be given a conceptual-semantic grounding in terms of interactions between different nominal types and the contexts in which they are embedded (see also Harris and Bates 2002). This account is based on the accessibility theory notion (Ariel 1990) that nominal expressions form a continuum that reflects the relative information status of a referent in a given context. Full nominals form one end of

this continuum in that they generally introduce new information and are most appropriate when the referent cannot be recovered from context. Pronouns, on the other hand, profile a referent that is recoverable from the context—i.e., more accessible. So, although they might refer to the same entity, full nouns and pronouns convey different meanings concerning the relationship between the referent and the (discourse or extra-linguistic) context. It is the need to choose an appropriate referring expression in terms of accessibility for the given context that underwrites van Hoek's constraints on anaphora. Unacceptable co-reference occurs when, for example, a full noun phrase, assumed to be low in accessibility, is embedded in a context in which the referent is in fact highly accessible. This would be the case if co-reference were assumed in sentence 3 (see van Hoek 2003: 176).

3. He saw a skunk near Ralph.

In sentence 3, the initial pronoun indicates that we already know to whom "he" refers. The subsequent use of a full name "Ralph" indicates that we must now be talking about a less accessible referent, which must be a different person - else there is an anomaly in that the same person is simultaneously accessible and not accessible.

In order to precisely determine when co-reference is and is not allowed, the account of noun phrase accessibility must be combined with an adequate notion of context as it is the interaction between the two that defines when co-reference is anomalous. To achieve this, van Hoek introduces a *reference point model*, which presents sentential conceptual structure in terms of reference points and dominions. Reference points are similar to topics in that they are salient entities that the conceptualizer makes contact with and that form the background from which subsequent entities are understood. Dominions are conceptual structures that are construed in relation to a reference point. Within the clause, the subject is the primary reference point and thus all other nominals fall in its dominion. The direct object is the second reference point and all subsequent nominals fall within its dominion. Nesting of reference points within dominions continues down a hierarchy from one nominal to the next, from main clauses to embedded clauses. It is in terms of this embedding that van Hoek defines the following anaphora constraints (van Hoek 1997: 57):

i) A full nominal cannot appear in the dominion of a reference point that it corresponds to.
ii) The antecedent for a pronoun must be sufficiently salient within the context in which the pronoun appears that it can plausibly be construed as a reference point with the pronoun in its dominion.

These constraints constitute knowledge of (cognitive) grammar to the extent that specific reference point/dominion structures become entrenched in the language in the form of conventionalized grammatical structures. These conventionalized structures are referred to as the *complement chain*, which is essentially a grammatical relations hierarchy determined by semantic prominence, with the subject at the top of the head/complement structure and complements successively further down.

Note that, so far, van Hoek's account can explain non-co-reference in Principle C type sentences, such as "*He adores Joe*". However it does not explicitly handle the impossibility of co-reference in sentences such as "*Joe adores him*", which we are currently concerned with. An explanation for this is provided in van Hoek's account of reflexivity (van Hoek 1997: 174–178). The critical concept here is *point of view*. Whereas a pronoun in the sentence *"Joe adores him"* indicates that the speaker considers the referent of *"him"* accessible (part of the interlocutors' shared background knowledge), a reflexive indicates not only that the referent is accessible but also that is perceived from the point of view of some participant in the scene. Van Hoek gives the following illustrative example (taken from Cantrall 1974):

4. I can understand a father wanting his daughter to be like **himself** but I can't understand that ugly brute wanting his daughter to be like **him**.

The use of the reflexive *himself* invokes the concept of the father from his own perspective, the pronoun *him* invokes a concept of the father as viewed objectively, by somebody else. Van Hoek explains this distinction further in terms of Langacker's stage model (Langacker 1985). As far as anaphora constraints are concerned, we can presume that co-reference in sentences such as "*Joe adores him*" is not so much ruled out as pre-empted by sentences like "*Joe adores himself*".

This pre-emption hypothesis is in many respects very similar to Reinhart's Rule I, which essentially states that pronouns should not refer back to the subject of the same clause except in cases where replacing the pronoun with a reflexive would result in a change of meaning. The main difference is that on a Cognitive Grammar account, these constraints are assumed to be learnable. Indeed, combining van Hoek's approach to anaphora with a constructivist account of language acquisition, we suggest that the following three abilities are central to mastering the use of anaphoric pronouns and could replace the need for innate syntactic constraints.

1.2.1. *Noun phrase accessibility.* Appropriate use of anaphoric pronouns requires an appreciation of the information status (accessibility,

givenness) associated with both pronouns and other referring expressions. An example hierarchy that summarises what might be learnt is that put forward by Gundel and colleagues (Gundel et al. 1993; Gundel et al. 2001). That children need to learn such properties of referring expressions is assumed to be uncontroversial, even if how this is achieved is currently a matter of debate (Matthews et al 2006; Matthews et al. 2007).

1.2.2. *Discourse and sentential context: The complement chain.* To make co-reference inferences one needs to understand the pronoun's relation to the surrounding discourse. Within-sentence reference resolution requires an appreciation of the complement chain or similar grammatical hierarchy such as that put forward by Keenan and Comrie (Keenan and Comrie 1977):

Subject > Direct Object > Indirect Object > Oblique > Genitive > Object of comparison

The assumption here is that any such hierarchy would be a conventionally established organization of reference points. Thus it would emerge from a structured inventory of constructions that map phonological form to communicative function (Croft 2001; Goldberg 1995; Goldberg 2006; Langacker 2000).

1.2.3. *Point of view and the pronoun/reflexive contrast.* Mastering the contexts in which anaphoric pronouns can be used requires an appreciation of contrastive forms, in particular reflexives. The pronoun/reflexive contrast is understood in terms of perspective, such that reflexives are used to convey information about a referent as it is seen by that referent, whereas pronouns refer to an accessible referent viewed from a more objective perspective.

The above three linguistic abilities are by no means proposed to be an exhaustive list of those required to process anaphoric pronouns. Indeed, numerous other factors affect anaphor resolution, for example, the morphological properties of the pronoun (gender, number, animacy, case), lexical properties of the verb (causal bias verbs, perceiver/action verbs, highly reflexive/transitive verbs), contrastive stress, real-world probabilities and discourse level factors (global topic, local focus and so on). We single out these three abilities simply in order to provide an alternative account of how children might learn structural constraints on anaphora resolution that have traditionally been assumed to require innate syntactic knowledge.

