J. Child Lang. **34** (2007), 701–724.
doi:10.1017/S0305000907008124 Printed in the United Kingdom

Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task*

CARMEL HOUSTON-PRICE

School of Psychology & Clinical Language Sciences, University of Reading, UK

EMILY MATHER

School of Psychology & Clinical Language Sciences, University of Reading, UK

AND

ELENA SAKKALOU

School of Psychology, Cardiff University, UK

(*Received* 24 *July* 2006. *Revised* 31 *October* 2006)

ABSTRACT

Two experiments are described which explore the relationship between parental reports of infants' receptive vocabularies at 1;6 (Experiment 1a) or 1;3, 1;6 and 1;9 (Experiment 1b) and the comprehension infants demonstrated in a preferential looking task. The instrument used was the Oxford CDI, a British English adaptation of the MacArthur-Bates CDI (Words & Gestures). Infants were shown pairs of images of familiar objects, either both name-known or both name-unknown according to their parent's responses on the CDI. At all ages, and on both name-known and name-unknown trials, preference for the target image increased significantly from baseline when infants heard the target's label. This discrepancy suggests that parental report underestimates infants' word knowledge.

[*] The research reported in this article was supported by a project grant from the British Academy (SG-35100). Portions of the data were presented at the conferences of the International Society for Infant Studies (1999) and the Society for Research in Child Development (2005). We would like to express our thanks to the parents and infants who participated in the studies reported and to Graham Schafer, Letitia Naigles and an anonymous reviewer for their comments on a draft of this paper. Address for correspondence: Dr Carmel Houston-Price, School of Psychology, University of Reading, Earley Gate, Whiteknights, Reading, RG6 6AL, UK. Email: c.houston-price@reading.ac.uk. Tel: 0118 378 5378. Fax: 0118 9316715.

INTRODUCTION

Until the 1980s, studies of vocabulary development involved expensive and time-consuming longitudinal observations of small numbers of children. Due to the difficulties of assessing comprehension in children under two years, research was also largely restricted to the study of productive vocabulary. During the past twenty years, however, other methods for assessing comprehension in the home and the laboratory have enabled researchers to track the language development of both younger age groups and much larger populations.

One of the tools that has been widely used by researchers is the parent report questionnaire, because it can provide 'quick and easy' data on the vocabulary development of large, cross-sectional groups of children. The report measures that are most widely used are the MacArthur-Bates Communicative Development Inventories (CDI/*Words & Gestures*, for infants aged 0;8 to 1;4 and CDI/*Words & Sentences*, for toddlers aged 1;4 to 2;6; Fenson *et al.*, 1993). The vocabulary sections of these instruments list several hundred early-learned words and ask parents to indicate which their child understands and which the child both understands and says (in the form for toddlers, parents report only which words are produced). From the responses provided by the parents of 1700 children, Fenson, Dale, Reznick, Bates, Thal & Pethick (1994) established norms and percentile scores for the development of receptive and productive vocabulary. Since its inception, the use of the CDI has been prolific, both in its original form, in its translations into other languages (e.g. Hebrew: Maital, Dromi, Sagi & Bornstein, 2000; New Zealand English: Reese & Read, 2000) and in its adaptations that assess both production and comprehension beyond the 0;8 to 1;4 age range (e.g. Infant Short Form: Fenson, Pethick, Renda, Cox, Dale & Reznick, 2000; Oxford CDI: Hamilton, Plunkett & Schafer, 2000).

The original aim of the MacArthur-Bates CDI was to provide a global measure of vocabulary that would allow a child's language level to be compared to norms for the appropriate age group (Fenson *et al.*, 1994). The instrument has since been used both for this purpose and to compare the developmental progress of distinct groups, including typically- and atypically-developing children (autism: Charman, Drew, Baird & Baird, 2003; developmental delay: Yoder, Warren & Biggar, 1997), children from different socio-economic classes (Fenson *et al.*, 1993; Arriaga, Fenson, Cronan & Pethick, 1998; Feldman, Dollaghan, Campbell, Kurs-Lasky, Janosky Paradise, 2000; Reese & Read, 2000) and children from different countries (Caselli *et al.*, 1995; Hamilton *et al.*, 2000; Maital *et al.*, 2000). Total scores on the CDI have also been used to compare the reports provided by parents and daycare teachers (Weitzner-Lin, 1996).

Users of the CDI believe that the instrument is valid and reliable, pointing out that parents have a unique opportunity to observe developments in their child's vocabulary. However, some researchers have expressed concerns about parents' ability to accurately assess their child's understanding of a word (e.g. Tomasello & Mervis, 1994; Golinkoff, Hirsh-Pasek, Cauley & Gordon, 1987). Estimates of comprehension are more subjective than estimates of production, as they require parents to notice children's non-verbal responses to words. Reports of comprehension may therefore be erroneously based on child behaviours that are cued by the context in which a word is heard, rather than on behaviours that are a direct response to the word itself.

Similarly, Tomasello & Mervis (1994) suggested that parents' comprehension attributions might depend on the child's familiarity with the referent of a word. Rather than answering the question 'Does your child know what the word *X* means?', they answer 'Does your child know what an *X* is?' Tomasello & Mervis propose that the difficulty of responding to the CDI's demand to mark WORDS YOUR CHILD UNDERSTANDS BUT DOES NOT YET SAY leads to over-reporting by parents, and as evidence they cite cases of parents claiming comprehension of up to 150 words for infants of only 0;8 to 0;10. Tomasello & Mervis' claims about parents' confusion are supported by the changes they observed in parents' reports when parents were given additional information about what was meant by 'comprehension' before they completed the CDI. When parents were told that the researchers were interested in whether their child understood the meaning of each word on the list, regardless of any accompanying behaviour the parent might make when uttering the word, there was a dramatic decrease in the number of words parents marked as known.

Given the widespread use of report instruments, we think it is vital to establish that parents are providing accurate information. That is, to what extent do reports of comprehension map onto the child's actual understanding of the same vocabulary items? One approach to investigating parental accuracy has been to explore the reliability of their responses on the CDI. Yoder *et al.* (1997) used the stability of reports across a two-week test delay as an indirect measure of accuracy. They found a high level of 'summary-level' stability in the total number of words that parents of 17 developmentally-delayed children reported to be comprehended, with correlations averaging 0·93. However, stability was markedly poorer at the 'item-level', with the best correlation at 0·65 for nouns. Moreover, as Yoder *et al.* point out, 'stability is not synonymous with accuracy' (p. 64). Given the rapid gains in vocabulary that are seen during the second year, when several words are learned every day, a degree of instability would be expected from accurately-reporting parents. That is, parents who are sensitive to their child's word learning SHOULD produce somewhat discrepant reports over a two-week delay, as new words are added to the child's repertoire. The precise value of the reliability quotient that would support parental accuracy is therefore unclear.

