Word learning in preschoolers:
Are bilingual 3-year-olds less guided by mutual exclusivity than their monolingual counterparts?

Madeleine Campbell

Supervisors: Barbora Skarabela and Monica Tamariz

Master of Science in Developmental Linguistics

School of Philosophy, Psychology and Language Sciences

University of Edinburgh

2007

Abstract

A fundamental question in developmental linguistics and developmental psychology is how young children learn new words. While some researchers suggest that words are primarily learned through experience, others argue that the acquisition process is guided by innate lexical biases. One of the most widely studied biases is the Mutual Exclusivity Bias (ME), which describes children's preference for just one label per concept. The disambiguation effect in ME has been demonstrated extensively with ostensive paradigms requiring young monolingual children to choose between familiar and novel labels in identifying unfamiliar objects. However, evidence for ME within languages in bilingual children is mixed. In the present study, a productive naming paradigm was used to assess 3-year-olds' tendency to adopt novel labels for familiar items (a variant on Merriman and Bowman's (1989) rejection/correction effect)**.** Five monolingual and 5 bilingual children aged 2;11-3;6 were tested in English. Following a training session when the experimenter applied novel labels to 3 of 12 pictures of familiar objects, the children played two successive naming games. The first game involved further reinforcement of the novel labels by the experimenter while the second game did not. In the first game, the bilingual children adopted novel labels more frequently (*Mdn*=.40) than the monolingual children (*Mdn*=.13) and Mann-Whitney's (one-tailed exact) U=3.0, was significant, $p<0.05$ with a large effect size (r=-.63). In contrast, only one bilingual produced a novel label in the second game. Measures of receptiveness in the training session displayed asymmetries between production and comprehension. Overall the results suggest that experience of two languages plays an important role in learning novel labels. The findings are consistent with an account of ME as a heuristic learned from monolingual input, the application of which varies in bilingual preschoolers according to both ambient language and socio-pragmatic context. The results are discussed in the context of what insights can be gained from possible extensions to the experiment.

Acknowledgements

This study follows a pilot research project funded by a small research grant from the University of Edinburgh in 2006 and awarded to Barbora Skarabela, entitled 'Early word learning in bilingual children: Mutual exclusivity as perspective taking or linguistic bias?'.

I am grateful to my supervisors, Barbora Skarabela and Monica Tamariz, for their invaluable feedback and guidance in project conception, data analysis and interpretation of the results.

I would also like to thank Frances Wilson for her patient and friendly help and support in the provision of materials, formulating games, co-piloting the procedure and countless fruitful discussions in the Linguistics common room. My thanks also to Holly Branigan, who advised on appropriate experimental design for the target age group in the initial stages of the pilot project.

This project would not have been possible without the interest and support shown by parents and nursery staff, nor the cheerful if somewhat puzzled cooperation of the preschoolers who took part in the study, to which I am especially indebted.

Finally, my thanks to all the Lecturers and fellow students at the School of Philosophy, Psychology and Language Sciences, University of Edinburgh, whose bright and lively debates have been an inspiration throughout the Master's course.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text. This work has not been submitted for any other degree or professional qualification except as specified.

Madeleine Campbell

Table of Contents

Page

Table of Contents
(continued)

Page

List of Figures

List of Tables

Word Count

15,778

(pp i & 1-56, excludes references and appendices)

Chapter 1

Introduction

From birth to the end of their pre-school years, children face the formidable task of acquiring one or more languages without explicit tuition yet typically developing preschoolers normally accomplish this feat with remarkable ease. Infants utter their first words around 12 months. The rate of learning increases dramatically from one or two per week to one or two words a day when infants reach around 50 words at about 18 months. Between the ages of 2 and 6 this rate increases to an average of 10 a day (Clark, 1993). Children face three key challenges in learning words: They need to learn that there is a relationship between sounds and meaning, to identify words in a stream of speech and to identify what these words designate in the world – the problem of referential indeterminacy (Quine, 1960). This dissertation is concerned with the third of these challenges.

Explanations of word learning can be broadly described as either based on experience or on linguistic mechanisms which are thought by some to be innate. Amongst those who believe word-learning is essentially based on experience, word learning is described in terms of cognitive, exemplar-based or socio-pragmatic processes modulated by attentional factors, which are thought to interact to varying degrees. Those who adhere to the school of innate linguistic mechanisms subscribe to a language acquisition model based on lexical biases which guide the developing child to apply ME through a process of induction and elimination. In contrast, input or exemplar-based theory favours emergentist processes whereby children learn from the input in a monolingual environment to apply the 'one-word-one meaning' rule. Whether innatist or emergentist, both schools of thought, concede that word learning evolves in the context of experience, aided for example by syntactic cues to distinguish between word classes There is evidence to suggest that ME is operational in monolingual infants from as early as 12 months. However, investigations of ME in children ranging from 2 to 6 years have yielded mixed results. While some studies suggest that ME increases with age, others find the opposite. Importantly, recent studies suggest that cognitive factors and use of socio-pragmatic cues such as the ability to take the perspective of others, may impact the developmental path of ME.

One way to investigate different accounts of early word learning is to compare the behaviour and developmental paths of monolingual versus bilingual children. Previous studies indicate that bilingual children do not tend to apply ME across languages. It is less clear, however, whether ME is applied within each individual language in a bilingual child.

The aim of the present study is thus to provide a comparative account of ME in mono- and bilingual 3-year-olds, using a productive naming paradigm to investigate to what extent they are willing to apply two labels to one concept within a language.

Chapter 2

Literature Review

2.1 The Theoretical Framework

Theoretical accounts of early word learning have put forward linguistic, input-based and socio-pragmatic accounts to explain the phenomenon of mutual exclusivity. However most explanations, whether linguistic or based on experience, acknowledge the need for a pluralistic account while focussing on specific theoretical stances and, as we shall see below, the theoretical framework does not divide up categorically into one or the other school of thought.

2.1.1 Linguistic mechanisms

The ME hypothesis postulates that objects, concepts or categories have only one name and that labels have only one referent, that is, that there is a one-to-one mapping between labels and referents (Markman, 1989). Although accounts of this linguistic mechanism have evolved since its original formulation and interpretations vary, broadly 4 phenomena have been taken as indication of ME: two in respect of novel labels, disambiguation and rejection and two in respect of familiar labels: correction and restriction (Merriman & Stevenson, 1997). The disambiguation task provides the most commonly observed manifestation of ME: when children are faced with two objects, one familiar (e.g. a cup) and one novel (e.g. a melon scoop), they tend to assume that a novel label (e.g. a *dax*), refers to the novel item. Further strategies adopted by children when faced with ambiguity involve narrowing of the category, either through correction (the child applies a novel label to a known object and no longer uses the familiar label) or rejection (the child resists a novel label and continues to use a familiar label). The restriction effect describes children's tendency to avoid generalizing a familiar name to atypical exemplars. Evidence for the correction/restriction effect in respect of familiar labels is more equivocal than the disambiguation effect. This is due in part to widely differing methodologies, but also perhaps to different mechanisms operating in respect of novel versus familiar labels.

Proponents of mutual exclusivity (ME) have postulated that children must build on other biases in order to choose between possible referents when faced with novel labels. For example, Markman (1992) proposed that two initial *constraints* (Markman's terminology) guided children's attention when mapping words to referents: words denote whole objects ('the whole object assumption') and labels refer to categories of objects rather than to individual objects: 'the taxonomic constraint' (see Markman & Hutchinson, 1984). When picking out attributes to assist in categorization, Landau, Smith and Jones (1988) further found that children tend to assume labels refer to shape rather than material, colour, size or texture. Markman (1992) outlined several advantages to assuming ME, which would allow children to avoid redundant hypotheses about the meanings of categories and to infer referents without the need to have them pointed out. In addition, if children were guided only by the whole object assumption, they would never learn words for parts or attributes of objects. Thus when faced with a novel label for a familiar referent, ME allows children to learn terms for parts or substances (Markman & Wachtel, 1988). Mutual exclusivity can also help children narrow the overextensions characteristic of word learning in toddlers (Merriman & Bowman, 1989). Thus, explanations of ME as an innate constraint or bias are by necessity scaffolded upon a number of further lexical and perceptual biases which may or may not be innate, to guide children in identifying the target referent.

2.1.2 Linguistic Mechanisms: Constraints or biases?

Nelson (1988) flagged up the theoretical implications that constraints entail an all-or-none response, while linguistic mechanisms and principles pertain more to comparative preferences, strategies or biases. Although this distinction between constraints and biases is not used consistently in the research literature, most proponents of innate mechanisms agree that they are tendencies rather than absolute markers of binary mechanisms and this is echoed in the statistical methods applied to empirical studies of ME, which often compare observed data to chance performance rather than predict responses approaching 100% occurrence. Merriman and Bowman (1989) acknowledge this as a natural imperative of word learning: "The [ME] bias cannot have absolute strength; if it did, people would never learn the correct relations between words that are supposed to share referents" (p. 7). One disadvantage of all-or-nothing ME, is that it would impede the ability to learn hierarchies beyond the basic level, yet preschoolers do learn super- and subordinate terms. Markman (1992) surmises that violations of the ME bias do not constitute evidence that the bias does not exist. Rather, the ME assumption, in the same way as the whole object assumption and taxonomic assumption, serves to support a good 'first guess', and is applied in the context of environmental cues.

However observations of the ME phenomenon *per se* disclose nothing about its putative origin and the evidence and methodology for investigating linguistic mechanisms has been the subject of much controversy (see for example Woodward and Markman, 1991). Proponents of linguistic biases have surmised that ME may be supported by a process of induction. Ishida, Kosugi and Itakura (2003), for example, suggested that preschoolers used the taxonomic bias and ME in combination, to infer the referent of a novel label**.** More recently Halberda (2006), using eye tracking experiments, showed that both adults and preschoolers applied 'disjunctive syllogism' in disambiguation tests, that is, a process of elimination, looking first to familiar items in order to infer the referents of novel labels. On the other hand, Gathercole (1987), Nelson (1988), Deák (2000), MacWhinney (2005) and others, concluded that there was little or no evidence for internal constraints, biases or principles. Rather, the ME phenomenon could be the result a learned heuristic, perspective-taking, or a combination of these factors.

2.1.3 Learning from the input

MacWhinney (1989), for example, proposed that children's word learning "mirrored the statistical configuration of the input" (p. 126) and argued for an exemplar-based theory of word learning, in which a child's initial categories are based directly on experience of labels for objects in these categories. MacWhinney thus saw ME strategies essentially as a process of competition between words, whereby repeated associations make the child's mapping of a label with its referent category grow in strength and compete successfully with other labels for the same category. Explanations based on the configuration of the input have also been put forward by innatists, who suggested that children build on lexical biases by using morphosyntactic cues to distinguish between word classes (Pinker, 1984). For example Macnamara (1982) noted that articles such as 'a, the' denote nouns, while suffixes such as 'ed' denote verbs and Gathercole (1997) noted that objects are typically denoted in English by count nouns and preceded by an article while mass nouns occur in the singular (e.g. 'rice', not 'rices'), suggesting the application of ME to whole objects in the first instance may be guided by the nature of the ambient language. Gathercole, Thomas and Evans (2000) found that English, Welsh and Spanish speaking preschoolers as young as 2 years old responded to novel labels in ways that corresponded to the structure of their different languages, and contended that this finding undermined the notion that universal biases like ME could guide early word learning.

More recently, Monaghan, Christiansen and Chater (in press) postulated that lexical co-occurrence statistics and phonological properties could aid the process of acquisition by providing cues to word class and syntactic categories, based on cross-linguistic corpora of between-adult speech in the presence of children and child-directed speech. Of specific interest to ME, 98% of English nouns were classified correctly according to distributional cues (p. 33) compared with 58.6% according to phonological cues, unlike verbs where both types of cues offered similar predictive values (between 82% and 88%). These findings suggest that distributional cues may help to distinguish between actions and objects as referents, though lexical co-occurrence alone cannot explain how, having surmised that a label refers to a noun, children choose amongst possible referents in the world. This process would need to be guided either by the lexical and perceptual biases invoked by Markman (1992) and others, or by other experience-based mechanisms. For example, language simulations have shown that agents can infer the meaning of signals with a high degree of communicative success when presented with target-signal pairs in multiple contexts (Smith, 2003, 2005), suggesting that statistical co-occurrence coupled with cross-situational consistency is in principle enough to learn one-to-one signal meaning pairings. Houston-Price, Plunkett and Duffy (2006), using a preferential looking paradigm, found that 15-month-old infants could use a combination of digitally controlled social and perceptual cues to determine word-meaning, but that learning could occur based purely on consistency of perceived referent attributes across situations without the support of direction of gaze. However, the relative contributions of associative learning and socio-pragmatic cues in experience-based accounts present a complex web of interrelated factors, a fundamental issue to which we shall return after reviewing the main tenets of socio-pragmatic accounts of word learning.

2.1.4 Socio-pragmatic explanations of word learning

While children as young as 14-months-old appear to learn words even in the absence of cues regarding the speaker's intentions (see for example, Werker, Cohen, Lloyd, Stager & Cassosola, 1998), Tomasello (1999, 2003) argues that intention-reading skills, coupled with pattern-finding skills, are general prerequisites for children to learn the communicative conventions of their culture. Children's ability to engage in joint attention with their interlocutor emerges between 9 and 12 months. According to Tomasello, the development of their language is then based on their involvement in joint attentional frames with adults, which involve sensitivity to the speaker in a number of ways: Studies by Tomasello (1988) and Baldwin (1991, 1993) have suggested that children use the speaker's direction of gaze (SDG) to choose amongst likely referents when hearing a speaker's utterances. Children have also been shown to be sensitive to speaker's intentions, assumptions and perspectives to infer the meaning of words (Diesendruck, 2005; Diesendruck and Markson, 2001). In addition infants as young as 18 months-old have been shown to be sensitive to emotional and behavioural expressions when gauging whether communicative intent is achieved (Tomasello, Strosberg and Akhtar, 1996). According to this socio-pragmatic account, then, ME is not an innate bias and word learning is primarily a function of perspective-taking.