To the extent that anaphora resolution relies on knowledge of the complement chain as emergent from an organized inventory of constructions, a usage-based account of language acquisition would predict that knowledge of anaphora would develop gradually. Until this knowledge becomes fully robust, children's ability in tasks testing anaphora constraints might vary according to the individual lexical items (verbs and pronouns) and constructions being tested. The more familiar children are with the specific lexical items involved, the easier and less error prone reference resolution should be. In contrast, on a generativist account of co-reference, one would expect children to have abstract knowledge of co-referencing constraints, be they syntactic or pragmatic, which would apply to all lexical items as soon as they had been correctly categorized.

1.3. *Testing generativist and usage based accounts*

The current study set out to test three predictions following from the accounts reviewed above: 1) children will not make co-referencing errors with cases of bound anaphora (Chien and Wexler 1990) 2) children make co-referencing errors because they create guises where adults would not (Thornton and Wexler 1999) 3) children will make fewer errors with sentences containing higher frequency lexical items.

To investigate the first of these predictions, we simply tested children on sentences such as "*Every boy is washing him*", which generativist accounts state should not be open to co-reference errors. The first aim of doing so was to replicate Chien and Wexler's (1990) findings and to test whether children's performance was statistically indistinguishable from ceiling performance. This is of particular concern because, although both generativist and cognitive grammar approaches treat bound anaphora as a special case (see Ch. 6 of van Hoek 1997), only generativist approaches to language acquisition predict that children will have knowledge of the relevant syntactic constraint (Principle B) from the outset of language development, thereby leaving no room for development. We also wanted to see if resistance to error reported in previous research might be due to a conflict between the plural concept of lots of boys and the singular marked pronoun. We thus predicted that children might be more liable to erroneously accept these sentences if the third person plural pronoun were used (e.g., "*Every boy is washing them*").

To test the second prediction, we presented children with a truth-value judgment (TVJ) task combined with a production task. Following Hestvik and Philip (1999/2000) the TVJ task took the form of a guessing game. Thus, a first experimenter showed each child a set of simple drawings (e.g., of a girl washing herself beside her mother) and a second

experimenter, who could not see the drawings, guessed what was happening in them (e.g., saying "*The girl is washing her*"). The child was asked to say whether E2 had guessed correctly or not and to give/repeat the correct answer to E1. If, in judging E2's guess, a child assumed the sentence required guise creation because it encoded surprise, we would expect them to repeat the sentence with the guise-creating pronoun to E1. If children are simply unsure as to the referential properties of pronouns they may accept the infelicitous pronouns in comprehension but nonetheless prefer a reflexive in production.

To test the third prediction, we tested sentences of the form "*The girl is washing her*" and varied them according to pronoun gender, pronoun number and verb frequency. We predicted that, because cognitive grammar accounts propose that anaphora resolution relies on knowledge of the complement chain as emergent from an organized inventory of constructions, children's ability in tasks testing anaphora constraints would vary according to the individual lexical items (verbs and pronouns) and constructions being tested, with higher frequency lexical items making sentences easier to judge. We thus predicted that children would make more accurate judgments about sentences with higher frequency verbs than lower frequency verbs. In addition, we expected that pronoun gender and number would affect children's judgments, although it is difficult to predict precisely which pronouns should be easiest for children to judge.[2] On the one hand, the third person plural pronoun *them* is more frequent than either *him* or *her* and so might be mastered earlier. On the other hand, Tanz (1976) found that children aged between 3 and 5 years have significant difficulties understanding the use of "*them*" in comparison to the singular pronoun "*it*" and mass nouns such a "*playdough*". Such difficulties might therefore make "*them*" more susceptible to error. Our manipulation of pronoun number and gender is therefore more for exploratory purposes.[3]

2. The frequencies of the pronouns in the speech directed to the twelve children in the Manchester corpus (Theakston et al. 2001) available on the CHILDES database (MacWhinney 2000) are as follows: *him* (2805), *her* (4129), *them* (5176). The high frequency of *them* as opposed to *him* and *her* is probably due to the fact that *them* can refer to animate and inanimate objects. The frequency count of *her* includes uses of the possessive pronoun. It is difficult to estimate the frequency of *her* as an accusative pronoun only and worth noting that the pronouns that any individual child hears most often are likely to vary greatly according to the gender of their siblings and other family members.
3. In a pilot study, we also manipulated the verbs according to reflexiveness (inherently reflexive: *wash*, *dry*, *hurt* vs. transitive: *hit*, *tickle*, *kiss*). We found no effect of verb group in this pilot study.

2. Method

2.1. *Participants*

63 normally developing, monolingual, English-speaking children participated in the study (30 boys, 33 girls). There were 21 four-year-olds (range 4;3–4;10, mean age 4;7), 21 five-year-olds (range 5;3–5;11, mean age 5;7) and 21 six-year-olds (range 6;3–6;9, mean age 6;6). A further four children were not included either because they i) did not complete the testing session or ii) made errors on more than one control filler question. The children were tested in a quiet room in their school library.

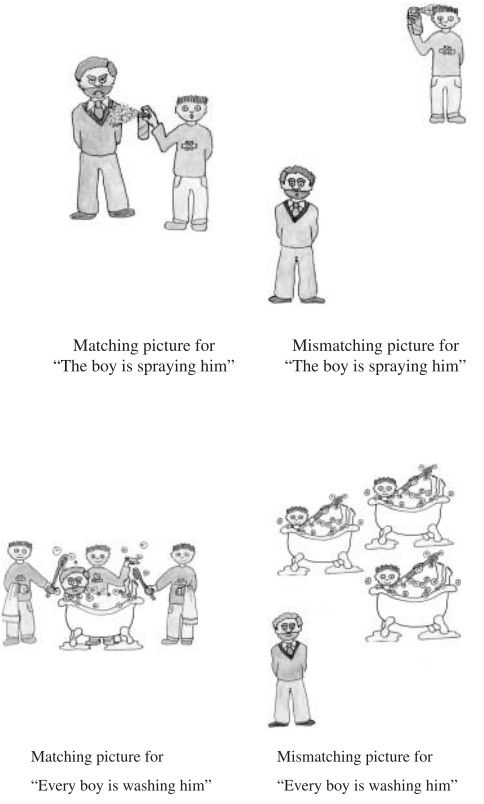

Figure 1. *Examples of matching and mismatching sentence-picture pairs*

2.2. *Materials and design*

Eighteen "non-quantifier" test sentences of the form "[*X is*/*The Xs are*] *VERBing* [*him*/*her*/*them*]" were generated by combining each of the six