A more direct approach to exploring the CDI's validity has been to seek relationships between reported vocabulary and infants' performance on secondary assessments of linguistic competence, such as the Peabody Picture Vocabulary Test (PPVT) or the receptive language subscales of the Bayley Scales of Infant Development. With the exception of Reznick (1990), such studies have found a strong relationship between total CDI scores and standardized assessments (Bates, Bretherton & Snyder, 1988; Dale, Bates, Reznick & Morisset, 1989). Longitudinal studies also suggest predictive relationships between CDI scores at 1;6 to 2;0 and productive language at 3;0 (Feldman *et al.*, 2005).

However, correlations involving total CDI scores do not provide an index of report accuracy. This is because CDI scores and alternative language tests may correlate highly even if parents are poor estimators of their child's total vocabulary size, as long as parents are able to rank their child's ability RELATIVE TO PEERS. That is, if a group of parents systematically overestimated children's vocabulary by even a large margin of error, the strength of the correlation between their reports and the second language assessment would remain unaffected (for more on this, see Evans & Wodar, 1997; Yoder *et al.*, 1997). Thus, correlational studies say nothing about the absolute accuracy of parents' estimates.

Nor do such 'total scores' correlations speak to the accuracy of reports at the ITEM LEVEL, for they need not reflect concurrence between report and test performance with respect to individual words on the CDI. Indeed, even where a strong relationship is found between total CDI scores and scores on an alternative vocabulary task, there may be little or no overlap between the individual words that the two measures declare to be known. As long as parents are able to estimate the child's total vocabulary size (either in absolute terms or relative to their peers), correlations between reports and test scores could be found even if parents' marking of specific words on the CDI were COMPLETELY RANDOM. This problem is exemplified by Ring & Fenson's (2000) finding that two-year-olds' scores on a production task were as highly correlated with productive vocabulary reports for a random set of 35 nouns, verbs and adjectives as they were with reports for the 35 items actually tested.

We think it is important to establish the accuracy of reports at the item level because, as well as providing a global measure of a child's vocabulary size, the CDI has been used to identify specific words or classes of words a child does (or does not) understand. Such use is seen in both research (e.g. Mills, Plunkett, Prat & Schafer, 2005) and clinical settings (see Yoder *et al.*, 1997), where the CDI is used to determine targets for remediation. If report instruments are to be validated for these types of use, a relationship must be demonstrated between the individual items that are marked as understood on the CDI and those for which comprehension is shown in a controlled setting. Yet few studies to date have explored this relationship. For example, when Ring & Fenson (2000) compared parents' comprehension judgements for 35 test items with children's performance in a picture book pointing task, they reported a discrepancy in the total scores produced by the two tasks but provided no information about the correspondence at the item level.

Of those studies that have compared infants' responses to specific words marked as known or unknown by parents, the majority have measured electro-physiological (rather than behavioural) responses (e.g. Molfese, 1990; Mills, Coffey-Corina & Neville, 1993; Mills *et al.*, 2005). For example, Mills *et al.* (1993) compared 20-month-olds' event-related potentials (ERPs) to words rated as 'definitely known' or 'definitely unknown' by parents. The finding that ERPs to the two sets of stimuli were reliably different is often cited in support of the accuracy of parental report.

However, the demand made of parents in these studies – to select a few definitely known and definitely unknown words from a list – is not akin to the CDI's requirement to mark every listed word as known or unknown. Greater accuracy would be expected in Mills *et al.*'s (1993) procedure, as parents were not required to report on items about which they may have been uncertain. Furthermore, Mills *et al.* selected unknown words with very low input frequency, meaning that the differences in ERP activity on hearing the known and unknown words may be explained by infants' differing levels of familiarity with the words' sounds rather than differences in their understanding of the words' meanings. It is also worth noting that the ERP responses to known and unknown words reported by Molfese (1990) differed markedly from those found by Mills *et al.*, in their latencies, distributions and polarity. Mills *et al.* attribute the discrepancy to Molfese's use of unknown words that were more likely to be familiar to infants, thereby acknowledging that responses may have been driven by word familiarity, not comprehension. Thus, while these studies offer insights into the neural bases of early lexical representations, they do not demonstrate the accuracy of parents' reports.

How else might one explore whether words marked as known or unknown on the CDI are indeed known or unknown by the child concerned? Over the past twenty years, the Intermodal Preferential Looking procedure (Golinkoff *et al.*, 1987) has proved increasingly useful for assessing word comprehension. In this paradigm, infants are simultaneously presented with two images and a word or sentence referring to one image, and cameras record the duration of the infant's looks towards each image. Comprehension of the word or sentence is typically reflected in longer looks towards the named image. Numerous studies have demonstrated the use of the paradigm in detecting comprehension of words learned at home (e.g. Hirsh-Pasek & Golinkoff, 1996; Tincoff & Jusczyk, 1999) and words

https://doi.org/10.1017/S0305000907008124 Published online by Cambridge University Press

learned in the laboratory (Schafer & Plunkett, 1998; Graham & Poulin-Dubois, 1999; Houston-Price, Plunkett & Harris, 2005; Houston-Price, Plunkett & Duffy, 2006). While the mechanisms that drive eye movements in this paradigm are relatively unexplored, it is necessarily the case that a systematically positive result (significantly longer looking at the named target) will only occur if the infant has some understanding of the link between the word heard and its referent.

Of course, if parents are to report accurately on their child's understanding of a word, they must detect evidence of the same referential link. That is, the two measurements necessarily tap the same core aspect of word knowledge. One might therefore expect the two indices to coincide in indicating the same words as understood. However, there are also important differences between the tasks. For example, parents have the opportunity to observe their child's understanding of a word in a range of contexts, while preferential looking studies assess knowledge in a context divorced from the infant's usual experiences of the word. One might therefore expect a preferential looking task to fail to detect comprehension of some words that parents report to be known.

A number of preferential looking studies to date have corroborated parents' reports of known words (e.g. Thomas, Campos, Shucard, Ramsay & Shucard, 1981; Behrend, 1988; Robinson, Shore, Hull Smith & Martinelli, 2000). For example, Thomas *et al.* asked parents to select one word that they were sure their infant understood, and one they were sure their infant did not understand, from a list of twenty words. In a subsequent preferential looking task, infants aged 1;1 preferentially fixated the target image when they heard the known word, but not when they heard the unknown word, in line with parents' predictions. However, as we discussed earlier, the requirement to select one definitely known or definitely unknown word from a list does not make the same exacting demands on parents' vocabulary knowledge as the CDI.

Robinson *et al.* (2000) explored the relationship between infants' preferential looking behaviour and parents' responses to a telephone interview about their child's vocabulary. Parents were given three response options relating to comprehension: words were rated as 'understood', 'heard before but not understood' (which the authors term 'frontier words') or 'never heard before'. In the preferential looking task, infants of 1;3, 1;6 and 1;10 showed understanding of words marked as understood, and no understanding of words never heard before, supporting parents' reports. However, the oldest group also showed comprehension of the frontier words. Thus, parents knew which words their children were familiar with, but underestimated the number of these that their children understood (cf. Tomasello & Mervis, 1994). However, it is, as yet, unknown whether underestimation might similarly be elicited by the CDI's two response options relating to comprehension. This is the aim of the studies reported in this article: to systematically explore infants' behavioural responses to items marked as known or unknown by their parents on the CDI.