2.1.5 ME as a product of experience or linguistic bias?

From this review of apparently conflicting accounts it would be tempting to adopt one or the other theoretical stance at the risk of oversimplifying the issue. Whether ME is interpreted as a product of experience or the result of an innate linguistic bias may depend on implicit methodological assumptions or isolated developmental snapshots. Arguing that there is evidence and counter-evidence for linguistic, socio-pragmatic and associative accounts, Hirsh-Pasek, Golinkoff, Hennon and McGuire (1994) offered an explanatory platform that takes a developmental perspective whereby different theories take prominence at different stages in the child's acquisition of words. For example, object novelty might play a more decisive role in early infancy, while a socio-pragmatic cue such as monitoring the speaker's direction of gaze is more likely to develop later as part of the child's ToM. Deák (2000) proposed the need for "an explanatory framework that is sufficiently rich to capture the flexibility and diversity of word learning" (p. 30). MacWhinney (2005) noted that 'the shift in recent years from reliance on constraints to an emphasis on cues has opened up space for considering the time course of the processes supporting initial mapping' (p. 200). According to this explanatory framework, in contrast with Quine's (1960) hypothetical linguists, children are immersed in a varied and dynamic socio-pragmatic context, with abundant cues taken from the perspectives of both the hearer and the speaker that serve to guide appropriate mappings between words and referents.

One way to elucidate the relative contributions of experience and linguistic biases in guiding word learning is to compare monolingual and bilingual children within a developmental framework while taking into account the empirical evidence gleaned from monolingual samples to date. The bulk of theories on word learning rely on research performed in a monolingual environment. It is therefore possible that ME is a construct arising from research in a monolingual environment, with the implication that ME may be a learned heuristic which is largely a function of monolingual experience. The key contributing factors to this experience may be purely attributable to a comparatively higher exposure to one-to-one mappings of labels and referent in monolingual children, or to other cognitive and socio-pragmatic factors associated with interacting in just one language. In the following section, therefore, we shall compare experimental evidence to date based on different methodological approaches and discuss implications in the context of experience-based explanations.

2.2 Review of Experimental Evidence

2.2.1 The developmental path of ME in monolingual preschoolers

Some researchers have found evidence of ME well before the vocabulary spurt, while others found that it does not become apparent until 2-years-old and increases with age in preschoolers. In disambiguation tests, Markman, Wasow and Hansen (2003) found that ME started to operate in 15- to 17-month olds, as did Halberda (2003) while Xu, Cote and Baker (2005) found similar evidence in 12-month old infants. Clark (1987) argued that, once children acquire the ability to categorize hierarchies, they no longer apply ME but continue to rely on the principle of contrast, while Markman and Wachtel

(1988) postulated that the ME bias may weaken with time but continues to operate into adulthood.

In contrast Merriman and Bowman (1989) found that ME emerged some time after 25 months, with older children disambiguating more than the younger children. Further, several disambiguation experiments in monolingual children suggested that ME increases with age (Merriman, Marazita & Jarvis, 1993, in 4-year-olds; Merriman & Schuster, 1991, comparing 2 and 4-year olds) while others, for example Deák, Yen and Pettit (2001), found no age differences in disambiguation between 3- and 4-year-olds. Merriman and Bowman (1989) found that correction and restriction effects were weaker than disambiguation effects and, though they also showed a tendency to increase with age, the three 3 exhibited 'developmental asynchrony' (p. 78), while Merriman and Stevenson (1997) found a significant restriction effect in 24-25 month-olds' tendency to avoid generalizing a familiar name to atypical exemplars. However Woodward and Markman (1991) noted that the same tests in different age groups may yield results that are not comparable, arguing that Merriman and Bowman's (1989) attempts to control for the novelty effect led to habituation, which 'overcompensated' for novelty in 2-year-olds, but not in older children. Further, if it is evident before 2-years of age, as Markman, Wasow and Hansen (2003) and others have suggested, then it is possible that ME may be indicative of a different underlying process than its manifestation in older children.

2.2.2 Comparing ME in monolingual and bilingual children

The discrepancies in developmental evidence for ME in monolingual children are further complicated by comparisons with bilingual children. Au and Glusman reported that bilingual English/Spanish pre-school children (age 3;4 to 5;7) seemed able to accept lexical overlap across languages when they had no doubt that the labels came from different languages, while Diesendruck (2005) found that acceptance of lexical overlap across languages related to children's expectations of different referential intents in 3-5-year-old Hebrew-speaking and English-Hebrew bilinguals.

Further, the acquisition of cross-language equivalents early in the lexicons of bilinguals has been demonstrated extensively and in consistent proportions, even in cross-modal settings (Deuchar and Quay, 2000; Pearson, Fernández, and Oller, 1995; Holowka, Brosseau-Lapré and Petitto, 2002).

It is less clear, however, whether bilingual children readily adopt synonyms within languages in a monolingual discourse context and there is scant empirical evidence regarding ME within languages in bilingual preschoolers. Davidson, Jergovic, Imami and Theodos (1997) showed that 5-6-year old monolinguals applied ME in disambiguation tests significantly more than younger mono- and bilinguals (3-4-year-olds), while there were no significant age differences in the bilingual group of English/Greek and English/Urdu speaking children. Interestingly, older bilingual children also applied ME in the disambiguation test (that is, significantly above chance), but 'less so' than their monolingual peers.

In a further restriction test, Davidson et al. (1997) found a complex pattern of results (as did Merriman and Bowman, 1989), which made it difficult to draw categorical conclusions for these more complex indicators of ME. Their restriction test consisted of two sets of 5 hybrid pictures (fork/knife or fish/bird hybrids). After being told that the name of the first hybrid (for example, a picture of a knife-fork) was a *knife*, children were then asked whether 2 of the remaining hybrids were knifes (that is, the name given by the experimenter) and whether the other 2 were *forks*. Rejection was measured by the children's propensity to reject the other whole object label (in this case the *fork*), as measured by 'no' answers when asked 'is this a(n) X?'. Acceptance was measured by the children's propensity to answer 'yes' to the name originally given by the experimenter. Using a yes/no paradigm, Davidson et al. did find a significant result in that bilingual children rejected the 'other' label less often than both younger and older monolingual children. Yet age, but not language, was found to be a significant factor in how readily the children accepted the given (restricted) names for the hybrid pictures, with older children showing greater acceptance.

In another study, Davidson and Tell (2005) noted Davidson et al.'s (1997) findings that older bilingual children were in fact less likely than their younger counterparts to apply ME in tests of disambiguation and rejection, in contrast with some findings in monolingual children that ME becomes more pronounced with age. They compared monolingual (English-speaking) and bilingual (English-Urdu speaking) children aged 3 and 6-years-old. When they assessed children's tendency to attribute novel labels for familiar objects to parts of the same objects, the older monolingual group was more likely than the bilingual group to select the spare part when given a new name for a familiar object; while overall bilingual children in both age groups showed less preference for selecting spare parts. Thus at least from age 3 onwards, monolingual children applied ME to a greater extent than bilingual children in respect of the whole object assumption, suggesting that experience of two languages may impact children's word learning, and in particular their application of ME.

2.2.3 The developmental path of ME: methodological considerations

As we have seen, empirical research has yielded discrepant results about the onset and developmental path of ME in both mono- and bilingual children and a central question about ME remains the timing of its activation (Merriman and Bowman, 1989). These inconsistencies may stem from the variation in methodological approaches, in particular from the difference in indicators of ME measured. For example, the possibility that novelty guides children's choices is a potential confound in disambiguation tests. Merriman, Marazita and Jarvis (1993, 1995) coined the term 'feeling of novelty' (FN) to underline their findings that it was not the new name, but the feeling that an object was new to the child, that led the disambiguation effect. Another possible explanation of the novelty effect put forward by Golinkoff, Hirsh-Pasek, Bailey and Wenger (1992), is that new names are for categories with no names so far – the novel-name, nameless category principle (N3C). Shifting the focus from an child-centric view to considering the child's interaction with the people around her and the assumptions she makes about their intentions, Akhtar, Carpenter and Tomasello (1996) proposed that discourse novelty also guides word learning and demonstrated that 24-month-olds were able to discern the intended referent of a label from a knowledge of which referent was new to the discourse context. Thus, although Merriman and Bowman's (1989) disambiguation test was originally invoked as evidence of the ME bias, disambiguation has been shown to be a function of novelty of object or discourse rather than name, thereby undermining the disambiguation effect as a reliable indicator of naming strategy.

2.2.4 Does comprehension mirror production in tests of ME?

Another methodological consideration is that most studies of ME have involved ostensive paradigms ('show me the X'; 'can this be called an X?') rather than naming paradigms ('say X'; 'tell me what this is'). That is, the variables measured in tests of disambiguation, correction/rejection and restriction, and those involving features or whole *versus* parts of objects, have been based on how children selected or responded to questions about target items. Subjects were generally not asked to produce novel labels.

Underlining this important shortcoming, Piccin and Blewitt (2007) used a naming paradigm to compare the performance of preschoolers aged 3;2–3;11 under two naming conditions. In the shared label condition, two puppets knew both novel labels for a novel object, while in the distinct label condition, one puppet knew one label and the other puppet knew the other label. Children in the distinct label conditions produced two labels more often than the children in the shared label condition in both a verbal and tone-based task, suggesting that their tendency to apply ME was influenced by general cognitive resources. Echoing the research literature on ME, Piccin and Blewitt predicted that children in the shared label condition would show evidence of learning just one label, while children in the distinct label condition would learn both labels. A follow-up comprehension test involved both the identification of previously learned novel label items *wug* and *fam* and a disambiguation task using a novel item labelled *dax*. The children were presented with 3 novel artefacts, plus the novel object (a sticker) for which they had previously learned the two novel labels. They were then asked to point to the *wug*, the *fam* and the *dax*. To their surprise, Piccin and Blewitt found that only 5 of 12 children in the distinct label condition associated at least one previously learned label with the sticker. That is, although they were able to produce two labels in order to achieve their communicative goals in the preceding test of production, the children in the distinct label condition did not achieve a bidirectional mapping between label and referent, implying a dissociation between the processes of comprehension and production. Further, synonymy (as evidenced in the distinct label condition in the production task) 'co-existed' with a strong disambiguation effect, with all 12 children in the distinct label condition successfully identifying the *dax* as one of the novel artefacts in the comprehension task. In other words the ostensive disambiguation effect, coupled with the productive use of multiple labels, implied an asymmetry in children's application of ME.

In another naming experiment, Deák, Yen and Pettit (2001) tested 3- and 4-year olds with objects that had the appearance or function of other objects (e.g. a dinosaur-shaped crayon). Subjects were taught 3 novel terms for these hybrid objects then prompted to produce more than one label either in super- or subordinate categories, for example: 'what kind of thing is a dinosaur?'; or by demonstrating the object's function – thus producing an effect they described as 'polynomy'. Four-year-olds produced significantly more novel words per object than 3-year-olds and learning a new word reduced the number of familiar words produced per referent, which Deák et al described as a 'partial correction effect'. Thus polynomy in word learning situations was shown to be present in 3-year-olds and to increase with age, which was contrary to the one-to-one mapping hypothesis, at least in respect of multiple categories or different category levels and this echoed Piccin and Blewitt's (2007) findings regarding co-existence of synonymy and disambiguation.

It is generally attested that comprehension precedes production on a developmental timescale in the acquisition of the lexicon (see for example, Clark 1993) and an intuitive assumption would expect this learning sequence to be replicated in the micro-timescale of an experiment. Yet the methodological approaches adopted in many studies of ME do not state explicitly how their measures of receptiveness translate to an interpretation that word learning has taken place. This shortcoming is compounded by widely different approaches to training and testing. At one end of the spectrum, Halberda and Goldman (submitted) have shown that 2-year-olds can learn new words 'in three seconds flat' from a single ambiguous exposure by testing fast-mapping in a series of discrete disambiguation tasks. Similarly, Houston-Price, Plunkett and Duffy (2006) measured infants' responses in a test phase which came before the training phase for each trial. In contrast, other empirical studies have used extensive training prior to testing comprehension (for example, Merriman and Kutlesic, 1993), thereby perhaps obscuring any initial effects attributable to ME. Consequently it is possible that discrepant results in ME research, in addition to being influenced by inherent methodological differences between experiments, are due in part to implicit assumptions that word learning has taken place. The findings reviewed here suggest that evidence for ME is based primarily on ostensive paradigms and has not been shown to be matched by production.

2.2.5 The developmental path of ME: possible explanations

While there is conflicting evidence as to the onset of ME and whether it increases or decreases with age and this is attributable in part to the methodological considerations just reviewed, there does seem to be a developmental trend and a difference between mono- and bilingual children, suggesting that the latter are less guided by ME than monolinguals of the same age. Alternative experience-based explanations are reviewed below, which lead to different comparative and developmental predictions for the application of ME.

Heuristics

A mechanisms based on learning from input, possibly modulated by processing requirements, offers one alternative explanation for ME. Davidson and Tell (2005) surmised that "learning more than one language may affect children's use of language-learning principles, particularly those principles relying on mutual exclusivity" (p. 41) and noted the greater flexibility displayed by bilingual children, due probably to greater exposure to instances where these constraints do not apply. Exposure to cross-language equivalents may lead to bilingual children developing a general word learning heuristic that maps multiple labels with one referent. The more specific heuristic associated with ME, that is, the mapping of one label with one referent would in effect be a subset of the general one. Monolingual children, in contrast, would not have the opportunity to develop the multiple-mapping heuristic. If this were the case, the prediction would be that ME would also emerge within languages in bilingual children, but possibly with less strength than in monolingual children.