test verbs (high frequency verbs: *hit*, *tickle*, *wash*, low frequency verbs: *pinch*, *spray*, *lasso*) with each of the three test pronouns (*her, him* and *them*). Verb frequencies are based on counts of the child directed speech of twelve English-speaking mothers from the Manchester Corpus available on CHILDES (MacWhinney 2000; Theakston et al. 2001). For each test sentence there was a picture that matched it and a picture that mismatched it (see Figure 1). The mismatching picture always represented the subject of the test sentence performing the action on itself and the other character standing still in the diagonally opposite corner. This ensured there was a plausible referent for the pronoun in these pictures. For these pictures, we systematically rotated the corner in which the subject of the test sentence was seen. Pictures always had two characters clearly of the same gender (either a mum and a girl or a dad and a boy). Pictures corresponding to test sentences with the third person plural pronoun, "*them*", always had two of each type of character in them (e.g., two dads and two boys for the sentences "*The boys are hitting them*").

Nine quantifier test sentences of the form "*Every boy is VERBING him/them*" were created. To keep the experiment short, and since we were interested in whether children would correctly reject mismatching sentence-picture pairs of this type, six of these sentences had a mismatching picture and three had a matching picture. The six sentences with mismatching pictures were generated by combining each of the three high frequency verbs (*hit*, *tickle*, *wash*) with each of two pronouns (*him* and *them*). In each case the picture depicted three male characters performing an action on themselves (e.g., three boys washing themselves) and in the opposite corner either one character or three characters standing still. In sentences with the pronoun "*him*" there was one character standing still, in sentences with the pronoun "*them*" there were three characters, such that a grammatical, non-matching antecedent was always available.

Finally, six control sentences were also included, three with "full name" sentences (e.g., "*The boy is hitting the dad*") and three with reflexive pronouns (e.g., "*The boy is hitting himself*"). Half of these control sentences mismatched their picture. This allowed us to check that children did not demonstrate a strong tendency to accept incorrect representations of sentences when no pronouns were present.

In total each child was presented with 51 sentence-picture pairs. The pictures were presented in three blocks. Each block contained simple test (T), quantifier test (Q) and control (C) sentences in the following fixed order: (TTQTTCTTQTTCTTQTT). This ensured that the control and quantifier sentences were spread evenly through the experiment. Sentences were pseudo-randomly assigned to blocks such that there were never more than three matching or mismatching pictures in a row and

matching and mismatching sentences for the same picture never appeared consecutively. The blocks were counterbalanced such that each block appeared first for a third of the children in each age group. All three blocks were tested in immediate succession. The full list of test sentences can be found in the appendix.

2.3. *Procedure*

Throughout the experiment, the child sat at a table next to the first experimenter, E1. The book of pictures was placed in front of the child and E1 with a screen behind it so that it could not be seen by the second experimenter, E2, who sat across the table. The screen also hid E2's script from the child. E1 explained to the child that E2 could not see the pictures but that E1 would give some clues and then the E2 would guess what was happening in the pictures. It was explained that sometimes E2 would guess right and sometimes she would guess wrong because she couldn't see, so she was only guessing. The child was asked if s/he could tell E2 if she had guessed right. Then s/he was asked if s/he could tell E1 what the right answer was so she could write it down.

Before beginning the test phase there was a short warm up phase with intransitive sentences. First E1 presented pictures of the four characters to the child and asked if s/he could point to each character in turn to check that each child understood the terms "*girl*", "*boy*", "*mum*" and "*dad*". If a child was unsure, E1 named each character and the two would practice naming until the child could confidently name every character. Next, the guessing game was introduced with four warm up pictures. These pictures consisted of one character either dancing or eating, with another character of the opposite sex standing still in the diagonally opposite corner. E1 gave clues about the pictures to E2, for example, "*There's a boy and a mum and the clue word is dancing*". Clues always consisted of the names of the two characters and the test verb in the present progressive. E2 then guessed what was happening in the picture. Half of E2's guesses were correct, the other half had the other character as the subject. Guesses were always in declarative forms preceded by "*Hmm*", "*I know…*" or "*Could it be*" (e.g., "*Hmmm … I know … The boy is dancing*."). E2 was blind to whether their scripted guesses were correct or not. Once the guess was made, the child was asked if it was right or not and was helped to respond correctly if necessary. E1 then asked "*Can you tell me the right answer (too) so that I can write it down?*". We added "*too*" in cases where the guess had been judged correct.

The test phase continued with the transitive test sentences in the same fashion as the warm up phase, apart from this the child was never helped

to guess and was always given positive feedback. On the rare occasions when children asked E2 to clarify her guess, E1 said that E2 was only allowed one guess.

Both experimenters transcribed the child's responses and an audio recording of the test was made. If children changed their minds, their final response only was used in coding. The children's responses were coded by the first author. 10 percent of the data were transcribed and coded by a third experimenter, who was blind to the hypotheses of the experiment. For the TVJ task, Cohen's Kappa was 0.92 (96% agreement). For the production task, Cohen's Kappa was 0.75 (99% agreement). All discrepancies in coding were easily resolved (almost all involved the coding of a handful of responses where children correctly produced reflexives but marked them incorrectly for number. We resolved to accept these responses as correct).

3. Results

We first present the results for test sentences containing the quantifier *every* to assess whether children are at ceiling when judging these sentences and to see whether judgments are affected by pronoun number. We then present results of the remaining test sentences to compare performance on comprehension and production (thereby investigating the guise creating hypothesis), and to test for lexical effects on judgments (thereby testing usage-based predictions). Last of all we consider the results of the control sentences and a qualitative analysis of individual response patterns.