We compared parents' reports of infants' word comprehension with the preferential looking behaviour of infants aged 1;3, 1;6 and 1;9. The original CDI omitted a comprehension scale beyond 1;4 on the grounds that, at a time of such rapid vocabulary growth, parents might confuse which words are understood and which produced. However, recent research has concluded that reports may be useful for assessing receptive vocabulary well beyond this age, and suggests that a comprehension checklist should be added to Toddler Forms of the CDI (Ring & Fenson, 2000). Ring & Fenson also recommend that parents' responses should be compared with an experimental measure of child performance, to establish the validity of reports for individual items. A second aim of our studies was therefore to establish whether report accuracy changes during the child's second year. If underestimation is common for older infants with larger vocabularies, as Robinson *et al.*'s (2000) data suggest, it may be inappropriate to incorporate a receptive vocabulary scale for children beyond a certain age. We also explore whether parents report more accurately on their sons' or daughters' vocabulary. Girls are reported to have significantly larger vocabularies throughout the second year (Fenson *et al.*, 1993; Feldman *et al.*, 2000, 2005; Reese & Read, 2000) and, if the accuracy of parents' reports is inversely proportional to vocabulary size, greater accuracy would be expected for boys.

In our studies, parents were asked to indicate which of the words on a British English CDI were 'understood' or 'both understood and produced' by their child. The Oxford CDI (Hamilton *et al.*, 2000) was ideal for our purposes as it combines the vocabulary sections of the MacArthur-Bates CDI/*Words & Gestures* and the CDI/*Words & Sentences* in a single checklist, allowing reports of receptive vocabulary (as well as productive vocabulary) to be obtained throughout the second year. After parents completed the CDI, infants took part in a preferential looking task in which they were presented with pairs of images of objects while they heard one of the objects' labels. Trials were either 'name-known', when the names of both presented images were reported to be understood, or 'name-unknown', when the name of neither image was reported to be understood. The words to be tested were therefore selected individually for each infant on the basis of their parent's responses to the CDI. Test items were restricted to object names, as these form a large proportion of a child's vocabulary during the second year (Fenson *et al.*, 1994) and are easily depicted for test purposes.

Unlike studies requiring overt behavioural responses (e.g. Ring & Fenson, 2000), the preferential looking task makes minimal response demands, meaning that competence is unlikely to be underestimated due to a failure to elicit the required behaviour. To be certain that competence is not overestimated, the study must also be carefully designed to control for any spontaneous visual preferences. For, if infants prefer to fixate familiar objects, and if parents use their child's familiarity with an object as a heuristic when reporting comprehension, as Tomasello & Mervis (1994) suggested, then an item-level relationship between reported comprehension and preferential looking behaviour might be found that would tell us nothing about the validity of parents' vocabulary reports. If looking times differing from chance are to be attributed with confidence to comprehension of the word presented, all such potential confounds must be eliminated.

The studies reported here employed three such controls. First, in addition to the CDI, parents were asked to complete an Object Familiarity Questionnaire (adapted from Schafer's, 1998, Pointing Out Questionnaire). A word was selected for testing only if the parent reported that the infant regularly came into contact with an item similar to that used to depict the word. This control enabled us to match infants' level of familiarity with the images presented, both within each trial, and across known and unknown word trials. Second, by presenting each object as the named target and unnamed distracter equally often, any spontaneous preferences for one object over another would result in no overall preference. Finally, a preference for the target image was not on its own considered evidence of comprehension. Instead, the impact of hearing each word over and above infants' baseline preference was assessed by comparing looking times towards the two images during the 2·5 seconds before and after the word was heard (Swingley, Pinto & Fernald, 1998).

We report two experiments that explore whether words marked as understood or not understood on the Oxford CDI elicit different preferential looking responses from infants. Infants in Experiment 1a were aged 1;6, while those in Experiment 1b were aged 1;3, 1;6 and 1;9 and came from a second region of the UK. Our hypothesis in both cases was that reports of word knowledge would coincide with evidence of comprehension in the preferential looking task. That is, infants would show an increase in target preference on hearing known words but not on hearing unknown words.

EXPERIMENT 1A

METHOD

Participants

Thirty participants were recruited from a database of infants whose parents had volunteered to participate in developmental studies. Participants came from English-speaking households in the Oxford area of the UK and were in good health. One infant failed to complete the experiment due to fussiness and was excluded from all analyses. The remaining 29 infants were 15 boys and 14 girls with a mean age of 1;6.10 (range=1;5.19–1,10.13). The majority of infants were brought to the test session by their mothers, with a small number accompanied by fathers, grandparents or childminders. Questionnaires were completed by parents in all cases.

Materials

Parents were sent two questionnaires approximately two weeks before their visit, which they completed and returned by post before the test session. The Object Familiarity Questionnaire (OFQ) was adapted from Schafer's (1998) Pointing Out Questionnaire, a set of 80 prototypical black and white pictures of common objects, including animals, vehicles, toys, food, clothes, household objects and things found outside. Parents were asked to rate each picture according to how often their child came into contact with something similar to the object depicted. Response options included 'hardly ever' ('more than a month since last came into contact'), 'rarely' ('more than a week goes by between coming into contact'), 'sometimes' ('comes into contact more than once per week') and 'often' ('comes into contact more than once per day'). For the purposes of this study, items were classed as 'familiar' to the child if the parent gave the item either the 'sometimes' or 'often' rating.

Parents were also asked to complete the Oxford CDI[1] (Hamilton *et al.*, 2000), a list of 416 of the first words children use and understand. Parents were instructed to mark each word that they thought their child 'understood' or 'understood and produced'. If a word was marked in either of these ways, it was classed as 'known' for the purposes of this study. Items that were not marked were classed as 'unknown'. By comparing parents' responses to the OFQ and CDI it was possible to randomly select four familiar name-known objects and four familiar name-unknown objects for each child. These were randomly grouped into two pairs of known items and two pairs of unknown items, and were presented in these pairs in the preferential looking task.

The preferential looking task used the same set of images as the OFQ, but picture editing software was used to present each object as a 320×200 pixel, 256-colour image against a 5% grey background.

Tokens of the word *Look!* and the name of each object on the OFQ were produced by a female native English speaker using infant-directed speech, and digitally recorded at 22·05 kHz into signed, 16-bit files. Tokens were edited to remove head and tail clicks and background noise and matched for peak-to-peak amplitude. The word *Look!* was inserted into each naming sample in the format 'Look! … Object name!', so that the onset of the name was 2·5 s after the onset of the word *Look!*

[1] The Oxford Communicative Development Inventory (Hamilton *et al.*, 2000) can be downloaded from: http://www.psy.ox.ac.uk/babylab/cdi.html

Procedure

Testing took place in an infant language laboratory at the University of Oxford. On arrival, infants spent a few minutes playing with the experimenter and their parent. Infants were then seated on their parent's lap in a preferential looking booth approximately 1 m in front of two 17″ eye-level monitors. A loudspeaker situated centrally above the monitors delivered the auditory stimuli. A red light and buzzer mounted between the monitors were used to centre infants' attention between trials. Video cameras positioned above the two monitors recorded infants' looking behaviour.