Vocabulary size

The predictions of a heuristics-based explanation contrast with MacWhinney (1989) when he surmised that ME should become more useful as a child's vocabulary size increases. According to this line of reasoning, ME should be stronger in bilingual children, whose vocabulary is potentially double that of monolingual children, although this interpretation may depend on whether bilingual children operate within one or two lexical systems, and this remains an open question (for a review see, for example, Deuchar and Quay, 2001, p. 63). Interestingly, in Deák, Yen and Pettit's (2001) experiment (see Section 2.2.4), the size of children's receptive vocabulary was found to be a predictor of number of words produced per referent in monolinguals, accounting for some one-third of the variance in polynomy. Such findings would not seem to support MacWhinney's contention, at least in terms of production.

Resource conservation

Looking at ME from a processing perspective, Liittschwager and Markman (1994) suggested that 24-month-olds learned more second labels for familiar items than 16-month-olds, probably because of a greater processing capacity. Similarly, Piccin and Blewitt's (2007) account predicts that "if there is no apparent communicative benefit to learning a new word […children…] will conserve resources and avoid doing so" (p. 20). If it could be assumed that word learning tasks place equivalent cognitive demands on mono- and bilingual children, a processing account would predict that there should be no difference between them in the application of ME.

Socio-pragmatic experience

The different socio-pragmatic experiences relating to language in mono- and bilingual children offer another, possibly complementary, explanation. While monolingual children have an expectation of dealing with their interlocutors from the same (linguistic) perspective as their own across most situations, bilingual children learn from a young age to distinguish between interlocutors and situations in making appropriate language choices. Importantly, Deuchar and Quay (2000) placed the adoption of cross-language equivalents around 1;0, which is well before their ability to match the linguistic context as defined by interlocutor or location at age 1;7-1;8 (Deuchar and Quay, 1999). According to their findings, then, an early form of contextual perspective-taking and the absence of the ME bias *across* languages, did not emerge at the same time. However, it is possible that this early ability in bilinguals to switch languages according to context does confer, in turn, an early ability to apply synonyms *within* a language, facilitated by taking another's perspective *across* languages within the broader framework of the child's developing theory of mind (ToM).

A further developmental milestone in ToM emerges between 3- and 4-years of age in preschoolers' reported ability to pass false belief (FB) tests. This benchmark was investigated by Goetz (2003), who noted that the ranges of Mandarin-speaking children passing the False Belief test were comparable to data reported for Western cultures of 35% for 3-year-olds, 69% for 4-year-olds, and 85% for 5-year-olds (Goetz, 2003, p. 3, citing Lee, Olson and Torrance, 1999). Goetz found that 4-year olds performed significantly better than 3-year olds using a ToM score based on 4 tasks: an appearance-reality test, a level 2 perspective-taking task, a false belief unexpected contents task and an unexpected transfer task. Goetz found a bilingual advantage in English/Mandarin speaking 4-year olds[1].

Using the findings of Deuchar and Quay (2000) as developmental benchmarks for an early manifestation of ToM, if cross-language perspective-taking alone can impact the development of ME within languages, then a lesser tendency to apply ME should be evident in bilinguals as early as 20 months of age. Indeed, Goetz (2003) postulated that a possible explanation of the bilingual advantage in FB could be a greater sensitivity to interlocutor's language, noting that this is evident in children between 1 and 2-years old (Lanza, 1992; Genesee, Boivin and Paradis, 1996; both as cited in Goetz, 2003).

A further implication of Goetz' (2003) findings on FB would be that ME, whether innate or learned, should guide 3-year-old monolingual children until they develop the ability to pass a false belief test, but that the ME effect should then become more flexible in relation to the socio-pragmatic context; while bilingual children of the same age should display a lesser tendency to apply ME. However with the exception of a study by Doherty and Perner (2005), which found a correlation between ability to monitor and produce known synonyms and ability to pass a false belief test, there is a

---

[1] Though it should be noted that the between-group difference with English and Mandarin speaking monolinguals only approached significance when the test was administered a second time, a finding worthy of further investigation

paucity of empirical data drawing direct developmental comparisons between ME and ToM.

Chapter 3

The Research Question

3.1 Problem Statement

In summary, if ME is an innate and absolute linguistic constraint, there should be no difference between monolingual and bilingual children: ME should be evident in both groups of children in the context of novel words within a language. If, however, there is less evidence of ME in bilingual children, this would suggest that perspective-taking and/or experience in two different languages play a role in word learning. The difference could simply be due to greater exposure to cross-language equivalents, which would lead to a general rule for 'several words-one meaning', a multiple-mapping heuristic which may in turn facilitate learning synonyms within a language; or more generally to a greater amount of practice in word learning, modulated perhaps by differential processing loads and vocabulary size. Alternatively or perhaps in addition, ME could be influenced by the ability to take another's perspective, or to a child's theory of mind. If monolingual preschoolers apply the ME bias to a greater extent than bilinguals of the same age, it is possible that this developmental difference correlates with an earlier realization of ToM in bilingual children.

According to experience-based interpretations, then, mono- and bilingual children in the same age group, *ceteris paribus*, should differ in their treatment of novel labels. Monolingual children should be less willing to apply novel labels to familiar items than their bilingual counterparts. The present study aimed to test this prediction by looking at a cross-section of children within a tight developmental window, namely 3-year-old preschoolers. Specifically, the study sought to establish any differences between mono- and bilingual children within the same age group in their willingness to use two labels for one thing.

Further, if ME were a function of input and/or socio-pragmatic interaction rather than age, covariance with such experience-based factors could in turn explain disparate results concerning the developmental path of ME. If a bilingual advantage were found in 3-year-olds, who are on the threshold of the key ToM developmental milestone of FB, this would then provide grounds for further research, outwith the scope of this project, to disentangle the relative contributions of heuristics, early perspective-taking and the later manifestations of ToM associated with FB.

3.2 Proposed Approach

A productive paradigm to measure ME

As we have seen, most basic tests of ME to date have utilized ostensive paradigms. Further, novelty confounds have undermined the disambiguation effect as an indicator of naming strategy. Restriction tests are also prone to confounds, since they lend themselves to interpretations of multiple category levels or multiple categories because they involve either hybrid objects (as used by Merriman and Bowman, 1989; Davidson, Jergovic, Imami and Theodos, 1997) or objects with conflicting appearance and function (as used by Deák, Yen and Pettit, 2001).

In the present study, the aim was to assess children's acceptance and production of novel labels for unambiguous familiar items using a proactive indicator of ME, which we termed adoption. It will be recalled that correction occurs when the child applies a novel label to a known object and no longer uses the previous (familiar) label and rejection occurs when the child resists the novel label and continues to use the previously known (familiar) label. Adoption occurs when the child actively adopts a novel label for a familiar object and produces it intermittently with its familiar name (this is what happens with synonyms). That is, the child sometimes applies correction (of the familiar label) and at other times applies rejection (of the novel label). Thus the adoption effect is a combination of Merriman and Bowman's (1989) rejection/correction effect, with the important difference that it is tested in terms of production, not comprehension. Further, like other ME effects, the adoption effect is primarily relevant as a comparative tool, with measures of its relative strength between groups indicating their relative tendency to apply ME.

Adoption is comparable to the partial correction effect observed by Deák et al. (2001), whereby learning a new word reduced the number of familiar words produced for a referent. However, unlike Deák et al.'s polynyms, in the present experiment children were taught novel labels for familiar items, that is, the adoption effect related to their ability to learn and produce novel second labels or synonyms for familiar items of the same category. Further, while adoption is seen as a proactive word learning strategy, rejection is seen as a more passive fall-back strategy, and it is predicted that adoption, rather than rejection, is likely to be indicative of different word learning mechanisms in mono- and bilingual children.

Finally, in light of the methodological shortcomings discussed previously, the present study measured both receptive and productive behaviour on first exposure to target labels during the training phase as well as productive behaviour during the test phase, in order to elucidate further the extent to which production mirrors comprehension.

3.3 Hypotheses and Predictions

Main working hypothesis

The hypotheses to be tested for each measure of ME, whether measures of receptiveness or production, were as follows:

H_0 : There is no difference in mono- and bilingual 3-year olds' ability to use two labels for one thing within a language
H_1 : 3-year old bilinguals are more willing to use two labels for one thing within a language than their monolingual counterparts

Comparative indicators of ME

It was predicted that if children's word learning was guided by experience rather than linguistic mechanisms, monolingual children should be more guided by ME than bilingual children. Based on a rigorous interpretation of ME as a one-to-one bidirectional mapping between label and referent, that is, one in which comprehension mirrors production, comparative indicators of ME would be as follows: if children are guided by ME, they should show a tendency to produce standard labels instead of novel labels. Further, comprehension should mirror production, in other words children should show a reluctance to match novel labels to familiar items, a poor ability to identify novel label items, and a tendency to correct the experimenter's use of novel labels. On the other hand, if bilingual children are comparatively less guided by ME than monolingual children, they should be more willing to produce novel labels instead of standard labels for familiar items at least some of the time (the adoption effect). In terms of receptive behaviour, bilingual children should be more willing to match novel labels to familiar items, better able to identify novel labels items by pointing to them, and less disposed to correcting the experimenter's use of novel labels.

A more flexible interpretation of ME

In line with the findings of Deàk et al. (2001) and Piccin and Blewitt (2007), and contrary to the rigorous interpretation of ME outlined above, we further hypothesized a more flexible interpretation of ME as one that can change over time in the course of an experiment, and one in which productive and receptive behaviour may be indicative of different underlying mechanisms.

Chapter 4

Method

4.1 Materials

A laminated sheet (42.0 cm by 30.5 cm) was prepared with large colourful pictures depicting 12 familiar inanimate items together with sets of cards (6.0 cm x 6.5 cm) depicting the same objects (see colour pictures in Appendix I and II).

Three of the 12 pictures were given novel labels in the training and production tasks. The novel label items (a pink sock, a green chair and a blue drum) were named '*mana*, *poggle* and *wug*' respectively. These were pronounced as British English phonemes with the characteristic stress on the first syllable for the disyllabic words and preceded by the indefinite or definite articles 'a' or 'the'.

4.2 Pilot Study

An initial pilot study[2] was undertaken with 3 monolingual and 3 bilingual children (age range 2;0 to 4;3) in order to test the feasibility of the experiment, which also served to streamline experimenter behaviour. The flower picture originally used for a novel label item was changed to a chair because of a potential confound of the novel label as a superordinate for the term flower. Changes were made to shorten the Training Tasks, as the children seemed to get bored and this increased the likelihood of non-completion of the Production Tasks. Further, a test was introduced to indicate whether the children effectively associated the novel labels with the corresponding pictures (the mapping task described in Section 4.4.1). The production tasks revealed that children produced novel or standard labels but also null responses and other types of response. This led to altering the original binary coding system for these tasks to accommodate the complexity of child responses with the aim of shedding further light on their labelling behaviour.

[2] This pilot was part of a research project funded by a small research grant from the University of Edinburgh in 2006 and awarded to Barbora Skarabela

4.3 Recording

Sessions were recorded with a digital audio-video recorder and a separate digital audio-recorder. It was sometimes necessary to compare the videotaped audio file with the audio recorded file to decipher child utterances. Coding involved categorizing utterances for each game according to the coding system described below.

4.4 Experimental Design and Coding

Pre-test (to test knowledge of standard labels)

The aim of the Pre-test was to establish that children knew the standard labels in English and also that the bilingual children knew the labels for these items in their other language.

The bilingual children were first asked by their parents to label 6 items on the laminated sheet and to point to the other 6 in their parent's native (non-English) language. Children in both groups (5 BL, 5ML) were then asked in English by the experimenter to label 6 items, including the 3 novel label items, with the question 'can you tell me what this is?'. The experimenter then asked children to point to the other 6 standard label items by saying: 'can you show me the X?'). The experimenter was careful to avoid producing either standard or novel labels for the novel label items in the pre-test (see order of items in Appendix III). Having thus established the English language mode for the bilingual children in the pre-test, all subsequent tasks were performed in English.

4.4.1 Training Tasks

The aim of the Training Tasks was to introduce the subjects to the 3 novel label items and test their understanding of the new mappings by establishing whether they would match a card with the corresponding novel label item on the chart when asked to do so. A subsidiary aim was to evaluate the children's receptiveness to the novel labels by logging their spontaneous behaviour in terms of repetitions and contradictions. All children were exposed to the same 3 novel label/item pairs. The procedure is described below.

Matching Task

Children were asked to place cards on their corresponding pictures on the laminated sheet. The experimenter labelled each card 3 times by saying: ‘I have an X, this is an X, can you put the card on the X?’ The experimenter thus named 9 standard and 3 novel labels items.

Mapping Task

Children were then asked, for each of the 3 novel label items, to point to their corresponding pictures, for example: ‘can you show me the *wug*?’ Responses were categorized as follows:

(a) Point to correct novel label item
(b) Point to standard label item
(c) Other (including no pointing, or wrong novel label item, or any combination of items)

The number of instances when the child pointed to the correct novel label item was quantified as a proportion of total responses (a/(a+b+c)).

Contradictions

This variable measured the number of spontaneous contradictions produced by children when novel labels were initially introduced. Responses in this category included: ‘it’s not a *poggle*’; ‘No, this is a *drum*’; and ‘it’s a *drum*’. Given that the experimenter consistently introduced the same number of novel label items to all children, the total number of contradictions were then compared as raw scores across groups.

Repetitions

As it was observed that children sometimes repeated labels, we counted the number of times they repeated the labels during the Training Tasks as follows:

(a) standard label used (including contradictions of the type: 'that's a *drum'*)
(b) novel label used (excluding contradictions of the type: 'that's *not a poggle'*)

Repetition was quantified as the number of instances when the child spontaneously repeated either the novel label or standard label item as a proportion of total number of repetitions (a/(a+b) and b/(a+b)).

4.4.2 Production Tasks (Memory and Snap games)

The production tasks sought to establish whether and to what extent the children would produce novel labels in answer to a question after hearing the novel label in the Memory game and without further reinforcement in the Snap game.