3.1. *Sentences with the quantifier "every"*

This section tests the generativist prediction that children should be at ceiling when required to reject test sentences containing the quantifier *every*. It also tests our prediction that judgments will be affected by pronoun number. Table 1 presents the results of the TVJ task and the production task in terms of mean proportion of responses correct for matching and mismatching quantifier sentence-picture pairs that had either a third person singular or third person plural object (*Every boy is washing* ***him*** vs. *Every boy is washing* ***them***). For the TVJ task, correct responses are acceptances of matching sentences and rejections of mismatching sentences. For the production task correct responses are simply correct descriptions of the picture. To aid comparison, the last row of the table summarizes the results of the analysis of non-quantifier test sentences (Non-Q), averaging across verb frequency and pronoun conditions.

Table 1. *Mean proportion correct responses for the TVJ task and production task as a function of age, pronoun type and sentence-picture pair status ((mis)match)*

	TVJ task				Production task			
	Matching		Mismatching		Matching		Mismatching	
	Him	Them	Him	Them	Him	Them	Him	Them
4 years	0.95	0.90	0.37	0.54	0.95	1.00	0.65	0.75
5 years	0.90	0.81	0.60	0.63	0.95	0.93	0.83	0.89
6 years	0.86	0.79	0.51	0.56	0.95	1.00	0.78	0.90
Mean	0.90	0.83	0.49	0.58	0.95	0.98	0.75	0.85
Non-Q	0.83		0.39		1.0		0.83	

The most remarkable result is that children made a substantial number of errors when required to reject mismatching sentence-pictures pairs in the TVJ task, although they fare better than with non-quantifier sentences. Single sample t-tests revealed that, when required to reject mismatching quantifier sentences, children in all age groups were significantly below ceiling performance (Pronoun = him: four-year-olds: $t(20) = 7.684$, $p < 0.001$, five-year-olds: $t(20) = 3.998$, $p = 0.001$, six-year-olds: $t(20) = 5.413$, $p < 0.001$. Pronoun = them: four-year-olds: $t(20) = 6.501$, $p < 0.001$, five-year-olds: $t(20) = 5.043$, $p < 0.001$, six-year-olds: $t(20) = 7.260$, $p < 0.001$) and in fact did not differ significantly from chance performance (Pronoun = him: four-year-olds: $t(20) = -1.633$, $p = 0.118$, five-year-olds: $t(20) = 1.040$, $p = 0.311$, six-year-olds: $t(20) = .087$, $p = 0.931$. Pronoun = them: four-year-olds: $t(20) = 0.560$, $p = 0.581$, five-year-olds: $t(20) = 1.864$, $p = 0.077$, six-year-olds: $t(20) = 0.907$, $p = 0.375$). A 3 (age) × 2 (pronoun) ANOVA performed on the mean proportion of correct responses in the TVJ task for matching sentence-picture pairs revealed no significant effects and no interaction. An equivalent ANOVA for mismatching sentence-picture pairs showed that, contrary to our prediction, children were significantly more likely to correctly reject mismatching sentences with the pronoun *them* than with the pronoun *him* ($F(1,60) = 4.7$, $p = 0.033$, partial $\eta^2 = 0.074$). There was no effect of age and no significant interaction.

For the production task, a 3 (age) × 2 (pronoun) ANOVA on the proportion of correct comments revealed no significant effects and no interaction. However the equivalent ANOVA performed on the proportion of correct comments for mismatching sentence-picture pairs revealed a significant effect of pronoun such that children were better at providing correct comments after a sentence containing the pronoun *hem* than after a sentence containing the pronoun *him* ($F(1,60) = 7.3$, $p = 0.009$,

Table 2. *Mean proportion correct responses for the TVJ task as a function of age, pronoun type, verb frequency and sentence-picture pair status ((mis)match)*

	Matching sentence-picture (correct acceptance)						Mismatching sentence-picture (correct rejection)					
	High freq verbs			Low freq verbs			High freq verbs			Low freq verbs		
	Her	Him	Them	Her	Him	Them	Her	Him	Them	Her	Him	Them
4yrs	.97	.93	.71	.94	.92	.83	.37	.39	.22	.35	.37	.24
5yrs	.87	.83	.69	.92	.86	.83	.56	.43	.42	.49	.50	.44
6yrs	.83	.81	.68	.83	.78	.77	.46	.27	.34	.36	.47	.33

$\eta^2 = 0.108$). This result is not predicted by any of the reviewed accounts. There was no effect of age and no interaction.

3.2. *Non-quantifier sentences*

This section tests the extended guise creation hypothesis by comparing children's performance on the TVJ task and the production task. It also tests the prediction that children's responses will be subject to effects of lexical items. We present results for the main set of test sentences, which took the form "*The X is/are VERBing him/her/them*". Table 2 presents the results of the TVJ task in terms of the mean proportion of matching sentences that the children correctly accepted (*correct acceptances*) and the mean proportion of mismatching sentences that the children correctly rejected (*correct rejections*). Table 3 presents the results of the production task in terms of the mean proportion of responses that were accurate after matching sentences (*comments match*) and after mismatching sentences (*comments mismatch*)[4].

With respect to the predictions of the guise-creation hypothesis, comparison of tables 2 and 3 shows that children performed considerably better on the production task than on the TVJ task. Thus, individuals would frequently judge an incorrect, mismatching sentence as acceptable but then spontaneously offer the correct version of the sentence in the production task. A t-test comparing mean scores on the TVJ task and the production task for mismatching sentence-picture pairs confirmed this difference was significant ($t(62) = 9.72$, $p < 0.001$).

4. All analyses of proportional data presented here are based on untransformed proportions. All analyses were re-run with arc sine transformed proportions but this made no significant difference to the results.