The experiment was carried out in semi-darkness. The parent was asked to sit quietly with her eyes closed to ensure that she did not influence the infant's direction of gaze. Parents also listened to instructions over headphones; these prevented them from hearing the auditory stimuli and reminded them to keep their eyes closed and to ensure their child faced forwards throughout the study. The experimenter remained out of sight in an adjacent control room during testing.

The experiment consisted of 12 trials. Four trials tested comprehension of known words; each pair of name-known images was presented twice, so that each image was the target once and the distracter once. Eight trials tested comprehension of unknown words; each pair of name-unknown images was presented four times, so that each image acted as the target and distracter twice.[2] The first two trials were always known word trials and the order of the remaining 10 trials was randomized by the presentation software. The side of presentation of the target was counterbalanced for both trial types.

Each trial was initiated by the experimenter when the infant was facing forwards. Trials lasted for exactly 5 seconds, throughout which the two images were displayed. The word *Look!* began 100 ms after the onset of the trial and the target word began 2600 ms after the onset of the trial.

Scoring

Infants' looks towards each monitor were scored off-line using a button-press apparatus. Each trial was scored twice for looks towards the left monitor and twice for looks towards the right monitor. All scoring was blind to the trial type and the side of the target. Analysis software determined

[2] There were unequal numbers of 'known' and 'unknown' word trials because the study reported here formed the pretraining test phase of a word learning experiment. We assessed infants' understanding of known and unknown words before and after teaching them two of the four unknown words. Our primary concern was therefore to accurately ascertain any changes in infants' comprehension of the unknown words with training, and to this end we doubled the number of unknown word trials. Of course, our pre-training results made the rest of the study redundant.

whether each look scored by the experimenter fell before or after the onset of the target word. Looks that started prior to the onset of the target word and continued until after its onset were split into two looking times, one for the period prior to the onset of the target label and one for the period after its onset. The first and second sets of scored times were then averaged. Intra-scorer reliability was assessed for a random sample of 20% of infants ($N=6$). The mean Pearson's correlation between the first and second sets of scored times was $r=0.97$ (range: 0·95–0·99).

RESULTS AND DISCUSSION

The measure of target preference reported is the 'percentage of target fixation', the amount of time infants spent fixating the target image out of the total time they spent fixating either image (e.g. Reznick, 1990; Tincoff & Jusczyk, 1999; Houston-Price *et al.*, 2006). Target preference was calculated for the periods of each trial before and after the onset of the word. Data were averaged across known and unknown word trials, and checks for normality were satisfied. During the pre-naming trial phases, looking behaviour was expected to be at chance (means of 50%). When infants heard a known word, attention to the target image was expected to rise above 50%, and when unknown words were heard, looking behaviour was expected to remain at chance.

Mean percentage of target fixation before and after the target word was heard on each trial type can be seen in Figure 1. On both types of trial, infants looked randomly towards the two images before the words were heard, and longer at target images after their labels were heard. The data were entered into a three-way mixed ANOVA with two levels of 'word type' (known and unknown), two levels of 'onset of target word' (pre- and post-onset) and two levels of gender. There were no effects of gender. There was a main effect of onset of target word ($F(1,27)=11.27$, $p=0.002$, partial $\eta^2=0.30$); infants looked longer at the target after hearing it named than before (pre-naming: mean = 49·2%, $SD=4.6$; post-naming: mean = 54·3%, $SD=6.6$). No other main effects or interactions were found. Most notably, the expected interaction between word type and onset of target word fell well short of significance ($F(1,27)=0.03$, $p>0.05$, partial $\eta^2=0.00$). That is, infants increased their attention to the target on hearing both known ($t(28)=2.46$, $p=0.02$) and unknown words ($t(28)=2.46$, $p=0.02$).

Participants in this study showed understanding of words reported to be unknown as well as words reported to be known. One explanation of this finding is that their parents underestimated the number of object names infants understood when they completed the CDI. However, it is also possible that parents recorded their children's vocabulary accurately,

https://doi.org/10.1017/S0305000907008124 Published online by Cambridge University Press

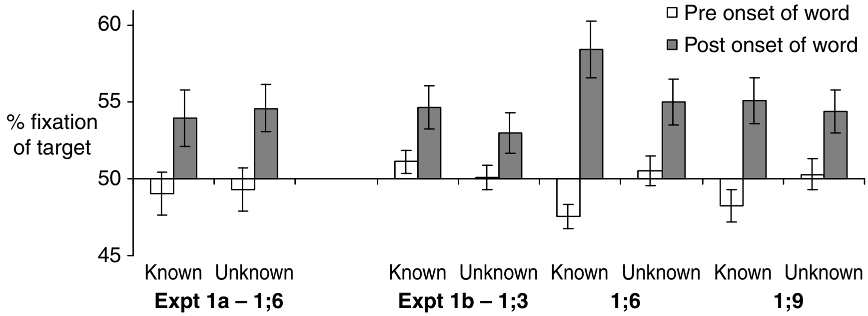

Fig. 1. Mean percentage of target fixation before and after hearing 'known' and 'unknown' words in Experiments 1a ($N=29$) and 1b, for infants aged 1;3 ($N=37$), 1;6 ($N=35$) and 1;9 ($N=41$).

and that infants learned some words marked as unknown between completion of the CDI and the laboratory task. Recall that parents returned the completed questionnaires by post prior to the test session, to allow the experimenter to tailor each infant's test stimuli to their parent's responses. As a result, testing took place on average 9·3 days ($SD=4{\cdot}9$) after CDI completion. We therefore investigated the possibility that infants' understanding of unknown words reflected post-CDI gains in vocabulary knowledge.

If such learning was a factor in infants' behaviour, one would expect to find a relationship between the length of the CDI-test delay and infants' understanding of the unknown words, as indexed by their mean increase in target fixation on hearing these words. No such relationship was found (Pearson's $r(29)=-0{\cdot}17$, $p>0{\cdot}05$). In the same way, when infants were split into groups according to whether their test delay exceeded the median of 9 days ($N=14$) or not ($N=15$), and whether their target preference increased on hearing unknown words ($N=19$) or not ($N=10$), a chi-squared analysis found no relationship between these two factors ($\chi^2(1)=2{\cdot}89$, $p>0{\cdot}05$). Finally, when infants with a longer than average test delay were excluded from the analysis, those who remained (mean delay = 5·6 days, $SD=2{\cdot}4$) still showed a significant naming effect for 'unknown' words ($t(14)=2{\cdot}18$, $p<0{\cdot}05$).

The data therefore fail to reveal any evidence that infants' understanding of unknown words was due to their learning of these words after the CDI had been completed. Regardless of the length of the CDI-test delay, infants showed a significant naming effect for words rated as unknown. Indeed, known and unknown words were indistinguishable on the basis of infants' behaviour in this study; infants showed equivalent increases in target preference on hearing both word types. The results of Experiment 1a therefore show that 'false negatives' are to be found in parents' vocabulary reports.