For the Memory game, the experimenter put down and labelled two cards, asked the child to label them, then turned one over – the child was asked to guess which card was turned over (if the child could not guess, she was encouraged to look at the card that had been turned over). For each card held by the experimenter, the experimenter said saying 'I've got a(n) X' before the child was asked 'what is it?' or a similar question. The question was asked again after a card was turned over: 'what did you turn over?', with the further prompt 'what was it?' if necessary. Then the roles were reversed, with the experimenter prompting the child with 'what do you have?' and 'what was it?'. The child and the experimenter had 14 cards each, 3 depicting novel label items and 11 depicting standard label items. The cards were pre-ordered so that the 3 novel label cards were first produced by the experimenter.

The second productive task was a traditional game of Snap. The experimenter pack contained only standard label items (14 cards), while the child's pack contained both standard label and novel label items (8 and 6 respectively). The child was asked to name each of the cards in her pack, if necessary prompted by the question: 'what have you got?'. Novel labels were not produced by the experimenter unless requested by the child. When two cards matched, the child was enthusiastically encouraged to say 'snap' and won the round.

The order of cards listed in Appendix IV, was not always preserved as the children sometimes dropped their packs. Further, the number of cards actually played for each game depended on individuals' enthusiasm and other distractions. Hence the number of opportunities to produce a novel label differed across subjects and a range of responses was possible. In order to reflect this variability while maintaining data comparability, coding involved categorizing children's responses as follows:

Productive Responses in answer to 'What is it?' (novel label items only)

This analysis compared the use of novel and standard labels to refer to a novel label item after an eliciting question such as: 'what is it?' or equivalent. Children also sometimes produced standard or novel labels as part of ambiguous answers, indicating a degree of uncertainty about the labels they used. Examples from bilingual children included: 'it's *like* a chair'; or, in answer to the experimenter asking, following a null response, 'what do you want to call it?', the child's response was: 'I *want to call it* a chair'. Such responses, which were obviously not 'null responses', were categorized as 'other' productive responses.

Productive responses were categorized as follows:

(a) standard label used
(b) novel label used
(c) other response (other than null responses[3])

The adoption rate **(x)** was quantified as the number of instances the child produced a correct novel label as a proportion of his/her total responses (x=(b/(a+b+c))). The rejection rate **(y)** was the number of instances the child produced a standard label as a proportion of his/her total responses (y=(a/(a+b+c))).

[3] The reasons categorizing null responses separately are given in the description of the Null Responses analysis in the following paragraph

Null Responses in answer to 'What is it?'

This analysis compared whether the children said nothing or produced no label when asked 'what is it' in the standard and novel label item conditions. Examples included either silence, or responses such as: 'I can't say it' or 'laughter', which could indicate an inhibition effect; or responses from monolinguals such as 'that one', which could indicate memory deficit. None of these responses gave an objective indication of which type of label, if any, was under consideration by the child. Such null responses were excluded from total productive responses and analyzed separately, as it would not be meaningful to interpret silences or avoidance strategies in the same way as responses which clearly involved the production of either a standard or novel label. While the latter, even in the context of an ambiguous answer, are positive indicators of a productive strategy, silences may be due to lack of interest, attentional or memory factors or to avoidance of naming due to confusion. Given that it was not possible to establish which of these explanations lay behind null responses, their inclusion in the total number of productive responses would have confounded the results regarding productive indicators of ME. Nevertheless, a comparison of null responses between the two groups may indicate the extent to which these other factors, taken as a whole, may differ between the two groups.

Coding categories were as follows:

(a) Total number of questions asked about standard items
(b) Number of null responses to questions about standard items.
(c) Total number of questions asked about novel items
(d) Number of null responses to questions about novel items.

The dependent variable was the number of null responses divided by the total number of questions for each condition (b/a; and d/c).

4.5 Experimenter Utterances

The number of novel label items produced by the experimenter was counted for each game. A two-way mixed ANOVA found a task effect but confirmed that there were no differences between groups. Overall, the mean number of novel labels produced by the experimenter during the Training Tasks was higher at 14.5 than in the Memory Game (*M*=5.5*)* and the number of novel labels produced in the Snap game was the lowest (*M*=0.5). Mauchly's Test of Sphericity was not significant (*p*=.75) and so the assumption of sphericity was met. Levene's test of equality of variance was also not found to be significant for each within-subject condition, so homogeneity of variance was assumed (*p*=.65, .13, and .10 for the 3 Tasks). In conformance with the experimental design, there was a within-subject contrast across the tasks, with F=108.923, (*df*=1,6)*,* $p<.001$. There was no interaction between language status and task, $F(1,6)=.92$, $p>0.05$, indicating that there was no difference between bilingual and monolingual groups in their exposure to novel label items.

Chapter 5

Results and Statistical Analyses

5.1 The Sample and Pre-Test

The main experiment included 10 preschoolers aged between 2;11-3;6, with 5 monolingual and 5 bilingual children. All 10 children completed the Training Tasks and the Memory Game. Only 8 children, 4 in each group, completed the Snap game. Sessions lasted between 10 and 20 minutes depending on the child's need for breaks.

Each group comprised 4 boys and one girl. The bilingual group included two Italian/English, two Danish/English and one German/English speaking preschooler. According to their parents, all bilingual children were exposed to both languages from birth. The original sample included 14 subjects. One child in the monolingual group was dropped due to poor audio quality, one for non-completion and another who was later found to be too young. A Chinese/English subject was dropped because he didn't recognize the word 'drum' (one of the novel label items) in English.

Three bilingual children were identified by word-of-mouth in the local bilingual community and tested in a quiet room in their homes in the presence of their non English-speaking parent. The two English/Danish children were interviewed at a Danish playgroup in the presence of their Danish-speaking parent. The monolingual children were tested in Scottish nurseries in open plan settings in the presence of peers and nursery staff. Signed consent was obtained from the parents. Children received a small gift for participating. All tests were undertaken by the same female experimenter and author of this paper.

Pre-test

Nine children passed the pre-test by indicating that they could name or point to at least 10 of the 12 items in both languages if bilingual and in English if monolingual, including all 3 of the items intended to be named with novel labels during the

Productive Tasks. One Danish/English bilingual (K.) did not complete the pre-test, but his mother confirmed that he knew all the items in Danish and English and so he was included in the analysis. The threshold for proceeding was set at 10 known labels (including all novel label items) because the picture of the spoon was sometimes interpreted as a fork (due perhaps to the vertical black lines inside the drawing (see Appendix I) and the apple was sometimes misinterpreted as a tomato, but it was clear that the child knew the names for these once the meaning of the pictures was clarified.

5.2 Statistical Analyses of Data

Non-parametric tests were undertaken in most cases for the following reasons:

1) A small sample size (4-5) participants per group, extensive ranges and/or the presence of one or two outliers for each test resulted in skewed distributions where the means were not representative. Non-parametric tests transformed the proportions into ranks, which afforded a way to be faithful to the data while using a test that was sensitive to the main differences under consideration, that is, the relative behaviours of one group compared to the other.

2) Levene's test of homogeneity of variance was significant for some of the ANOVAs carried out originally, for example in the analysis of repetitions and null responses for the memory game.

For the sake of comparability of findings across the entire experiment, the Mann-Whitney test was performed on all child responses for planned between group-comparisons while the Wilcoxon test was performed for repeated measures in the repetitions and null responses analyses. Exact tests were selected because of the small sample sizes. The median and range were taken as a statistical measure of central tendency. As medians only reflect rank ordering, their corresponding means are also quoted in tabular form to provide a cross-check of the internal consistency of reported data. In addition, where relevant, a more intuitive indication of children's behaviour is given by plotting the ranked individual data points in each group.

Statistical validity of results

According to Field (2005, p. 533), it is not possible to estimate the power of non-parametric tests, as power is a function of the probability of a Type I error – that is, of accepting that there is an effect when there isn't. However, power is estimated on the assumption that the underlying data distribution is normal and this, by definition, is not generally the case in non-parametric tests. Nevertheless, we can gain some indication of the effect size *r* for a significant effect in favour of the main hypothesis by using the z-score for Mann-Whitney's *U* and calculating *r*=z/square root of N and hence this has been quoted for the results of the non-parametric tests. According to Rosenthal (1991, p. 19 as cited in Field, 2005, p. 532), the threshold for a medium effect is .3 and for a large effect is .5.

The corollary is that with non-parametric tests there is only an increased chance of Type II error, that is, of accepting the null hypothesis when in fact there is an effect, if the data are normally distributed. As the data were not normally distributed for the non-parametric analyses, it is reasonable to accept the validity of those results that were found not to be significant. Nevertheless, overall the between-group analyses should be treated with caution because of the small sample size.

5.3 Production Task (Memory Game)

5.3.1 Adoption and Rejection

- **The bilingual children adopted (produced) more novel labels (*Mdn*=.40) than the monolingual children (*Mdn*=.13), $p<0.05$**
- **There was no significant statistical difference between groups in their tendency to reject the novel labels, though there was a slightly greater rejection rate in the monolingual children**

	Bilinguals	Monolinguals
Adoption (Novel label produced)	.47/.40(.56)	.20/.13(.63)
Rejection (Standard label produced)	.42/.50(.56)	.61/.57(.63)
Other	.11/.10(.22)	.19/.13(.43)

Table 1. Memory Game: Mean/median (range) of responses as proportion of total productive responses

Table 1 shows that bilingual children produced correct novel labels (*Mdn*=.40) as well as standard labels (*Mdn*=.50) and 'other' responses (*Mdn*=.10). In comparison, the monolingual children showed a slightly higher rejection effect (*Mdn*=.57) but a lower adoption effect (*Mdn*=.13) and comparable numbers of 'other' responses (*Mdn*=.13). None of the bilinguals and only one monolingual produced an incorrect novel label, using the word *mana* to refer to the *wug* (drum).

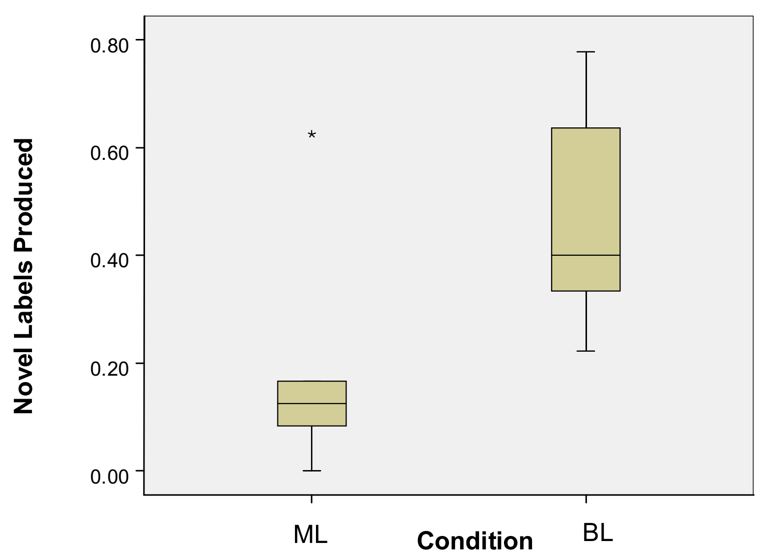


Figure 1. Memory Game Adoption Rate: Proportions of novel labels produced in response to 'what is this?'

When the adoption rate was compared between the two groups (see Figure 1), Mann-Whitney's (one-tailed exact) U=3.0, was significant, $p<0.05$, z=-1.98 with a large effect size, r=-.63[4], indicating that bilingual children produced novel labels more frequently (*Mdn(range)*=.40(.56)) than the monolingual children (*Mdn(range)*=.13(.63)). The monolingual outlier indicated by a star in Figure 1 produced 5 novel labels out of 8 productive responses but was not removed from the analysis as this would have increased the significance levels of the test results but reduced the sample size and hence effect size.

A between-group comparison for the rejection effect was not found to be statistically significant with *Mdn(range)*=.50(.56) for bilinguals and .57(.63) for monolinguals, Mann-Whitney's (one-tailed exact) U=7.5, $p>0.05$, z=-1.05, r=-.33[5]. Performing a second test in this manner increases the chances of a family-wise error, that is, of rejecting the null hypothesis when is should be accepted (a Type 1 error). Nevertheless, because this test suggests that there is no difference between the two groups in respect of the rejection effect, it should be reasonable to accept the null hypothesis (see Section 5.2 *re* Statistical Validity of Results).

5.3.2 Null Responses

- **Children in both groups provided significantly more null responses when asked to name novel label items (*Mdn*=.18) than when asked to name standard label items (*Mdn*=.08), *p* <0.05**
- **There was no significant statistical difference between language groups in their null responses**

[4] r=-1.984/3.1622=-.63
[5] r=-1.048/3.1622 = -.33

	Null responses for standard label items	Null responses for novel label items
Monolinguals	.17/.24(.33)	.25/.13(.64)
Bilinguals	.07/.07(.07)	.28/.18(.49)
Total across groups	.12/.08(.33)	.26/.18(.64)

Table 2. Memory Game: Mean/Median (range) of null responses as a proportion of total questions

The proportions of null responses for standard label items were compared with null responses for novel label items both within subjects and between groups. Measures of central tendency for null responses in both conditions are given in Table 2 and an illustration of within-subject comparisons in Figure 2a show that null responses were produced even in respect of standard items for which no novel labels had been given by the experimenter. The individual data points for both conditions were on a continuum and there were no significant outliers, however an ANOVA failed Levene's test of equality of variance (p=0.001) in respect of null responses to standard questions. Non-parametric tests showed that children across both groups produced significantly more null responses to novel label questions than to standard label questions, with *Mdn(range)* for novel labels=.18(.64) and for standard labels=.08(.33), Wilcoxon's z= -1.82, p<0.05, r=-.058[6], though effect size was negligible.