Table 3. *Mean proportion correct responses for the production task as a function of age, pronoun type, verb frequency and sentence-picture pair status ((mis)match)*

	Matching sentence-picture						Mismatching sentence-picture					
	High freq verbs			Low freq verbs			High freq verbs			Low freq verbs		
	Her	Him	Them	Her	Him	Them	Her	Him	Them	Her	Him	Them
4yrs	1.0	1.0	1.0	1.0	1.0	1.0	.83	.78	.76	.77	.72	.69
5yrs	1.0	1.0	1.0	1.0	1.0	1.0	.86	.89	.86	.89	.89	.87
6yrs	1.0	1.0	1.0	1.0	1.0	1.0	.89	.81	.85	.87	.85	.89

To test for predicted lexical effects on pronoun interpretation, we assessed children's accuracy on the TVJ task with a single measure of accuracy which reflects the ability to reject mismatching sentences and accept matching sentences. To do this we plotted an ROC (receiver operating characteristic) line for each child (x axis: probability of saying "yes" to mismatching sentences, y axis: probability of saying "yes" to matching sentences) and used the area under the line as a statistic for accuracy, A. An A score of 0.5 represents chance responding and 1.0 is ceiling accuracy.

A 3 (age) $\times$ 3 (pronoun) $\times$ 2 (verb frequency) ANOVA with A scores as the dependent measure revealed a significant interaction between pronoun and verb frequency ($F(2, 120) = 6.8$, $p = 0.002$, $\eta^2 = 0.102$), a significant effect of pronoun ($F(2, 120) = 22.8$, $p < 0.001$, $\eta^2 = 0.276$) and a borderline effect of verb frequency ($F(1, 60) = 3.4$, $p = 0.068$, $\eta^2 = 0.054$) but no effect of age. Pair-wise comparisons revealed that, for sentences with high frequency verbs, children performed significantly better with the pronoun *her* than with *them* ($p < 0.001$) or *him* ($p < 0.001$) and better with the pronoun *him* than *them* ($p < 0.001$). For sentences with low frequency verbs, the children performed better with sentences containing the pronoun *her* and *him* than *them* ($p = 0.001$, $p = 0.003$). Contrary to the usage-based prediction, children were significantly more accurate when judging the pronouns *them* and *him* when the sentence contained a *low* frequency verb ($p = 0.005$, $p = 0.044$). No such verb frequency effect was observed for sentences containing the pronoun *her*. These results are illustrated in Figure 2.

In the production task, all the children were at ceiling with their comments on pictures that had matching sentences. For mismatching sentence-picture pairs, a 3 (age) $\times$ 3 (pronoun) $\times$ 2 (verb frequency) ANOVA was performed with the proportion of comments that were correct as the dependent measure. There was a borderline interaction between age and verb frequency ($F(2, 58) = 3.01$, $p = 0.057$, $\eta^2 = 0.094$),

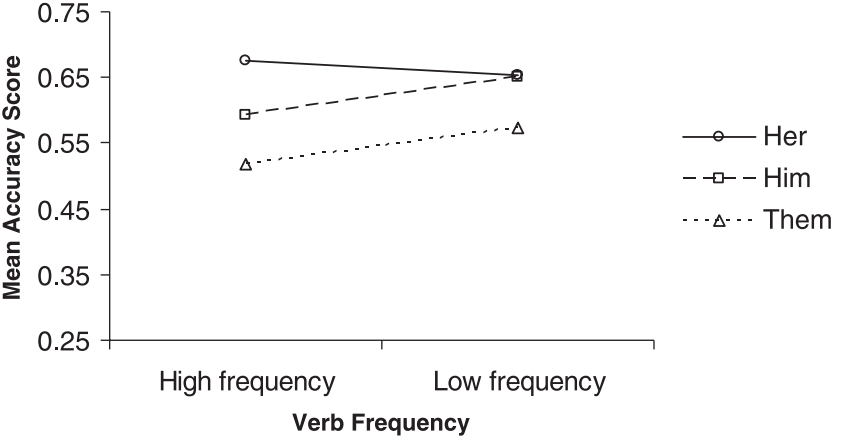

Figure 2. *Accuracy scores for the TVJ task as a function of pronoun and verb frequency*

such that the youngest children tended to perform better with the high frequency verbs ($p = 0.078$). There were no other significant effects or interactions.

3.3. *Control sentences*

Only four of the youngest children incorrectly accepted one of the two control sentences in which the experimenter guessed incorrectly with two full nouns (e.g., "*The mum is washing the girl*"). With control sentences that contained a reflexive pronoun (e.g., "*The mum is washing herself*") only two children incorrectly accepted the control sentence that should have been rejected. We can hence be confident that children did not show a strong "yes" bias with these types of sentences.

3.4. *Order effects*

Given the number of items tested it was important to test whether children began to say "yes" more (or less) as the experiment progressed. We thus performed a chi square test on the number of yes responses given for each block of sentences for children in the three counterbalanced order conditions (using a 3 (block) by 3 (order of presentation) table). This revealed no effect of order of presentation ($df = 4$, $\chi^2 = 5.9$ $p > 0.05$).

3.5. *Qualitative analysis*

In order to assess individual response patterns the following qualitative analysis was performed. For each sentence type (matching and mismatch-

Table 4. *Number of children following different qualitative response patterns as a function of age*

	Mat "yes" Mis "yes"	Mat "yes" Mis "no"	Mat "yes" Mis both	Mat "no" Mis "yes"	Mat "no" Mis both	Mat both Mis "yes"	Mat both Mis both	Total
4 yrs	10	4	5	0	0	2	0	21
5 yrs	5	8	2	0	0	5	1	21
6 yrs	7	5	5	3	1	0	0	21
Total	22	17	12	3	1	7	1	63

ing) each child was categorized as either generally responding "*yes*" (more than 75 percent of responses were "yes"), generally responding "*no*" (more than 75 percent of responses were "no"), or responding with "*both*" forms (i.e., neither yes nor no responses constituted over 75 percent of responses). The categories for matching and mismatching sentence types were combined to give an overall qualitative category. For example, one qualitative category would be for children who always said yes whatever the sentence was. This category is labeled "matching yes, mismatching yes" (Mat "yes", Mis "yes"). There were 7 different combinations of response types that the children gave, these are shown in table 4.

The vast majority of children generally responded *yes* to the matching sentences. In fact 22 children most often responded "yes" to both matching and mismatching sentences. We might take this as evidence of a general "yes" bias. However, it should be noted that these children did not show this bias in the control sentences so what we are observing here is a tendency to accept incorrect uses of pronouns and not a general tendency to accept the adult's guess.