Experiment 1b explores whether such underestimation is specific to the parents who participated in this study, or common to parent reports in general. For, if a tendency to misreport is specific to certain parent–child groups, then clinical and cross-cultural uses of the CDI are cast into doubt; population differences in vocabulary scores might be attributable to group differences in report style rather than genuine differences in vocabulary size. Indeed, previous research has reported links between socioeconomic status and parent report style (Reznick, 1990; Fenson *et al.*, 1993; Arriaga *et al.*, 1998; Feldman *et al.*, 2000). Population differences aside, it also remains to be seen whether vocabulary underestimation is specific to the parents of infants aged 1;6 or a feature of parents' reports throughout the second year. At the time of the vocabulary explosion, when the number of words in an infant's vocabulary is increasing rapidly, parents might struggle to keep track of the extent of their child's word knowledge. Before or after this time, parents might be able to state which words their child understands with more accuracy. For these reasons, a larger-scale study was carried out to explore whether the underestimation found in Experiment 1a was: (a) a systematic feature of parental report that could be replicated in a second UK population; and (b) a phenomenon seen throughout the second year or characteristic of a specific stage of vocabulary growth.

EXPERIMENT 1B

The design and procedure of Experiment 1b were almost identical to those of Experiment 1a, except that the numbers of known and unknown words selected for testing were doubled, and the number of trials assessing comprehension of each word type was balanced. Only the methodological details that differed between the two studies are described below.

METHOD

Participants

One hundred and twenty-one infants aged 1;3, 1;6 or 1;9 were recruited from the University of Reading's infant database. Eight infants were excluded from analyses: six failed to complete sufficient trials due to fussiness (see 'Results' for completion criteria) and two infants' data were lost due to experimenter error. Remaining participants were 37 infants aged 1;3 (18 boys and 19 girls; mean age = 1;3.6; range = 1;2.11–1;3.29), 35 infants aged 1;6 (19 boys and 16 girls; mean age = 1;6.3; range = 1;5.9–1;6.27) and 41 infants aged 1;9 (23 boys and 18 girls; mean age = 1;9.9; range = 1;8.13–1;10.0).

https://doi.org/10.1017/S0305000907008124 Published online by Cambridge University Press

Materials

The Object Familiarity Questionnaire used in Experiment 1a was extended by adding a further 28 items that children might encounter on a regular basis, including items in two new categories: body parts and people. Prototypical images of these items were taken from children's television programmes or picture archives. Parents were asked to complete only the noun, verb and adjective sections of the Oxford CDI (Hamilton *et al.*, 2000) in this study. On the basis of parents' responses to the two questionnaires, eight familiar items whose names were reported to be known and eight familiar items whose names were reported to be unknown were randomly selected for each child. These were randomly grouped into four pairs of known items and four pairs of unknown items.

Procedure

Testing took place in an infant language laboratory at the University of Reading. The procedure was as for Experiment 1a, except that images were presented on the left and right sides of a single back-projection screen, rather than on two monitors. The experiment consisted of 32 trials, 16 testing comprehension of known words and 16 testing comprehension of unknown words. Each pair of images was presented four times, so that each image was the target twice and the distracter twice. Trials lasted for 5 seconds, with the word *Look!* and the object name beginning 100 ms and 2500 ms after the onset of the trial respectively.

Scoring

A split-image digital recording of infants' looks towards each side of the screen was scored frame-by-frame off-line using Observer 4·1. Two experimenters each coded approximately half of the video recordings; all coding was blind to trial type and target side. As a check for scorer reliability, a third experimenter coded a random 10% of the videos for each age group ($N=13$). Mean Cohen's kappa for the concordance between scorers' ratings of look direction was 0·94 (range = 0·90–0·97).

Results and discussion

Six infants failed to complete at least 16 trials (out of a possible 32) due to fussiness and their data were excluded. Infants of 1;3, 1;6 and 1;9 who satisfied this inclusion criterion completed an average of 30·7, 30·1 and 29·7 trials respectively. As in Experiment 1a, the percentage of target fixation was calculated for the period of each trial before and after the target word was heard, and the data satisfied normality criteria.

Mean percentages of target fixation for each age group on known word and unknown word trials, before and after the target word was heard, are shown in Figure 1. It can be seen that, in each age group, infants looked approximately randomly at the two images before target words were heard. On hearing both word types infants increased their looking towards target images. The data were entered into a four-way mixed ANOVA, with two levels of 'word type' (known and unknown), two levels of 'onset of target word' (pre- and post-naming), three age groups and two levels of gender. There were no effects of gender. A highly significant main effect of onset of target word was found ($F(1, 107)=56{\cdot}80$, $p<0{\cdot}001$, partial $\eta^2=0{\cdot}35$); infants looked longer at the target image after it had been named than before (pre-naming: mean target fixation = 49·6%, $SD=3{\cdot}5$; post-naming: mean = 55·1%, $SD=6{\cdot}9$). There were no other main effects but, unlike in Experiment 1a, an interaction between word type and onset of target word was found ($F(1, 107)=6{\cdot}41$, $p=0{\cdot}01$, partial $\eta^2=0{\cdot}06$). Infants showed greater increases in fixation of the target on hearing known words (mean increase = 7·0%, $SD=10{\cdot}4$) than they did on hearing unknown words (mean increase = 3·8%, $SD=10{\cdot}5$). However, it is worth noting that the effect size for the interaction is very small. In fact, paired sample t-tests revealed that infants showed highly significant increases in target fixation on hearing both known ($t(112)=7{\cdot}18$, $p<0{\cdot}001$) and unknown words (t $(112)=3{\cdot}87$, $p<0{\cdot}001$). Thus, while infants showed a greater tendency to fixate the target on trials presenting known words and their referents, they also demonstrated comprehension of words parents reported to be unknown.

There was no interaction between age group, word type and onset of target word ($F(2, 107)=1{\cdot}52$, $p>0{\cdot}05$, partial $\eta^2=0{\cdot}03$), demonstrating that a similar pattern of understanding of known and unknown words was found for each age group. However, as a primary aim of this study was to investigate whether the relationship between parental report and preferential looking behaviour was the same for children of different ages, we entered the data from each age group into a separate mixed ANOVA, each with two levels of 'word type', two levels of 'onset of target word' and two levels of gender. For infants aged 1;3, we found a significant overall effect of onset of word ($F(1, 35)=7{\cdot}74$, $p=0{\cdot}01$, partial $\eta^2=0{\cdot}18$), and no other main effects or interactions. For infants aged 1;6, a main effect of onset of word was found ($F(1, 33)=26{\cdot}22$, $p<0{\cdot}001$, partial $\eta^2=0{\cdot}44$), qualified by an interaction between word type and onset of word ($F(1, 33)=5{\cdot}29$, $p=0{\cdot}03$, partial $\eta^2=0{\cdot}14$). Infants in this group significantly increased their target preference on hearing both known ($t(34)=5{\cdot}71$, $p<0{\cdot}001$) and unknown words ($t(34)=2{\cdot}18$, $p=0{\cdot}04$), but the size of the increase was greater for known words. For infants aged 1;9, a main effect of onset of word was found ($F(1, 39)=24{\cdot}00$, $p<0{\cdot}001$, partial $\eta^2=0{\cdot}38$), with no other effects or interactions. Thus, the overall interaction between word type and onset of word was driven primarily by the behaviour of infants aged 1;6, and all three age groups increased their fixation of the target on hearing both 'known' and 'unknown' words.