---

[6] r=-1.820/3.1622=-.058

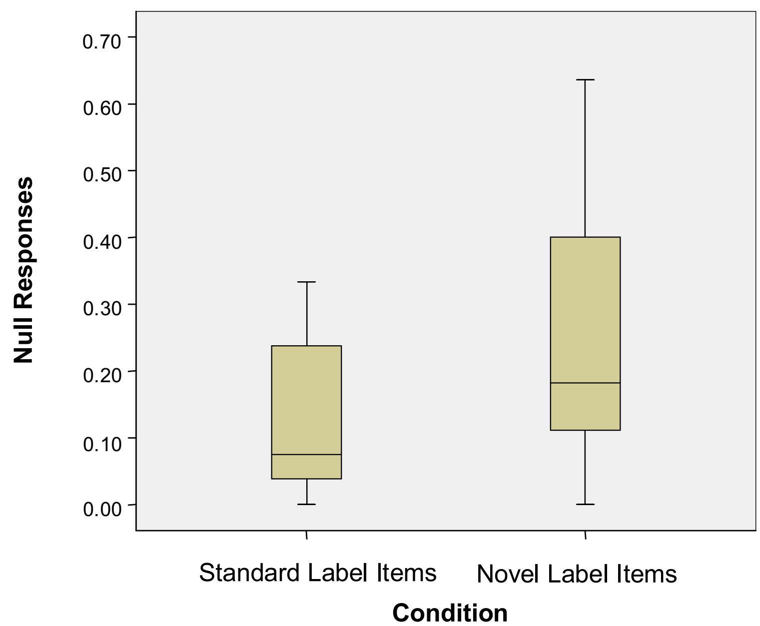


Figure 2a. Memory Game: Proportions of null responses to questions about standard and novel label items

Although the number of null responses for standard label items *Mdn(range)* for bilinguals=.07(.07) was less than for monolinguals *Mdn(range)*=.24(.33), there were no significant between-group differences for this condition nor for the novel label items. *Mdn(range)* for monolinguals was .13(.64) and for bilinguals=.18(.49), with equivalent mean ranks and hence identical Mann-Whitney's (one-tailed exact) U=10, z=-.52, r=-.17[7], p>0.05 for both conditions.

Interestingly, Figure 2b, which compares the trend lines for the ranked data, shows that 3 monolingual children produced null responses for standard label items ranging between proportions of .25 and .35 but the other two produced none. In contrast, the proportions for all 5 bilingual children ranged much lower but more consistently, between .05 and .10, with zero null responses.

[7] r=-.524/3.1622=-.165

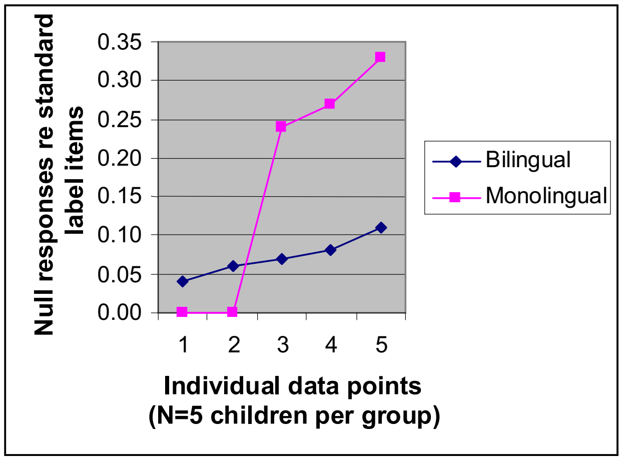


Figure 2b. Memory Game: Ranked proportions of null responses to questions about standard label items

5.4 Production Task (Snap Game)

5.4.1 Adoption and Rejection

- **There was no statistically significant difference between the bilingual and monolingual children in the number of labels produced, in terms of adoption or rejection**

Only one bilingual actually showed an adoption effect in the Snap game, producing one novel label (*mana*) correctly but also using *mana* to name the drum (*wug*) twice, while the monolinguals produced none. The differences between the groups for the adoption effect were not significant, with *Mdn(range)*=.00(.00) for monolinguals and .00(.17) for bilinguals, Mann-Whitney's (one-tailed exact) U=6, z=-1.00, r=-.35[8], p>0.05. All but one (bilingual) child displayed zero adoption effect. The rejection effect neared 100% in both groups, with *Mdn(range)*=1.00(.25) for monolinguals and 1.00(.67) for bilinguals.

[8] r=-1.00/2.828=-.35

5.4.2 Null Responses

- **Both groups in the Snap Game provided more null responses when asked to name novel label items than when asked to name standard label items, but the difference was not statistically significant**
- **There was no significant difference between groups in their null responses**

	Null responses for standard label items	Null responses for novel label items
Monolinguals	.31/.13(1.00)	.38/.42(.67)
Bilinguals	.03/.00 (.13)	.18/.17(.40)
Total across groups	.17/.00(1.00)	.28/.33(.67)

Table 3. Snap Game: Mean/Median (range) of null responses as a proportion of total questions

Measures of central tendency for null responses in the Snap Game are given in Table 3 and comparisons of the standard and novel label items conditions are illustrated in Figure 3a. Although there were no significant outliers, the data were skewed with wide ranges and sample size was reduced to 4 participants per group. There was no significant difference between the novel and standard label item conditions, with *Mdn(range)* for novel labels=.33(.67) and for standard labels=.00(1.00), Wilcoxon's z=-.96, p>0.05, r=-.34[9].

[9] r=-.962/2.828=-.34

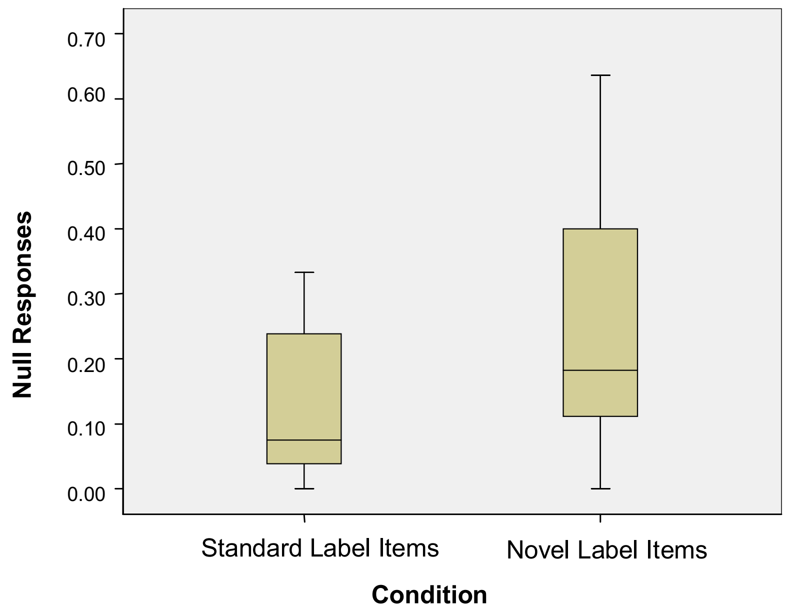


Figure 3a. Snap Game: Proportions of null responses to questions about standard and novel label items

Overall bilinguals produced fewer null responses to both types of questions than their monolingual counterparts. The between-group differences were not significant, however, with p>0.05 for both tests: *Mdn(range*) for bilinguals=.00(.13) and for monolinguals=.13(1.0), Mann-Whitney's U=5.0, z=-99, r=-.35[10] for standard label items; and for novel label items, Mann-Whitney's (one-tailed exact) U=4.5, z=-1.04, r=-.37[11], with *Mdn(range)* for bilinguals=.17(.40) and for monolinguals=.42(.67).

[10] r=-.992/2.828=-.35
[11] r=-1.042/2.828=-.37

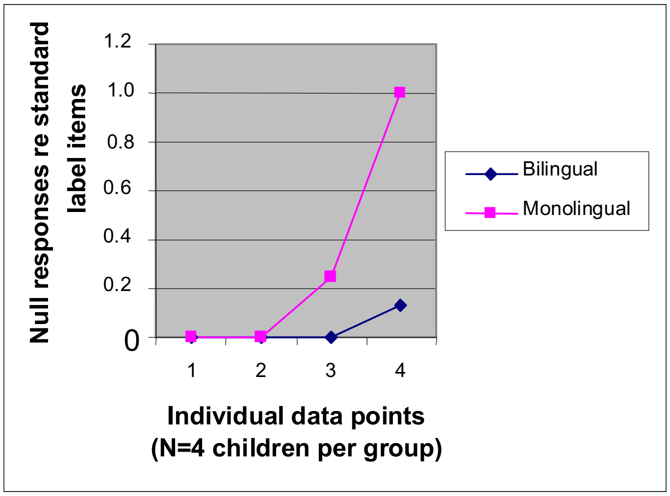


Figure 3b. Snap Game: Ranked proportions of null responses to questions about standard label items

Interestingly, the ranked null responses to standard label questions in Figure 3b show that two monolingual children produced zero null responses for standard label items and these are the same children as in the Memory game (see Figure 2b, Section 5.3.2). Thus across the two games the novel naming paradigm consistently highlighted two different types of behaviour in the monolingual group, with a stronger element of confusion in some of the monolingual children than the bilingual children, but with two monolingual children showing no hesitation in naming the standard items. The monolingual child who scored 100% null responses to standard items was not simply refusing to cooperate, because she had a score of .67 for the corresponding novel label item condition, suggesting perhaps that she felt the labels for all the standard items, were obvious and did not require naming again.

5.5 Training Tasks

5.5.1 Matching Task

- **All 10 children succeeded in placing all 9 standard- and 3 novel-label cards to their corresponding pictures when these were first named by the experimenter**

5.5.2 Mapping of Novel Label Items

- **All but one child in each group mapped at least one novel label item correctly, with median(range) values of .33(.67) and the distribution of correct mappings was identical for both groups**

	(a) Pointed to correct novel label item	(b) Pointed to standard label item	(c) Other
Monolingual	0.33/0.33(0.67)	0.20/0.00(0.67)	0.47/0.33(1.00)
Bilingual	0.33/0.33(0.67)	0.20/0.33(0.33)	0.47/0.67(0.67)

Table 4. Training Tasks - Mapping: Mean/median (range) of proportions of total selections

The measures of central tendency in Table 4 show that comparable median values were obtained for the monolingual and bilingual groups in the number of novel labels items correctly mapped. The distribution of raw scores for category (a) was identical for monolinguals and bilinguals, with 3 in each group correctly mapping one of the 3 novel label items.

Figure 4 shows 2 outliers in each group (marked by stars) who achieved two and zero correct matches respectively. Some children in both groups (*Mdn(range)*=.33(1.00) for monolinguals and =.67(.67) for bilinguals, pointed to other items (novel or standard), in addition to pointing to a novel label as indicated in the 'other' category, which reflected comparable degrees of confusion in both groups and reduced the proportional values of the novel label selections accordingly. Three bilinguals chose one standard label each, compared with just two monolinguals selecting standard labels. Interestingly, one of the monolingual children said 'I thought this was the *wug'* when pointing to a standard label, indicating that he was actually trying to map the novel label to the novel label item.

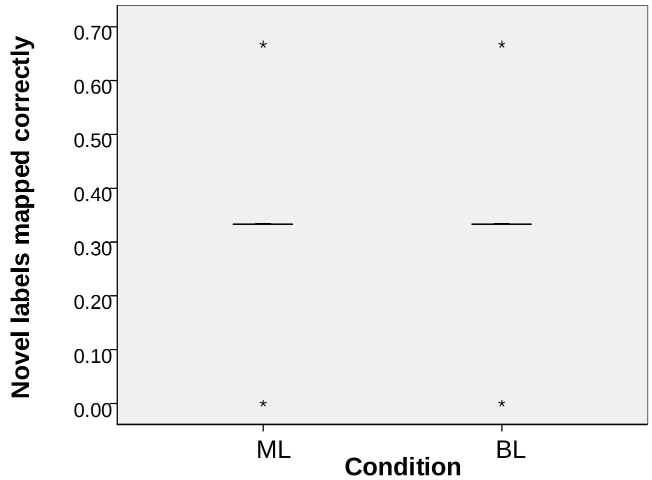


Figure 4. Training Tasks - Mapping: Proportions of novel label items mapped correctly when asked to point to each of 3 novel label items

5.5.3 Repetitions

- **Monolingual children spontaneously repeated the novel label items when first introduced to them, but the bilingual children did not ($p<0.05$)**

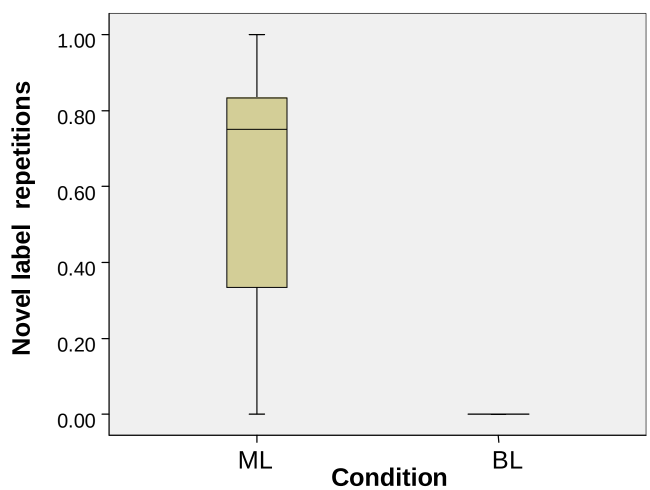


Figure 5a. Training Tasks: Proportions of novel labels produced as spontaneous repetitions

The monolingual children repeated novel labels more frequently (*Mdn(range)*=.75(1.0)) than the bilingual children (*Mdn(range)*=.00(.00)) and Mann-Whitney's (one-tailed exact) U=2.5, was significant, $p<0.05$, z =-.24, with a large effect size r=-.74[12].