It is also important to note that three of the eldest children responded *no* to matching sentences yet *yes* to mismatching sentences. These children's comments suggested that they thought the pronoun used in matching sentences should be replaced by a full name, as this was always their preferred response in the production task (one child also accepted control reflexive sentences on the TVJ task but switched the reflexive pronoun to him/her on the production task). In total 11 children were more likely to give a negative response to a matching sentence than to a mismatching one.

4. Discussion

The current study set out to test three key predictions following from generativist and usage-based accounts of children's co-referencing errors. The

results did not confirm any of the predictions and suggest that a more nuanced view of the task and what it is testing is necessary.

The first striking finding was that even the oldest children were well below ceiling in their judgments of quantifier sentences. These sentences were tested to check generativist predictions that they would be resistant to error and to see if pronoun number would affect error-rates. In fact children are not more likely to accept co-reference in sentences with a third person plural pronoun *them* as we predicted. Surprisingly, though, children did make a considerable number of errors with sentences of precisely the same form as those tested by Chien and Wexler (1990). Chien and Wexler found that five- and six-year-olds correctly rejected infelicitous uses of sentences such as "*Every bear is touching her*" 84 percent and 87 percent of the time respectively, which they judged to be ceiling performance. Whatever the criterion for ceiling performance may be (see Drozd 2004 for a discussion of this issue), the children in the current study did not conform to it: children aged 6 years were at chance when required to reject sentences of the type "*Every boy is washing/tickling/hitting him*". It is not clear what might have caused us to obtain such different results to those of Chien and Wexler and future experiments would be needed to replicate the current findings with more items. Nonetheless, this finding seriously calls into question whether six-year-olds do meet the acid test for knowledge of principle B (see also Matsuoka 1997).

The second relevant finding is that, in the TVJ task, children tended to accept sentences that mismatched pictures but then, in the production task, they would go on to produce appropriate descriptions by replacing the infelicitous pronoun with a reflexive. This does not support the guise-creating hypothesis, which proposes that children think the pronoun is marking an element of surprise. If this were the case we would expect the children to judge the pronoun to be correct and then to use it in production too—there would be no reason to produce a reflexive. Yet the children consistently produced appropriate reflexive forms (as they do in naturalistic speech—Bloom, Barss, Nicol and Conway 1994). This raises the question as to whether children simply have difficulty with the TVJ task. Two plausible explanations for such difficulties come to mind.

The first explanation is that children might not notice that the sentence final pronoun is not a reflexive. There is significant phonological overlap between reflexives and third person pronouns in English and it might be that children, having seen the picture, were expecting to hear "*The boy hit himself*" and did not notice that the "self" element was missing. Indeed in languages such as Italian, children make far fewer co-referencing errors (McKee 1992) perhaps because the offending pronoun occurs earlier in

the sentence and is phonetically very different to its reflexive counterpart (but see Avrutin 1999). Both of these factors would make the pronoun mismatch much easier to detect in the speech stream, something which connectionist modelers have argued is fundamental in anaphora resolution (Joanisse and Seidenberg 2003).

The second explanation is that children do notice the pronoun but are unsure whether it should be judged as incorrect. Making such a judgment requires meta-linguistic reasoning of the form "if she meant this, she really ought to have said *x* but she said something that means *y* and therefore her guess is wrong". The current results would suggest that most children aged 4 to 6 are not confident in making such inferences about pronouns and reflexives but do nonetheless contrast these forms appropriately in production (see Hendriks and Spenader 2005/2006 for a similar proposal in the Optimality Theory framework). This problem with rejecting pronouns in favour of reflexives might be compounded by the late development of reflexives, which are first learnt in English with an intensifier function (e.g., *I can do it by myself*), rather than a true reflexive function (Thomas 1994). Given the plausibility of this explanation, we conclude that the TVJ task is a problematic method, no matter what theoretical approach one wishes to test.

The sheer difficulty of the truth-value judgment task might go some way to explaining the third major finding of this study: the inverse effect of verb frequency on judgment accuracy where children were more accurate in judging sentences that contained low frequency verbs than high frequency verbs (for sentences containing the pronouns "*him*" and "*them*"). This effect does not conform to the predictions of usage-based accounts, which predict greater accuracy with items containing higher frequency verbs or generativist accounts, which do not predict such frequency effects at all. However, if children encounter serious difficulty performing truth-value judgment tasks, this inverse effect of frequency might be explained in terms of a processing effect. Children have often been shown to have trouble revising their initial interpretations of sentences, a phenomenon known as the kindergarten path effect (Trueswell, Sekerina, Hill and Logrip 1999). It might be that such an effect would be stronger with sentences containing lexical items that children know better. In the current experiments, then, children would be very quick to judge sentences containing high frequency verbs, perhaps not waiting to hear the pronoun before beginning the judgment or initially preferring a sentence internal referent (Sekerina, Stromswold and Hestvik 2004). In contrast, sentences containing lower frequency items would involve more initial uncertainty and therefore greater dependence on the pronoun downstream of the verb, which would ultimately increase accuracy. This explanation

is, of course highly speculative, and would benefit from being tested with online methods, such as reaction time tests or eye-tracking.

A processing explanation in terms of a speed/accuracy trade-off, may not explain all the results, however. The significant effects of pronoun are more difficult to explain purely in terms of frequency (and speed of processing). The highest frequency accusative pronoun, *them*, was generally judged less accurately than the pronoun *her* and *him* (accuracy: *her* > *him* > *them*), which is in line with the speed/accuracy trade-off hypothesis. However, the second most frequent accusative pronoun, *her*, was judged more accurately than *him*, the least frequent pronoun, which contradicts the trade-off hypothesis. Of course, to properly assess the role of lexical frequency it would be necessary to consider the frequency with which the relevant reflexive pronouns occur, the ratio between the reflexive and the corresponding accusative pronoun and their use with respect to specific verbs and constructions. However, since we know that children suffer from genuine confusion as to the referential properties of the third person plural pronoun (Tanz 1976), it seems plausible that frequency of form interacts with other areas of development, such as the concept of number, and is not a sole explanatory factor. Cross-linguistic studies would be extremely valuable here to properly weigh up the effects of language-specific effects and any more general cognitive developmental effects (c.f. Hickmann and Hendriks 1999).