As in Experiment 1a, there was a delay in this study between parents' completion of the CDI and the test date, to allow the experimenter to tailor each infant's test stimuli to reports of their word knowledge (all infants: mean = 10·4 days, $SD = 7{\cdot}2$; age 1;3: mean = 11·6, $SD = 8{\cdot}8$; age 1;6: mean = 11·0, $SD = 7{\cdot}4$; age 1;9: mean = 8·7, $SD = 5{\cdot}1$). Once again there was no relationship between the length of the CDI-test delay and infants' level of understanding of the unknown words, as indexed by their mean increase in target fixation on hearing these words, either for infants as a whole (Pearson's $r(109) = 0{\cdot}09$, $p > 0{\cdot}05$) or for any age group (1;3: $r(36) = 0{\cdot}00$, $p > 0{\cdot}05$; 1;6: $r(33) = 0{\cdot}11$, $p > 0{\cdot}05$; 1;9: $r(40) = 0{\cdot}21$, $p > 0{\cdot}05$). When infants were split into two groups according to whether their test delay exceeded the median for their age group ($N = 49$) or not ($N = 60$), and were categorized as increasing their target fixation on hearing unknown words ($N = 77$) or not ($N = 36$), a chi-squared analysis revealed no relationship between these two factors ($\chi^2(1) = 1{\cdot}27$, $p > 0{\cdot}05$). Finally, when those infants with a longer test gap than the median for the age group were excluded from analyses, those who remained (mean delay = 5·9, $SD = 3{\cdot}4$) still showed a significant increase in target preference on hearing unknown words ($t(59) = 2{\cdot}26$, $p = 0{\cdot}03$). Thus, infants' understanding of reportedly unknown words did not reflect the length of time that had passed between completion of the CDI and the test session.

Comparing the results of Experiments 1a and 1b

If parents are accurate reporters of their child's vocabulary knowledge, an interaction between word type and onset of word would be expected; infants should show an increase in target fixation on hearing known words but not on hearing unknown words. In Experiment 1a, no such interaction was found; infants showed equivalent and significant increases in target preference on hearing both word types, demonstrating that they understood both groups of words. In Experiment 1b, an interaction WAS found, particularly among infants aged 1;6, which might lead one might conclude that parents were reporting more accurately in this study. However, as can be seen in Figure 1, participants aged 1;6 behaved very similarly on unknown word trials in the two studies (Experiment 1a: mean increase in target fixation = 5·3%, $SD = 11{\cdot}6$; Experiment 1b: mean increase = 4·5%, $SD = 12{\cdot}2$; $t(62) = 0{\cdot}26$, $p > 0{\cdot}05$). The difference between the two studies' findings stems from infants' responses to known words: Infants aged 1;6 showed a much greater increase in target fixation on known word trials in Experiment 1b than in Experiment 1a (Experiment 1a: mean increase = 4·9%, $SD = 10{\cdot}8$;

Experiment 1b: mean increase = 10·9%, $SD = 11·3$, $t(62) = 2·14$, $p = 0·04$). In other words, the interaction between word type and onset of word seen in Experiment 1b was due to infants' particularly strong reaction to known words in that study, and not to their failure to respond to unknown words.

GENERAL DISCUSSION

In both studies reported here, infants showed comprehension of words their parents marked as unknown on the CDI as well as words they marked as known. That this finding was replicated in both genders, two different UK populations, and in three different age groups suggests that parental underestimation may be an inherent feature of British vocabulary reports.

A number of caveats must be made to this claim. First, we assessed comprehension only of object names in these studies; it is possible that reports for other word classes might coincide more closely with infants' behaviour in a laboratory test. It is important to establish the accuracy of parents' reports for word types other than nouns, as CDI measures are used to theorize about the make-up of early vocabularies in both typically-developing infants and those at risk of language impairment. If underestimation is specific to object names, the extent to which nouns predominate in infants' early vocabularies might be understated. However, preliminary data from our laboratory suggest a similar pattern for descriptive words: infants show a weaker target preference on hearing 'unknown' adjectives than 'known' adjectives, but an increase in target fixation on hearing both word types.

Second, the studies reported here assessed comprehension only for the names of objects with which infants were familiar, according to parents' reports. We expect that vocabulary reports would appear considerably more accurate if we selected test items from the full set of object names on the CDI, including those with which children have had no contact. For if a child has never encountered an object, they will be unlikely to demonstrate understanding of its label in the laboratory. However, while such a finding would appear to vindicate parents' 'no' responses on the CDI, infants' behaviour would rather reflect parents' ability to accurately identify the items the infant has not yet encountered. We therefore chose to control object familiarity in order to specifically target parental reports of children's word understanding. Of course, our control depends on parents providing accurate reports of their child's experience with the items depicted on the OFQ, but there are obvious reasons why an infant's exposure to an object should be more easily observed by parents than the infant's understanding of the object's label.

Third, the design of our study would have allowed infants who knew the name of one of the two objects presented on an 'unknown' word trial to use a principle of mutual exclusivity (Markman, 1989; Merriman & Bowman, 1989) or N3C (novel-name-nameless-category; Golinkoff, Mervis & Hirsh-Pasek, 1994) to identify the appropriate referent on trials presenting a label that was accurately described as unknown. That is, if infants knew the name of one of the two objects presented on 'unknown' word trials, they might assume that a second word must refer to the alternative object, leading them to fixate the alternative for longer, despite having no knowledge of the link between the label and referent concerned. The age at which the mutual exclusivity bias is employed has been the subject of considerable debate. Merriman & Bowman's (1989) studies suggest that the bias is not in place until around 2;6, while Markman, Wasow & Hansen (2003) interpret the findings of their study as evidence for its existence at 1;3 to 1;5. Results appear to very much depend on the precise method used to test for the presence of the bias. The only study to use a preferential looking method to explore the age at which the bias appears is that reported by Halberda (2003). Halberda found that when infants were shown one familiar and one novel object, infants of 1;5 (but not 1;4) increased their looking towards the novel object on hearing a novel label. If infants made similar use of the mutual exclusivity principle in our studies, which used a similar experimental procedure, our measure of infants' comprehension of unknown words could be inflated.