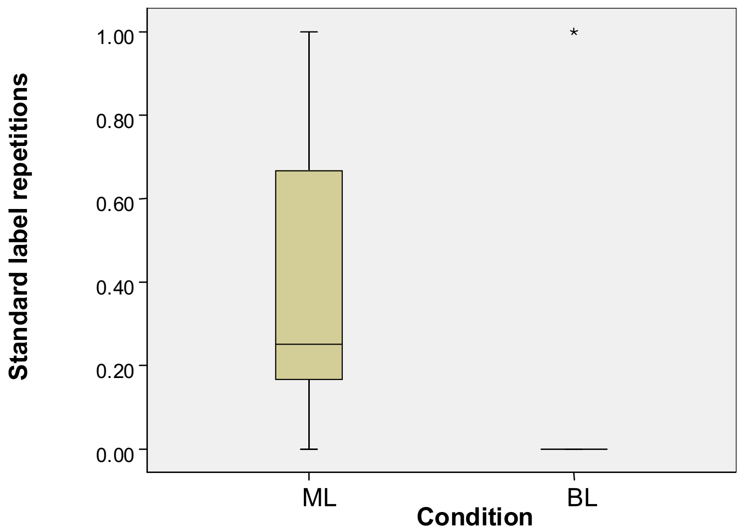


Figure 5b. Training Tasks: Proportions of standard labels produced as spontaneous repetitions

A comparison of novel and standard label repetitions between groups is illustrated in Figures 5a and 5b. Only two monolingual children produced zero repetitions of the standard labels, compared to all but one bilingual child, who is the outlier marked with a star in Figure 5b. A mixed ANOVA comparing the proportions of novel and standard label repetitions found that Levene's test of homogeneity of variance was significant for the novel labels (p=0.02) and hence this test could not be reported. Instead, a second Mann-Whitney test was performed to compare the proportion of standard labels produced by each group and was found to be close to significant, with *Mdn(range)*= .00(1.0) for bilingual children and *Mdn(range)*=.25(1.0) for monolingual children, and Mann-Whitney's (one-tailed exact) U=6.5, p=0.10, z=-3.14, r=-.99[13], a very large effect size. These findings suggest a tendency to repeat both novel and standard label items in the monolingual children, but not in the bilingual children.

[12] r=-.2353/3.1622= -.074

[13] r=-3.142/3.1622=-0.99

Kendall's tau[14] was performed to investigate whether children's mean length of utterances (MLU) correlated with a tendency to repeat novel labels. However, a low and non-significant correlation was found, with tau=.31 and p=1.23 (1-tailed). Further, there was no statistically significant difference in MLU between the groups, with *Mdn(range)*=1.75(3.0) for the bilinguals and 1.90(3.0) for the monolinguals, and Mann-Whitney's (one-tailed exact) U=9.5, p>0.05, z=-.63, r=-.20[15]. Thus the monolingual children's tendency to repeat labels was not associated with being more talkative than the bilingual children. It should be noted, however, that the number of utterances upon which the MLUs were sampled was low, ranging from zero to a maximum of 21 and hence this analysis can only be viewed as a qualitative indication of children's verbal output during the Training Tasks.

5.5.4 Contradictions

- **There was no significant difference between the groups in the number of times they contradicted the experimenter**

Only one monolingual and two bilinguals actually produced any contradictions in the Training Tasks, with the monolingual child producing two contradictions and the two bilingual children producing two and three contradictions respectively. Skewness values indicated that the group distributions were not comparable, at 2.24 for the bilingual and .89 for the monolingual group and hence a non-parametric test was conducted. The differences between the groups were not statistically significant, with *Mdn(range)*=0.0(2.0) for monolinguals and .00(3.0) for bilinguals, Mann-Whitney's (one-tailed exact) U=9.5, z=-.78, r=-0.25[16], p>0.05.

[14] Kendall's tau is recommended for small samples with tied ranks (Field 2005, p131)
[15] r=-629/3.1622=-.20
[16] r=-.775/3.1622=-.25

Chapter 6

Discussion

6.1 Review of Experimental Design

In order to provide a naturalistic setting the tests were not undertaken in a laboratory[17], the experimenter was not blind to conditions and the presence of siblings, peers, parents or nursery workers made it difficult to obtain uninterrupted sessions. These uncontrolled conditions, coupled with children's individual temperaments and attention spans led to variability in the number of opportunities to produce novel labels in the Production Tasks. However, post-hoc analysis showed that experimenter utterances did not significantly differ between groups, yet significant between-group differences in children's productive behaviour were obtained across these variable conditions**.** These findings would suggest that the results, though based on a small sample, indicate a potentially robust effect.

Coding of ME indicators and proportional measures

Coding involved categorizing behavioural indicators of receptiveness and production. For the measures of receptiveness, the number of opportunities to repeat or contradict, match and map the labels to the pictures was the same in both groups of children. Matching behaviour was recorded with a binary measure, with the child either placing the card on the correct picture or not. The identification of pictures when named involved either pointing to a novel label item, pointing to a standard label item, or some other behaviour, and so correct mapping of labels to pictures were counted as proportions of responses. Although ostensive behaviour has been used to indicate comprehension in other experiments (see Sections 2.2.3/4), it should be noted that the mapping task could not ascertain whether children were just focusing on matching the pictures or learning a relation between the novel labels and the items and hence these

[17] See Houston-Price, Plunkett and Duffy (2006) for an innovative approach to controlling test procedures and bias using digitized audio and visual materials

measures were better indicators of receptiveness, that is, of a somewhat weaker link between label and referent that that inferred by the term 'comprehension'.

The productive tasks involved prompting for responses. In the Memory game, the experimenter named each novel label item before asking the child 'what is it' to provide reinforcement, thus the number of times a child produced a novel label was by design related to the number of novel label items named by the experimenter. However, given that no significant difference was found in the number of novel label items named by the experimenter between groups, we can discount experimenter influence as a factor in the children's production of novel labels, at least as far as objectively observable cues are concerned.

Further, in both the Memory and the Snap games**,** if the child produced null responses to the question 'what is it?' the experimenter elicited an answer up to 4 more times. In order to ensure comparability between children, therefore, measures of adoption and rejection were quantified as proportions of total productive responses. It is possible that the number of experimenter elicitations, as guided by the children's null responses, may have somehow influenced the outcome of the productive tasks. However, as there was no significant difference in the number of null responses to elicitations for standard or novel questions between the two groups, we can discount the possibility that this procedure was a confounding factor in the productive tasks.

If the wrong novel label was produced, this was coded as 'other'. Interestingly, only one child in each group produced incorrect novel labels when naming a novel label item and this was '*mana*' in all instances, suggesting perhaps a recency effect as 'mana' was the last label produced by the experimenter in the Training Tasks. It is also possible that *mana* was simply phonologically the most memorable label.

Validity of comparing monolingual with more or less balanced bilinguals

Although the children were active bilinguals, parental questionnaires revealed that they were not equally balanced bilinguals. Figure 6 compares the bilingual children's adoption rates in the Memory game (expressed as percentages) with the amount of time they spent interacting in their non-English language. The German/English speaking child with minimal exposure to German (15%), appeared completely comfortable in German with his German-speaking parent and had the second highest adoption rate at 67%. In contrast, children with the highest degree of non-English interaction times of 60% and 80% (Italian/English and Danish/English), displayed the lowest adoption rates at 22% and 33% respectively. While the child with the highest adoption rate of 78% spent 50% of his time interacting in Italian. With the exception of the latter, then, there seemed to be an inverse correlation between these measures, suggesting that the more time they spent interacting intra-linguistically in their parent's native language, the lower the children's adoption rate.

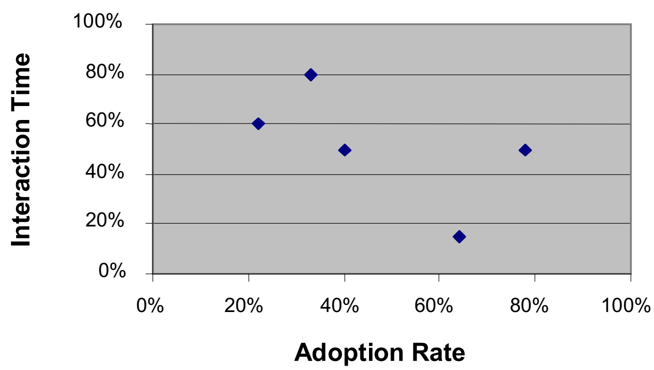


Figure 6: Memory Game Adoption rate compared with proportion of time spent interacting in non-English language
(N=5; data points represent individual bilingual children)

Arguably this trend would lend support to a heuristics-based interpretation, where greater exposure to a dominant language strengthens one-to-one mappings. However the qualitative nature of the parents' estimates, coupled with additional reports that most of the children had undergone periods of total immersion over the past year in their parents' native country, suggest that better quantitative data and control conditions would be needed to investigate any putative relationship in bilinguals between language dominance and ME.

Language tested only in English

The experimental tasks were only conducted in English which may have impacted the results if the bilingual children differed in their grasp of English. However, to some extent this was controlled by screening out any child who couldn't label all the novel label items in English. In addition, the different structures of the bilingual children's other languages, viz German, Italian and Danish, may have impacted the children's ability to follow syntactic cues that the novel labels referred to whole objects. In these languages, like English, major lexical items are salient by virtue of having 'at least one stressed syllable' (Peters, p. 173-174), though some are stress-timed and some are syllable-timed. Also like English, these languages are all Indo-European and non-tonal, with both definite and indefinite articles. However in Danish indefinite articles precede the nouns (Peters, p. 156) but definite articles follow them. This may have negatively impacted word learning in the Danish children as both article types preceded the novel label: 'this is *a wug*' and 'show me *the wug'*. Further, Scandinavian nouns are inflected for number and only the genitive is morphologically marked for case, but unlike English, singular nouns are marked for gender (Plunkett and Strömqvist, 1992), as are Italian nouns; while Italian and German articles are more inflected than English, with the former indicating number and gender and the latter also indicating case (Slobin, 1997 p. 281, citing Bates, Friderici and Wulfeck, 1987). The fact that a significant bilingual advantage was found on the basis of a sample comprising three non-English languages with different morphological markings for nouns and determiners, however, suggests that an effect attributable to the experience of bilingualism *per se* is perhaps more robust than one attributable to typological variation between English and these respective languages.

Other factors in experimental design

Despite the changes afforded by insights from the pilot study (see Section 4.2), perceptual confounds may have impacted the results. The proximity and likeness of objects depicted on the picture sheet (Appendix I) may have affected children's ability to discriminate between novel labels items and standard items. For example the drum (*wug*), a round solid blue picture with two drum sticks, was located immediately to the right and above the cake, which was also round and the same proportions as the drum, but solid green with candles and both items were on the right hand side of the sheet. This may have led two monolingual children to point to the cake or the cake *and* the drum when asked to point to the *wug* in the Training Tasks. Further, the sock (*mana*) and the chair (*poggle*) were located side by side in the bottom left-hand corner of the sheet, which may have led children to look to that corner but have difficulty in differentiating between the two novel label items.

Such perceptual confounds may have had an impact on the mapping task: whereas on the whole children mapped one item correctly when asked to point to each of the 3 novel label items, they tended to point to at least two out of the three target items. That is, although they did not achieve correct mappings in every case, they still tended to map novel labels to a novel label item more often than not, indicating a willingness to map the novel labels, tempered perhaps more by confusion and memory effects than an ME strategy. Further, any confusion may have persisted in the Memory and Snap games, leading perhaps to greater reluctance to produce novel labels.

Another possible confound was the use of novel label items for which common synonyms were already known. For example, the label '*poggle'* was used for the chair. However, during the pilot test one monolingual child named the chair a 'seat' in the pretest and did so throughout except in the game of Snap, where she called it a chair in two separate instances. Thus she already had a synonym for that word, and this may have affected her ability to acquire a third synonym – in fact there were zero tokens of '*poggle'* from this child. Given findings that children have difficulty in learning a third label (Liittschwager and Markman, 1994), future tests should avoid such confounds by selecting novel label items with no common basic level synonyms.

Finally, it should also be noted that the children were praised for getting the labels right (in both languages for the bilinguals) during the pre-test and this procedure may have self-primed them to use the standard labels, thus reducing novel label responses in subsequent tasks, though as the English pre-test immediately preceded the Training Tasks, such an effect should have been consistent across both groups.

6.2 Implications of findings in the context of previous research

Polynomy, synonymy and perspective-taking

In the present experiment, children from both language groups had no reason to believe that the label they knew in English for the novel label item was incorrect, other than making an inference based on the experimenter's use of a novel label. They were not encouraged by the experimenter or by the inherent look of the items to see the objects from a different analytic perspective. These conditions may explain the monolingual childrens' more extreme resistance to multiple labels relative to Deák, Yen and Pettit's (2001) experiment (p. 799, Table 2), which involved more extensive training and the association of ambiguous objects with novel labels, thus reducing the children's certainty in respect of target names. Yet none of the factors manipulated by Deák et al. to induce polynomy were necessary to induce the bilingual children in the present experiment to adopt the new labels, suggesting an inherently different approach to novel labels in bilinguals. Whether they saw new labels as polynyms representing different category or function labels, as in Deák et al.'s productive paradigm, or as synonyms for the same category, remains an open question. However, during the Memory game, several bilingual children provided spontaneous comments suggesting that they saw the novel label as representing the same category: 'it's *like* a chair', but bearing a different name in contrast with their own perspective: '*I want to call it* a chair'; '*I think* it was a sock'. Given that there were no inherent clues in the objects, it seems that the bilingual children were willing to use the novel labels simply because the experimenter named them so. That is, bilinguals were more willing to adopt the experimenter's perspective than the monolingual children, a finding that parallels those of Diesendruck (2005) and Diesendruck and Markson (2001).

In respect of vocabulary size, our results may also support the possibility that polynomy correlates with vocabulary size, if it can be assumed that bilingual children's vocabulary forms one and not two lexical systems[18]. However Deák et al. only noted this relationship post-hoc and further empirical research would be needed to elucidate any putative inverse correlation between vocabulary size and ME.