Taken together the current results pose serious challenges for both generativist and usage-based approaches to language acquisition. Generativist approaches predict and report ceiling performance on sentences where co-reference is only governed by Principle B. This ceiling performance was not replicated in the current study. Furthermore, results from the production test make it implausible that children are over-creating guises when making co-reference errors as Thornton and Wexler (1999) propose. On the other hand, usage-based accounts did not predict children to be more accurate with lower frequency items. A lack of frequency effect might have been explained by the fact that children have a fairly robust command of the transitive construction by the age of three, even on usage-based accounts. However, the *inverse* effect of verb frequency suggests that frequency effects take on quite a different appearance once test constructions are familiar to the child (Van Gompel and Majid 2004 for further inverse frequency effects in child and adult processing; c.f. Abbot-Smith et al. submitted). This suggests that if we construct theories purely on the basis of representations and their acquisition at the expense of the processes by which representations are used, we risk painting a picture of linguistic development that is incompatible with the data we find.

Nonetheless, the above findings provide substantial motivation for further developing usage-based accounts of language acquisition. This optimism is supported by demonstrations of the functional underpinnings of anaphora constraints (van Hoek, 1997; Harris and Bates 2002) and proposals that these need not be taken as a priori unlearnable (Akhtar et al. 2004; Lidz et al. 2003;; MacWhinney, 2004). The proposal put forward here is that children's knowledge of anaphora constraints depends on understanding i) the relative accessibility of different referring expressions ii) the contexts in which referring expressions occur, particularly in within-sentence grammatical hierarchies and iii) the contrastive values of pronouns and reflexives. We focused here on the second of these points by looking for gradual development of the complement chain, which we predicted would show up in frequency effects such that children would be more accurate in judging sentences made up of high frequency lexical items. These predictions were not confirmed, although interesting lexical effects were revealed.

Future work might investigate the other two proposed elements of the usage-based account of anaphora: understanding the relative accessibility of different referring expressions and the contrastive values of pronouns and reflexives. The discrepancy between the production and comprehension results that is apparent when children judge pronouns (but not full noun control sentences) suggests a deep-seated uncertainty as to the appropriate linguistic contexts for pronoun use. Given the pluri-functionality of these words, it is understandable that children might perceive pronouns as ''passe-partout'' words that apparently can refer to anything. The above qualitative analysis suggests that this was the case for the majority of children. However, a handful of the eldest children were clearly wrestling with the permissibility of pronouns and judging even ''matching'' uses as incorrect on the basis that the guesser should have used a full noun. Establishing how children come to understand the relative accessibility and contrasting functions of different referring expressions will thus be central to developing a usage-based account of anaphora.

Received 11 December 2007
Revision received 14 July 2008

University of Manchester

Appendix: Test sentences

The order of presentation of the three blocks was counterbalanced.

Block 1

	Guess sentence	Match ?
1	The boy is tickling him	mismatch
2	The mum is pinching her	match
3	Every boy is washing him	mismatch
4	The girl is tickling her	match
5	The girls are lassoing them	mismatch
6	The girl is tickling the mum	match
7	The boys are pinching them	match
8	The boys are hitting them	mismatch
9	Every boy is tickling them	match
10	The girl is tickling her	mismatch
11	The girls are washing them	match
12	The dad is washing the boy	mismatch
13	The girl is lassoing her	mismatch
14	The dad is pinching him	match
15	Every boy is hitting him	mismatch
16	The boy is tickling him	match
17	The boy is spraying him	mismatch

Block 2

	Guess sentence	Match ?
18	The dad is washing him	mismatch
19	The girl is spraying her	match
20	Every boy is tickling them	mismatch
21	The mum is washing her	match
22	The boys are pinching them	mismatch
23	The mum is tickling herself	mismatch
24	The girls are spraying them	match
25	The boys are tickling them	mismatch
26	Every boy is washing them	match
27	The mum is washing her	mismatch
28	The boys are hitting them	match
29	The boy is washing himself	match
30	The mum is pinching her	mismatch
31	The boy is lassoing him	match

32	Every boy is hitting them	mismatch
33	The dad is pinching him	mismatch
34	The dad is washing him	match

Block 3

	Guess sentence	Match ?
35	The boy is hitting him	mismatch
36	The girl is lassoing her	match
37	Every boy is washing him	match
38	The girl is hitting her	match
39	The girls are spraying them	mismatch
40	The girls are hitting themselves	match
41	The girls are lassoing them	match
42	The girls are washing them	mismatch
43	Every boy is tickling him	mismatch
44	The girl is hitting her	mismatch
45	The boys are tickling them	match
46	The mums are hitting the girls	mismatch
47	The girl is spraying her	mismatch
48	The boy is spraying him	match
49	Every boy is washing them	mismatch
50	The boy is hitting him	match
51	The boy is lassoing him	mismatch

References

Abbot-Smith, K., M. Dittmar and M. Tomasello
(submitted) Frequency and reverse frequency effects in the acquisition of verb syntax.
Akhtar, N., M. Callanan, G. Pullum, and B. Scholz
2004 Learning antecedents for anaphoric one. *Cognition* 93, 141–145.
Ariel, M.
1990 *Anaphoric Antecedents.* London: Croom Helm.
Avrutin, S.
1994 *Psycholinguistic Investigations in the Theory of Coreference*. Cambridge, MA: MIT.
Avrutin, S.
1999 *Development of the Syntax-Discourse Interface.* Dordrecht: Kluwer.
Baauw, S. and F. Cuetos
2003 The interpretation of pronouns in Spanish language acquisition and breakdown: Evidence for the "Principle B Delay" as a non-unitary phenomenon. *Language Acquisition* 11, 219–275.