However, because 'unknown' words were always presented as a pair in our studies, the mutual exclusivity bias could have been applied by infants only if they understood the meaning of one word in each pair. It is therefore not possible for our finding of comprehension of unknown words to be 'explained away' by this principle. Furthermore, our youngest group also showed understanding of words reported to be unknown, at an age at which Halberda's (2003) findings suggest mutual exclusivity is not applied within this paradigm. Thus, while we do not claim that infants in our studies must have understood all the 'unknown' words tested (or indeed, all the 'known' words, for the same principle applies on known word trials), we do argue that understanding must have been present for a significant number of words in this category.

Finally, our findings do not speak to the accuracy of parents' estimates of infants' total vocabulary scores or to the appropriateness of using these scores as a global index of a child's language level. As discussed in the Introduction, parents might be accurate estimators of their children's total vocabulary size even if there are inaccuracies in their reports for individual items. The implications of our studies are therefore primarily for the use of instruments such as the CDI to identify which specific words or classes of words are in a child's comprehension vocabulary. To some extent, this reconciles our findings with previous claims that reports are good indicators of children's vocabulary knowledge (Bates *et al.*, 1988; Dale *et al.*, 1989).

Such claims rest largely on findings of correlations between total CDI scores and infants' performance on standardized vocabulary tests, and as we explained earlier, these correlations need not reflect the accuracy of parents' reports for individual words.

Previous research comparing reports at the item level to behavioural measures of comprehension has also found underestimation, at least in the latter stages of the second year (Robinson *et al.*, 2000). However, others have claimed that parents' reports are more at risk of overestimation than underestimation, especially at younger ages (Tomasello & Mervis, 1994). It is possible that among different populations, both over- and under-reporting may be found. For example, systematic differences might exist in the reporting styles of parents from different countries: North American parents may be more prone to exaggerate their child's level of language development, as Tomasello & Mervis claim, while British parents may be more likely to understate their child's knowledge. This hypothesis is supported by Hamilton *et al.*'s (2000) comparison of the vocabulary development of children from England and the US. Hamilton *et al.* found significantly lower comprehension and production scores on the Oxford CDI throughout the second year compared to those reported in Fenson *et al.*'s (1994) norming study. While it is feasible that the language acquisition of British children is delayed in comparison to children in the US, our results suggest an alternative explanation: that report styles differ between the two countries.

The finding that some parents, at least, are overly cautious in their reporting of comprehension has implications for the development and use of instruments such as the CDI. Ring & Fenson (2000) suggest that a comprehension scale should be incorporated into toddler versions of the CDI in order to assess the validity of receptive vocabulary reports beyond the middle of the second year. Indeed, some researchers already use the CDI/*Words and Gestures* form to measure comprehension beyond 1;4 (e.g. Stennes, Burch, Sen & Bauer, 2005), while others have created adaptations that assess receptive vocabulary throughout the 0;8–2;6 age range (Hamilton *et al.*, 2000). The studies reported here suggest that underestimation of comprehension is a feature of parent report throughout the second year, including at 1;3, when receptive vocabulary reports are widely collected.

Replication of our findings among further populations is therefore needed, and we suggest that the preferential looking method provides an ideal tool for such investigations. Our studies add to the growing literature demonstrating the paradigm as a means of detecting quite subtle differences in word knowledge. In Experiment 1b, comprehension of both known and unknown words was evident, but the stronger effect on known word trials suggests that the paradigm is sensitive to differing levels of understanding

https://doi.org/10.1017/S0305000907008124 Published online by Cambridge University Press

of different sets of words. But, can we be sure that longer looking on hearing certain words reflects 'better comprehension' of those words? Or indeed that preferential looking measures comprehension at all? These important questions require consideration.

The fundamental basis of the Intermodal Preferential Looking paradigm is that systematically longer fixation on a named image can be interpreted as evidence of comprehension. But is it necessarily comprehension that the paradigm is detecting? If the study is well-designed and carefully balanced, we would argue that the answer to this question is 'yes'. A target preference that is significantly different from chance can only result from the activation of the associative link between a word and its referent. This associative link is at the heart of word comprehension – knowing which phonological forms 'go with' which meanings. While it is true that everyday language uses a rich lexical understanding that is not required by the looking paradigm, it is also true that every real-life interaction also involves the activation of these same associative links between words and meanings. Thus, preferential looking captures a process that is fundamental to every instance of comprehension.

Can we be similarly certain that a significantly greater target fixation in this paradigm reflects 'better comprehension' of the words concerned? If lexical understanding is not absolute, but cumulative with experience, it seems feasible that an infant might demonstrate differing degrees of comprehension of words in the lengths of their target fixations. Alternatively, a greater preference for the target in one condition might reflect more consistent looking behaviour in that condition; for example, an infant might understand more of the words in that group. Third, a larger target preference might result from more individual infants in the group understanding the words heard. Our finding that infants in Experiment 1b showed greater target fixation on hearing known words than unknown words could therefore indicate EITHER that infants better understood each of the known words OR that infants understood more of the known words OR that more infants understood the known words. Because we have only a limited understanding of the factors that cause differences between conditions in this paradigm, we cannot tease these accounts apart. As a consequence, definitive claims about comprehension in this paradigm must rely on significant differences from a baseline measure or chance. In our studies, infants significantly increased their target preference from baseline on hearing both known and unknown words, and this was true at all ages, for both boys and girls, and in two distinct populations. We can therefore state with confidence that infants had some understanding of both sets of words.

Why, then, is the preferential looking task more sensitive to a child's understanding of a word than parents? We consider three possible reasons.

First, parents might fail to notice their child's word understanding because they have not had the opportunity to observe their child interacting with the item(s) concerned; the relevant interactions may have occurred at daycare, for example. However, we selected items for testing only if parents reported them to be familiar to the child, and such reports imply that parents have seen their child observe or interact with the item. It is also worth noting that participants in Experiment 1b spent an average of only 1·2 days per week in daycare, with a median of 0 days at all three ages. That is, the majority of participants spent every day with their parents. Moreover, we found no relationship between the time infants spent in daycare and their understanding of the words reported to be unknown. It therefore seems unlikely that parents misreport comprehension because they are not present when their child learns the word.

Second, parents might simply be poorer detectors of infants' comprehension. Preferential looking provides an easy opportunity for a child to demonstrate their acquisition of a new word; the child is merely required to distinguish the word's referent from among two potential referents. The laboratory also provides ideal conditions for researchers to detect understanding; the tools available allow detailed, off-line examination of changes in infants' fixation durations towards target and distracter images. The resources available to parents are considerably less sophisticated. Parents presumably base their comprehension judgments on a broad range of behavioural indicators including gaze duration, but these behaviours may not be closely or obviously tied to the infant's understanding of the language used around them, and may provide parents with a poor heuristic.