[18] Though as noted in Section 2.2.5, this remains an open question

Are spontaneous contradictions indicators of ME or a function of metalinguistic awareness?

It will be recalled that only one monolingual and two bilingual preschoolers actually produced any contradictions in the Training Tasks. During the Memory game the same monolingual child produced a further 3 spontaneous contradictions, while another bilingual child produced one spontaneous contradiction. Merriman and Bowman (1989) found 'explicit resistance' in only 6% of trials (p. 77) when investigating the immediate correction effect, a low number consistent with the present findings for both groups. However, in the present experiment the monolingual child produced a proportion of 0.17 novel labels in the memory game, while the corresponding proportions produced by the bilingual children who contradicted the experimenter were .40, .64 and .78. Thus the bilingual children who contradicted the experimenter showed evidence of the adoption effect in the Memory game while the monolingual child did not, which suggests that spontaneous contradiction is not a reliable predictor of ME. Arguably, the bilinguals' behaviour could be attributable to a precocious ability to dissociate the labels from the meaning of the items, in other words, to see language as a formal system, which in turn is predicated on an understanding of the 'representational nature of mental states' (Doherty and Perner, 1998, p. 303).

Is spontaneous repetition of novel labels actually an indicator of ME?

In the Training Tasks the bilingual children were not prone to spontaneous repetition for any label type. Yet they were equally talkative during the session, as indicated by comparison of their MLU which found no statistically significant difference between groups. Further, spontaneous repetition was associated with the subsequent application of ME, as the monolingual children did not produce novel labels in the Memory Game, which echoes Merriman and Bowman's findings (1989) that 2-3½ year-olds imitated the novel labels on 42% of trials during the disambiguation task (second experiment, p. 77). In contrast, while the bilingual children did not repeat novel labels upon their first introduction to them, they were clearly less guided by ME than monolingual children, in the Memory game. Spontaneous repetitions were perhaps indicators of surprise on the part of monolinguals and thus may be prove to be reliable predictors of ME strength, while also lending further support to the notion that word learning triggers a different process upon hearing a novel label in mono- versus bilingual children[19].

Are bilingual children simply better at word learning?

An intuitive interpretation of the results could be that word learning places less processing demands on bilingual children, who are simply better at learning new words. However the bilingual children's greater adoption of novel labels in the Memory Game, but virtually 100% rejection rates across groups in the Snap Game, would not appear to be consistent with a simple explanation based on the different word learning capabilities of mono- and bilingual children, given that the experiment afforded the same opportunities to all children to form mappings between novel label and referent. Children in both language groups appeared to learn the novel labels to the same extent according to conventional measures of receptiveness, that is, they all successfully matched the correct pictures to the novel label items on the display sheet and there was no statistical difference between the groups in their selection of novel label items when asked to identify them by pointing.

Piccin and Blewitt's (2007) resource conservation account would predict that bilingual children and monolingual children should deliver similar performances in conditions of identical communicative benefit, while according to a heuristic account, bilingual children should show a lesser tendency to apply ME, which they did in the Memory Game. Alternatively, the observed differences in the Memory Game may be due to word learning actually placing different processing demands on mono- and bilingual children, even when communicative benefit doesn't vary. However, such a differential processing account would in turn need to explain the lack of differences between groups in the Snap game. One possible explanation would be to attribute the lack of differences in the second game to the levelling influence of short-term memory, fatigue or attentional factors, which would be corroborated by the fact that one child in each group lost interest and didn't play the final game.

[19] For example, Conboy and Mills (2006) found different ERP distributions between known and unknown words in simultaneous English/Spanish bilinguals aged 9-22 months and noted that word learning might place qualitatively different processing demands in bilinguals and monolinguals

With a larger sample it would have been possible to vary the order of the Memory game and Snap games in order to ascertain whether memory, attentional or contextual factors played a greater role in the children's responses. However, in the absence of such controls, a more probable explanation of the findings is that contextual cues played a stronger role in the bilingual children, who effectively displayed greater responsiveness to their interlocutor's perspective by showing a greater willingness to adopt the novel labels during the Memory Game, when the experimenter used the novel labels. In the Snap Game, when the experimenter did not use the novel labels, neither did the children in either group. The difference between the mono- and bilingual children in the Memory game could be attributable to the communicative dynamic of the game, with the bilingual children more willing to 'play along' with the experimenter's perspective. This interpretation is consistent with the well-attested finding that bilingual children adopt the language appropriate to the context and/or their interlocutor, and in this experiment would manifest itself in selecting the right word in the right context.

Chapter 7

Conclusions and Future Research

The present study aimed to compare ME in 3-years-old mono- and bilingual children using an innovative naming paradigm - adoption - in conjunction with more conventional measures of receptiveness. The focus on a tightly constrained age group, the use a productive paradigm and the observation of behaviour both on first exposure to target labels and during the experimental phase aimed to address discrepancies in the developmental literature regarding the application of ME during word learning experiments. Our findings concur with those of previous experiments in ostensive tests and in older bilingual children (Davidson, Jergovic, Imami and Theodos, 1997; Davidson and Tell, 2005), that on the whole bilingual preschoolers place less reliance on ME than their monolingual counterparts.

We predicted that the experience of learning two languages would impact the linguistic mechanisms of word learning. If mutual exclusivity was purely a linguistic bias, we predicted that bilingual children, just like monolingual children, would resist two labels per object in English. If, however, mutual exclusivity arose or was reinforced intra-linguistically from the experience of one-to-one mappings between label and referent, but in addition could be influenced by exposure to cross-language equivalents and socio-pragmatic factors such as interlocutor's perspective or ability to pass a false belief test, we then predicted that even 3-year-old bilingual children would apply two labels to one object in English more readily.

Although the main hypothesis was confirmed by a significant difference in the adoption effect for the Memory Game ($p<0.05$), indicators of ME were found in both groups and behaviours predicted for the monolinguals were not exclusive to that group. Rather, the bilingual children sometimes displayed an ability to produce multiple labels for a referent when the communicative dynamic required it, but otherwise showed a preference for the one-label, one-referent strategy which they would naturally have acquired as a result of their experience of a monolingual environment in each of their respective languages.

Given the cross-sectional nature of the experimental design, it was not possible to explore developmental manifestations of ME. The finding that a difference was evident in 3-year-olds concurs with an explanation based on the ability to pass a False Belief test, given that Goetz (2003) has shown a bilingual advantage in this age group. In contrast, whether the latter is purely associated with an ability to address interlocutors in their respective language would require a comparison of mono- and bilingual children in the 18-24 month age bracket, since this ability has been documented to emerge before the age of 2 (as reviewed by Deuchar and Quay, 1999). While to investigate a putative association between the developmental paths of ME and ToM would require comparable tests to be conducted for at least 3 developmental stages: in 18-24 month-olds (for the perspective-taking account), 3-year-olds (to investigate the false-belief threshold) and 5-year-olds (to investigate discrepant findings of ME increasing or decreasing with age in monolinguals and bilinguals).

Statistically significant discrepancies between comprehension (as measured by ostensive or selective behaviour) and production in the application of ME by monolingual children have been highlighted previously, notably those observed by Deák et al. (2001) and Piccin and Blewitt (2007) in monolingual children. Such discrepancies have also been noted anecdotally by Davidson et al. (1997) in both mono- and bilingual children in tests of acceptance following their restriction experiment. Similarly in the present experiment, mono- and bilingual children exhibited asymmetrical results in different ways with respect to receptive and productive measures of ME. Monolingual children did not tend to apply ME in the matching and pointing tasks during Training (both measures of comprehension) nor in their propensity to repeat novel labels (although as we have seen, the latter is arguably not an indicator of ME). They applied ME in their contradictions and in their propensity to produce standard labels in the Memory game and the Snap game. Similarly, bilingual children didn't tend to apply ME in matching and pointing during the Training Tasks. In contrast, they showed no propensity to repeat novel labels. Bilinguals also produced some contradictions in the Training tasks and produced slightly more standard labels than novel labels in the Memory game, yet displayed a greater propensity than monolinguals to produce novel labels in the Memory game. In other words, there was no linear relationship between productive and receptive indicators of ME in the bilingual children. These findings, compounded with those of

Deák et al. (2001), Piccin and Blewitt (2007) and Davidson et al. (1997), highlight the need for further research into productive indicators of ME in conjunction with more traditional indicators of comprehension and receptiveness based on ostensive paradigms.

The observed dissociation of comprehension and production does not sit easily within a rigorous interpretation of ME as a purely linguistic bias. However, such dissociations support a more flexible interpretation of ME as a developmental phenomenon that emerges as a product of experience and is responsive to context over time, as postulated by Hirsh-Pasek, Golinkoff, Hennon and McGuire (1994), MacWhinney (2005), Karmiloff-Smith (1992), Houston-Price, Plunkett and Duffy (2006) and others.

Our findings support an experience-based account of word learning, in which socio-pragmatic cues probably played a differentiating role during the course of the experiment. Bilingual children displayed greater responsiveness to their interlocutor's perspective by showing a willingness to adopt the novel labels during the Memory Game, when the experimenter used the novel labels. In the Snap Game, when the experimenter did not use the novel labels, neither did all but one child in either group. Nevertheless, bilingual children applied ME to the same extent as monolingual children in respect of the rejection effect, that is, in producing comparable numbers of standard labels instead of novel labels. While it is possible that a larger sample may have yielded a statistically significant difference for the rejection effect, it is likely that word learning, as opposed to the production of familiar words, is the subject of different mechanisms in bilingual and monolingual children. This interpretation is further supported by the finding that for the monolingual children, a comparable ability to identify novel label items through ostensive behaviour was not matched by ability to produce the novel labels.

In conclusion, the variability of responses observed in the present experiment is consistent with a flexible account of ME as a heuristic learned from monolingual input, the application of which varies in bilingual preschoolers according to both ambient language and socio-pragmatic context. In other words our findings support an interpretation of ME which is both sensitive to context in real-time and a function of cumulative experience within the broader developmental framework.

References

Akhtar, N., Carpenter, M., and Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, *67*, 635-645.

Au, T. K., and Glusman, M. (1990). The principle of mutual exclusivity in word learning: to honor or not to honor?. *Child Development*, *61,* 1474-1490.

Baldwin, D. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62,* 875-890.

Baldwin, D. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology*, *29,* 832-843.

Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.

Conboy, B., and Mills, D. L. (2006). Fast-track report. Two languages, one developing brain: event-related potentials to words in bilingual toddlers. D*evelopmental Science* (1), F1–F12

Davidson, D. and Tell, D. (2005). Monolingual and bilingual children's use of mutual exclusivity in the naming of whole objects. *Journal of Experimental Child Psychology*, *92*, 25-45.

Davidson, D., Jergovic, D., Imami, Z., and Theodos, V. (1997). Monolingual and bilingual children's use of the mutual exclusivity constraint. *Journal of Child Language*, *24*, 3-24.

Deák, G. (2000). Chasing the fox of word learning: Why 'constraints' fail to capture it. *Developmental Review*, *20*, 29-80.

Deák, G. O., Yen, L., and Pettit, J. (2001). By any other name: when will preschoolers produce several labels for a referent? [Electronic version]. *Journal of Child Language*, *28*, 787-804.

Deuchar, M. and Quay, S. (1999). Language choice in the earliest utterances: a case study with methodological implications**.** *Journal of Child Language, 26*, 461- 475.

Deuchar, M. and Quay, S. (2000). *Bilingual acquisition. Theoretical implications of a case study.* Oxford: Oxford University Press.

Diesendruck, G. (2005). The Principles of conventionality and contrast in word learning: an empirical examination. *Developmental Psychology*, *41*(3), 451–463.

Diesendruck, G., and Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*, 630-644.

Doherty, M. and Perner, J. (1998). Metalinguistic awareness and theory of mind: just two words for the same thing?. *Cognitive Development*, *13*, 279-305.

Field, A. (2005). *Discovering statistics using SPSS*. (2nd Ed.). London: Sage Publications.

Gathercole, V. C. (1987). The contrastive hypothesis for the acquisition of word meaning: A reconsideration of the theory. *Journal of Child Language*, *14*, 493-532.

Gathercole, V. C. (1997). The linguistic mass/count distinction as an indicator of referent categorization in monolingual and bilingual children. Child Development, 68(5), 832-842.

Gathercole, V. C. M., Thomas, E. M, and Evans, D. (2000). What's in a noun? Welsh-, English-, and Spanish- speaking children see it differently. First Language, 20(58), 55-90.

Goetz, P. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition 6*(1), 1-15.

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., and Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Journal of Developmental Psychology*, *28*(1), 99–108.

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

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults [Electronic version]. *Cognitive Psychology*, *53*, 310–344.

Halberda, J. and Goldman, J. (submitted). One-trial Learning in 2-year-olds: Children Learn New Nouns in 3 Seconds Flat. Retrieved May 1st, 2007 from Johns Hopkins University website:
http://www.psy.jhu.edu/~halberda/publications/HalberdaGoldman.pdf

Hirsh-Pasek, K., Golinkoff, R. M., Hennon, E. A., and McGuire, M.J. (2004). Hybrid theories at the frontier of developmental psychology: The emergentist coalition model of word learning as a case in point. In D.J. Hall and S. R. Waxman (Eds.), *Weaving a lexicon*. Cambridge, MA: MIT. Retrieved Aug 12, 2007 from http://copland.udel.edu/~roberta/pdfs/Hybrid%20theories.pdf.

Holowka, S., Brosseau-Lapré, F., and Petitto, L. (2002). Semantic and conceptual knowledge underlying bilingual babies' first signs and words. *Language Learning, 5*(2), 205-262.

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

Ishida, H., Kosugi, D., and Itakura, S. (2003). Mutual exclusivity and taxonomic bias in combination lead preschoolers' word learning. *Psychological Reports*, *93*, 649-652.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science.* Cambridge, MA: MIT Press.

Landau, B., Smith, L. B., and Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *59*, 299-321.