Bloom, P., A. Barss, J. Nicol and L. Conway
1994 Children's knowledge of binding and coreference: Evidence from spontaneous speech. *Language* 70, 53–71.
Cantrall, W.
1974 *View Points, Reflexives and the Nature of Noun Phrases.* The Hague: Mouton.
Chien, Y. and K. Wexler
1990 Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition* 1, 225–295.
Chomsky, N.
1981 *Lectures on Government and Binding.* Dordrecht: Forris.
Chomsky, N.
1995 *The Minimalist Program.* Cambridge, MA: MIT Press.
Croft, W.
2001 *Radical Construction Grammar: Syntactic Theory in Typological Perspective.* Oxford: Oxford University Press.
Drozd, K.
2004 Learnability and linguistic performance. *Journal of Child Language* 31, 431–457.
Foster-Cohen, S.
1994 Exploring the boundary between syntax and pragmatics: Relevance and the binding of pronouns. *Journal of Child Language* 21, 237–255.
Goldberg, A. E.
1995 *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago: University of Chicago Press.
Goldberg, A. E.
2006 *Constructions at Work: The Nature of Generalization in Language.* Oxford: Oxford University Press.
Grimshaw, J. and S. Rosen
1990 Knowledge and obedience: The developmental status of binding theory. *Linguistic Inquiry* 21, 187–222.
Grodzinsky, Y.
(in prep) Coreference and self-ascription. Ms., McGill University.
Grodzinsky, Y. and T. Reinhart
1993 The innateness of binding and coreference. *Linguistic Inquiry* 24, 69–101.
Gundel, J., N. Hedberg and R. Zacharski
1993 Cognitive status and the form of referring expressions in discourse. *Language* 69, 274–307.
Gundel, J., M. Hegarty and K. Borthen
2001 Information structure and pronominal reference to clausally introduced entities. In Kruijff-Korbay, I. and M. Steedman (eds.), *Workshop on Information Structure. ESSLLI*. Helsinki, 37–51.
Harris, C. L. and E. Bates
2002 Clausal backgrounding and pronominal reference: A functionalist approach to c-command. *Language and Cognitive Processes* 17, 237–269.
Hendriks, P. and J. Spenader
2005/2006 When production precedes comprehension: An optimization approach to the acquisition of pronouns. . *Language Acquisition* 13, 319–348.
Hestvik, A. and W. Philip
1999/2000 Binding and coreference in Norwegian child language. *Language Acquisition* 8, 171–235.

Hickmann, M. and H. Hendriks
1999 Cohesion and anaphora in children's narratives: A comparison of English, French, German and Mandarin Chinese. *Journal of Child Language* 26, 419–452.
Jackendoff, R.
2002 *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford: Oxford University Press.
Joanisse, M. F. and M. S. Seidenberg
2003 Phonology and syntax in specific language impairment: Evidence from a connectionist model. *Brain and Language* 86, 40–56.
Keenan, E. and B. Comrie
1977 Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 1, 63–99.
Langacker, R. W.
1985 Observations and speculations on subjectivity. In Haiman, J. (ed.), *Iconicity in Syntax.* Amsterdam: John Benjamins.
Langacker, R. W.
2000 A dynamic usage-based model. In Kemmer, S. and M. Barlow (eds.), *Usage Based Models of Language*. Stanford, CA: CSLI.
Lidz, J., S. Waxman and J. Freedman
2003 What infants know about syntax but couldn't have learned: Evidence for syntactic structure at 18-months. *Cognition*, B65–B73.
Lust, B., G. Hermon and J. Kornfilt
1994 *Syntactic Theory and First Language Acquisition: Cross Linguistic Perspectives.* Mahwah, NJ: Lawrence Erlbaum Associates.
MacWhinney, B.
2000 *The CHILDES Project: Tools for Analyzing Talk.* Mahwah, NJ: Lawrence Erlbaum Associates.
MacWhinney, B.
2004 A multiple process solution to the logical problem of language acquisition. *Journal of Child Language* 31, 883–914.
Matsuoka, K.
1997 Binding conditions in young children's grammar: Interpretation of pronouns Inside conjoined NPs. *Language Acquisition* 6, 37–48.
Matthews, D. E., E. Lieven and M. Tomasello
2007 How toddlers and preschoolers learn to uniquely identify referents for others: A training study. *Child Development* 78, 1744–1759.
Matthews, D. E., A. Theakston, E. Lieven and M. Tomasello
2006 The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics* 27, 403–422.
McDaniel, D. and T. Maxfield
1992 Principle B and contrastive stress. *Language Acquisition* 2, 337–358.
McKee, C.
1992 A comparison of pronouns and anaphors in Italian and English acquisition. *Language Acquisition* 2, 21–54.
Philip, W. and P. Coopmans
1996 The role of referentiality in the acquisition of pronominal anaphora. *North East Linguistic Society*, University of Massachusetts, Amherst.
Reinhart, T.
1983 *Anaphora and Semantic Interpretation.* London: Croom Helm.

Reinhart, T. and E. Reuland
1993 Reflexivity. *Linguistic Inquiry* 24, 657–720.
Runner, J., R. Sussman and M. Tanenhaus
2003 Assignment of reference to reflexives and pronouns in picture noun phrases: Evidence from eye movements. *Cognition* 89, B1–B13.
Sekerina, I. A., K. Stromswold and A. Hestvik
2004 How do adults and children process referentially ambiguous pronouns? *Journal of Child Language* 31, 123–152.
Tanz, C.
1976 Learning how it works. *Journal of Child Language* 4, 224–235.
Theakston, A. L., E. Lieven, J. Pine, and C. Rowland
2001 The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language* 28, 127–152.
Thomas, M.
1994 Young children's hypotheses about English reflexives. In Solokov, J. and C. Snow (eds.), *Handbook of Research in Language Development using CHILDES.* Mahwah, NJ: Lawrence Erlbaum Associates.
Thornton, R. and K. Wexler
1999 *Principle B, VP Ellipsis and Interpretation in Child Grammar.* Cambridge, MA: The MIT Press.
Tomasello, M.
2003 *Constructing a Language: A Usage-Based Theory of Language Acquisition.* Harvard: Harvard University Press.
Trueswell, J. C., I. A. Sekerina, N. M. Hill and M. L. Logrip
1999 The kindergarten-path effect: Studying on-line processing in young children. *Cognition* 73, 89–134.
Van Gompel, R. P. G. and A. Majid
2004 Antecedent frequency effects during the processing of pronouns. *Cognition* 90, 255–264.
van Hoek, K.
1997 *Anaphora and Conceptual Structure.* Chicago: University of Chicago Press.
van Hoek, K.
2003 Pronouns and point of view: Cognitive principles of co-reference. In Tomasello, M. (ed.), *The New Psychology of Language*. Mahwah, NJ: Lawrence Erlbaum Associates, 169–194.