Alternatively, parents might be sensitive to their infant's understanding of a word but set their threshold for reporting comprehension at a different level. For example, while a preferential looking study requires infants to discriminate the referent from a single, randomly-selected distracter (e.g. duck vs. table), parents might require their infant to discriminate the same referent from a set of taxonomically-related items (e.g. duck vs. goose vs. swan) before reporting comprehension. Which is more appropriate? The criterion that parents should use when completing the CDI is not clear; the instructions provide no advice about how parents should determine whether their child's behaviour indicates comprehension. It is therefore not surprising that parents find it a difficult task to report on comprehension. Several parents who participated in our studies described their difficulty in the space provided on the CDI for additional comments. A typical comment was provided by the mother of an infant aged 1;3, who wrote 'I am not sure of words X understands. I have ticked the ones most likely!' But perhaps the most telling comment came from the mother of an infant aged 1;9, who wrote 'I think X understands all these words. I may be a little overoptimistic but he responds well to requests and instructions.' Despite this mother's beliefs about her son's word knowledge, she left many words unmarked. If parental caution, rather than insensitivity, explains their under-reporting, ways of encouraging parents to overcome this tendency deserve exploration. The addition of a few instructions to those currently provided might be sufficient to reduce the number of false negatives. Alternatively, a three- or four-point scale incorporating frontier words as a response option might be needed (Robinson *et al.*, 2000), to allow parents to indicate that their child has heard a word before and might understand it.

In conclusion, our studies suggest that, throughout the second year, infants understand some words that are not reported to be understood on the CDI, as well as words that are reported to be understood. We think that the level of word comprehension in our studies is all the more noteworthy considering that our laboratory task required infants to recognize and understand words when spoken by an unfamiliar speaker and in a different context to any previous experiences the child might have had with the words. The implications are clear. When it matters that the contents of infants' vocabulary are measured accurately, the measurement instrument must provide infants with the appropriate opportunity to express their understanding.

REFERENCES

Arriaga, R. I., Fenson, L., Cronan, T. & Pethick, S. J. (1998). Scores on the MacArthur Communicative Development Inventory of children from low- and middle-income families. *Applied Psycholinguistics* **19**(2), 209–223.

Bates, E., Bretherton, I. & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge: Cambridge University Press.

Behrend, D. A. (1988). Overextensions in early language comprehension: Evidence from a signal detection approach. *Journal of Child Language* **15**, 63–75.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development* **10**, 159–99.

Charman, T., Drew, A., Baird, C. & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Developmental Inventory (Infant form). *Journal of Child Language* **30**, 213–36.

Dale, P. S., Bates, E., Reznick, J. S. & Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. *Journal of Child Language* **16**, 239–51.

Evans, M. A. & Wodar, S. (1997). Maternal sensitivity to vocabulary development in specific language-impaired and language-normal preschoolers. *Applied Psycholinguistics* **18**, 243–56.

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E. & Paradise, J. L. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development* **76**(4), 856–68.

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. & Paradise, J. L. (2000). Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Development* **71**(2), 310–22.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J. & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, Serial no. 242, **59**(5), 1–173.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S. & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing Group.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S. & Reznick, J. S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics* **21**, 95–116.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M. & Gordon, L. (1987). The eyes have it: lexical and syntactic comprehension in a new paradigm. *Journal of Child Language* **14**, 23–45.

Golinkoff, R. M., Mervis, C. B. & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language* **21**(1), 125–55.

Graham, S. A. & Poulin-Dubois, D. (1999). Infants' reliance on shape to generalize novel labels to animate and inanimate objects. *Journal of Child Language* **26**(2), 295–320.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition* **87**(1), B23–B34.

Hamilton, A., Plunkett, K. & Schafer, G. (2000). Infant vocabulary development assessed with a British Communicative Development Inventory. *Journal of Child Language* **27**, 689–705.

Hirsh-Pasek, K. & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm reveals emergent language comprehension. In D. McDaniel, C. McKee & H. Cairns (eds), *Methods for assessing children's syntax*, 105–124. Cambridge, MA: MIT Press.

Houston-Price, C., Plunkett, K. & Duffy, H. (2006). The use of social and salience cues in early word learning. *Journal of Experimental Child Psychology* **95**(1), 27–55.

Houston-Price, C., Plunkett, K. & Harris, P. (2005). 'Word-learning wizardry' at 1;6. *Journal of Child Language* **32**, 175–89.

Maital, S., Dromi, E., Sagi, A. & Bornstein, M. (2000). The Hebrew Communicative Development Inventory: language specific properties and cross-linguistic generalizations. *Journal of Child Language* **27**, 43–67.

Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.

Markman, E. M., Wasow, J. L. & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology* **47**(3), 241–75.

Merriman, W. E. & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, Serial no. 220, **54**(3–4).

Mills, D. L., Coffey-Corina, S. A. & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience* **5**(3), 317–34.

Mills, D. L., Plunkett, K., Prat, C. & Schafer, G. (2005). Watching the infant brain learn words: Effects of vocabulary size and experience. *Cognitive Development* **20**, 19–31.

Molfese, D. (1990). Auditory evoked responses recorded from 16-month-old human infants to words they did and did not know. *Brain and Language* **38**, 345–63.

Reese, E. & Read, S. (2000). Predictive validity of the New Zealand MacArthur Communicative Development Inventory: Words and Sentences. *Journal of Child Language* **27**(2), 255–66.

Reznick, J. S. (1990). Visual preference as a test of infant word comprehension. *Applied Psycholinguistics* **11**, 145–66.

Ring, E. D. & Fenson, L. (2000). The correspondence between parent report and child performance for receptive and expressive vocabulary beyond infancy. *First Language* **20**, 141–59.

Robinson, C. W., Shore, W. J., Hull Smith, P. & Martinelli, L. (2000). Developmental differences in language comprehension: What 22-month-olds know when their parents are not sure. Poster presented at the International Conference on Infant Studies, Brighton, July, 2000.

Schafer, G. (1998). Word learning in infancy. Unpublished doctoral thesis, University of Oxford.

Schafer, G. & Plunkett, K. (1998). Rapid word learning by 15-month-olds under tightly controlled conditions. *Child Development* **69**(2), 309–320.

Stennes, L. M., Burch, M. M., Sen, M. G. & Bauer, P. J. (2005). A longitudinal study of gendered vocabulary and communicative action in young children. *Developmental Psychology* **41**, 75–88.

Swingley, D., Pinto, J. P. & Fernald, A. (1998). Assessing the speed and accuracy of word recognition in infants. In C. Rovee-Collier, L. P. Lipsitt & H. Hayne (eds), *Advances in infancy research, Vol. 12*, 257–77. Stamford, CT: Ablex.

Thomas, D. G., Campos, J. J., Shucard, D. W., Ramsay, D. S. & Shucard, J. (1981). Semantic comprehension in infancy: A signal detection analysis. *Child Development* **52**, 798–803.

Tincoff, R. & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science* **10**(2), 172–5.

Tomasello, M. & Mervis, C. B. (1994). The instrument is great, but measuring comprehension is still a problem. Commentary on Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994. *Monographs of the Society for Research in Child Development* **59**, 174–9.

Weitzner-Lin, B. (1996). Assessing communicative and linguistic development through parent and teacher report. *Infant Toddler Intervention* **6**(3), 247–52.

Yoder, P. J., Warren, S. F. & Biggar, H. A. (1997). Stability of maternal reports of lexical comprehension in very young children with developmental delays. *American Journal of Speech-Language Pathology* **6**, 59–64.