Liittschwager, J. C. and Markman, E. M. (1994). 16- and 24-month-olds' use of mutual exclusivity as a default assumption in second-label learning. *Developmental Psychology*, *6*, 955-968.

Macnamara, J. (1982). *Names for Things*. Cambridge, MA: MIT Press.

MacWhinney, B. (1989). Commentary: Making words make sense. In: Merriman, W. E. and Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development, 54*(3-4, Serial no. 220).

MacWhinney, B. (2005). The emergence of linguistic form in time. *Connection Science*, *17*(3–4), 191–211.

Markman, E. M. (1989). *Naming and categorization in children.* Cambrigde, MA: MIT Press.

Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In M. R. Gunnar and M. Maratsos (Eds.): *Modularity and Constraints in Language and Cognition: The Minnesota Symposium on Child Psychology* (Vol. 25, pp 59-101). Hillsdale, NJ: Lawrence Erlbaum.

Markman, E. M. and Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, *16*, 1–27.

Markman, E. M. and Wachtel, G. A. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121-57**.**

Markman, E. M., Wasow, J. L., and Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*, 241-275.

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

Merriman, W. E. and Kutlesic, V. (1993). Bilingual and monolingual children's use of two lexical acquisition heuristics. *Applied Psycholinguistics*, *14*, 229-249.

Merriman, W. E. and Schuster, J. M. (1991). Young children's disambiguation of object name reference. *Child Development*, *62,* 1288-1301.

Merriman, W. E. and Stevenson, C. M. (1997). Restricting a familiar name in response to learning a new one: evidence for the mutual exclusivity bias in young two-year olds. *Child Development*, *68,* 2, 211-228.

Merriman, W. E., Marazita, J., and Jarvis, L. H. (1993). Four year olds' disambiguation of action and object word reference. *Journal of Experimental Child Psychology*, *56*, 412-430.

Merriman, W. E., Marazita, J., and Jarvis, L. (1995). Children's disposition to map new words onto new referents. In M. Tomasello & W. E. Merriman (Eds), *Beyond names for things: children's acquisition of verbs.* Hillsdale, NJ: Lawrence Erlbaum.

Monaghan, P., Christiansen, M. H., and Chater, N. (in press). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition [Electronic version]. *Cognitive Psychology (2007)*, doi:10.1016/j.cogpsych.2006.12.001

Nelson, K. (1988). Constraints on word learning?. *Cognitive Development*, *3*, 221-246.

Pearson, B. Z., Fernández, S., and Oller, D.K. (1995). Cross-language synonyms in the lexicons of bilingual infants: One language or two? *Journal of Child Language, 22,* 345-368.

Peters, A. M. (1997). Language typology, prosody, and the acquisition of grammatical morphemes. In D. I. Slobin (Ed.): *The crosslinguistic study of language acquisition.* (Vol. 5, pp. 135-197). Hillsdale, NJ: Lawrence Erlbaum Associates.

Piccin, T. B. and Blewitt, P. (2007). Resource conservation as a basis for the mutual exclusivity effect in children's word learning. *First Language*, *27*(1), 5–28 (200702). Retrieved May 14, 2007 from http://fla.sagepub.com at SWETS WISE ONLINE CONTENT, DOI: 10.1177/014272370706754.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: MIT Press.

Plunkett, K., and Strömqvist, S. (1992). The acquisition of Scandinavian languages. In D. I. Slobin (Ed.): The crosslinguistic study of language acquisition. (Vol. 3. pp 457-555). Hillsdale, NJ: Lawrence Erlbaum.

Quine, W. (1960). *Word and object.* Cambridge, MA: MIT Press.

Slobin, D. I. (1997). Origins of grammaticizable notions. In D. I. Slobin (Ed.): *The crosslinguistic study of language acquisition.* (Vol. 5, pp. 265-323). Hillsdale, NJ: Lawrence Erlbaum.

Smith, A. D. M. (2003). Intelligent meaning creation in a clumpy world helps communication, *Artificial Life, 9*(2), 559-574.

Smith, A. D. M. (2005). The Inferential Transmission of Language. *Adaptive Behavior*, *13*(4), 311-324.

Tomasello, M. (1988). The role of joint-attentional processes in early language acquisition. *Language Sciences*, *10*, 69-88.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M., Strosberg, R., and Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language, 23,* 157-176.

Werker, J. F., Cohen, L. B., Lloyd, V., Stager, C., and Cassosola, M. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, *34*, 1289-1309.

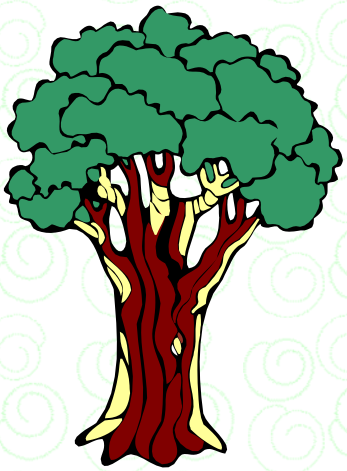
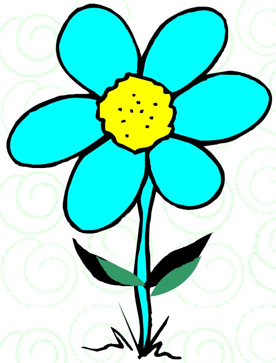
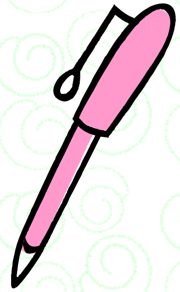
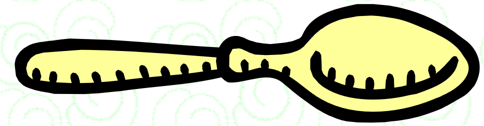
Woodward, A. L. and Markman, E.M. (1991). Constraints on learning as default assumptions: Comments on Merriman and Bowman's "The mutual exclusivity bias in children's word learning." *Developmental Review*, *11,* 137-163.

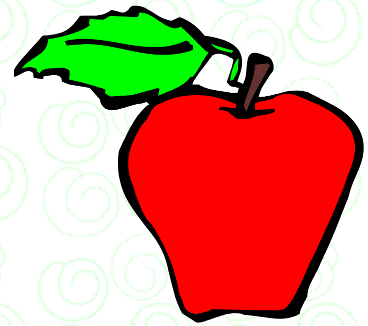
Xu, F., Cote, M., and Baker, A. (2005). Labelling guides object individuation in 12-month-old infants. *Psychological Science*, *16*(5), 373-377.

Appendices

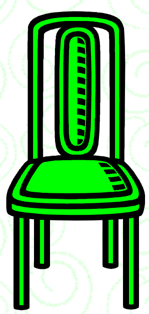
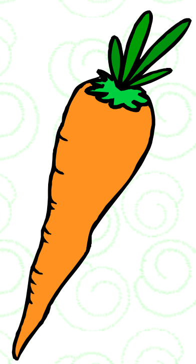
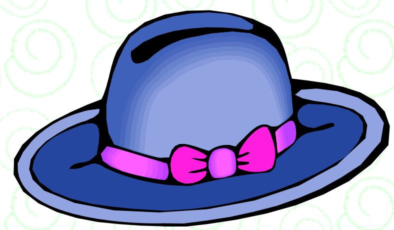
Appendix I

Laminated Picture Sheet

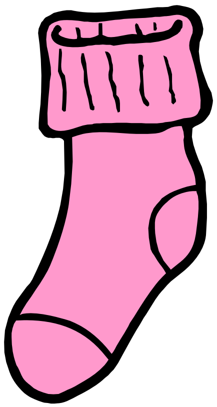


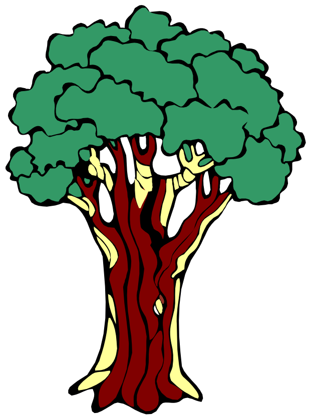


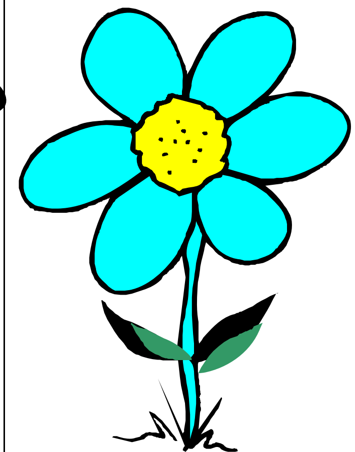




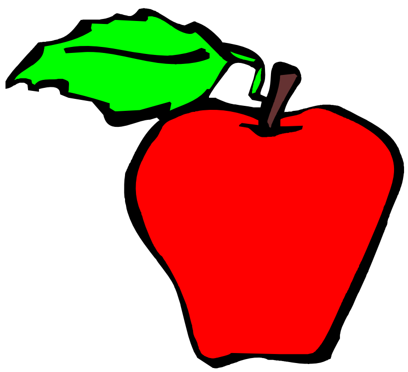
Appendix II

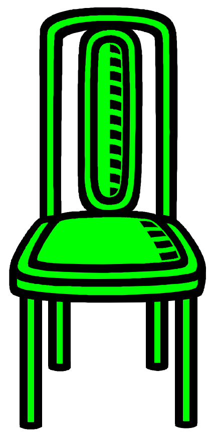
Laminated Cards

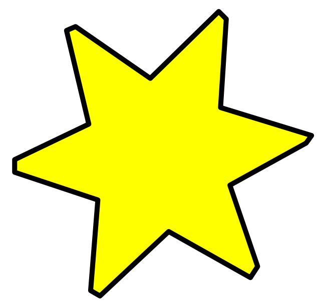


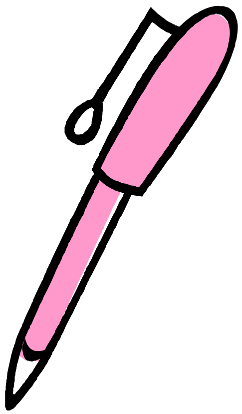
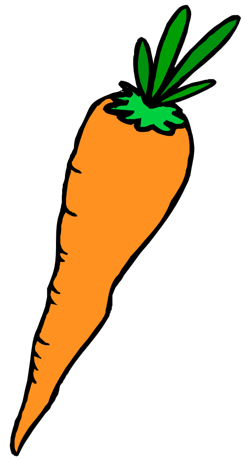
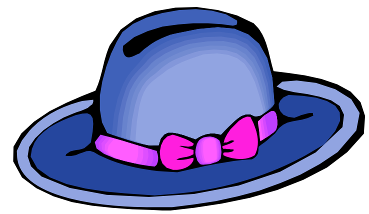
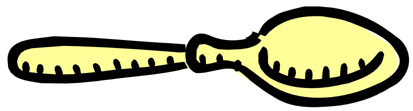










Appendix III

Order of Items, Pre-test and Training Tasks

Date: ____________

Child's Name: ____________

Pre-Test L2 (bilingual children only)

Parent names the following items in L2 and asks the child to point to them:

L2: ____________

1. Cake ____________
2. Carrot ____________
3. Pen ____________
4. Star ____________
5. Tree ____________
6. Flower ____________

Child is asked by Parent in L2 to name the following items:

L2: ____________

1. Hat ____________
2. Chair ____________
3. Apple ____________
4. Drum ____________
5. Spoon ____________
6. Sock ____________

Pre-Test/EN (all children)
Date: ____________ Child's Name: ____________
Date of Birth _____________ (Age: _______years ______months)

Experimenter names the following items and asks the child to point to them:

Correct Match?

1. Cake
2. Carrot
3. Pen
4. Star
5. Tree
6. Flower

Child is asked by Experimenter to name the following items:

Child Utterance?

1. Hat
2. Chair
3. Apple
4. Drum
5. Spoon
6. Sock

EXP 1 Thimbles Game Part 1- Experimenter holds cards

	Child Utterance?	Correct Match?	Rejection/Other Comments?
Cake			
Carrot			
Pen			
Star			
Tree			
Flower			
Hat			
Chair			
Apple			
Drum			
Spoon			
Sock			

EXP 1 Thimbles Game Part 2
Leaving the chart with the cards on the corresponding items, Experimenter asks child to point to each of the **novel labels** only

	Child Utterance?	Correct Match?	Rejection/Other Comments?
Cake			
Carrot			
Pen			
Star			
Tree			
Flower			
Hat			
Chair			
Apple			
Drum			
Spoon			
Sock			

Appendix IV

Order of cards, Productive Tasks: Memory and Snap Games

Date: ____________

Child's Name: ____________

Order of cards for productive task: Memory game - Target Items are in **bold**

Experimenter Experiment 2	Repeat?	Elicited Utterance ?	Child Experiment 2	Elicited Utterance ?	Reject?	Other Comments?
Flower						
Apple						
Cake						
Spoon						
			Apple			
			Flower			
Chair						
Spoon						
			Carrot			
			Pen			
Star						
Tree						
			Chair			
			Tree			
Sock						
Carrot						
			Hat			
			Cake			
Hat						
Pen						
			Sock			
			Star			
Carrot						
Drum						
			Spoon			
			Drum			
			Hat			
			Pen			

Date: ____________

Child's Name: ____________

Order of cards for Productive Task: Snap Game
Items which will elicit a "snap" are *italicised*. Target items are in **bold**.

Experimenter Experiment 3	Child Experiment 3	Elicited Utterance?	Reject?	Other Comments?
Cake				
	Pen			
Spoon				
	Carrot			
Carrot				
	Sock			
Spoon				
	Hat			
Flower				
	Chair			
Star				
	Pen			
Pen				
	Sock			
Hat				
	Spoon			
Carrot				
	Drum			
Tree				
	Tree			
Pen				
	Chair			
Carrot				
	Cake			
Spoon				
	Drum			
Apple				
	Apple			

Appendix V

Informed Consent Form

Language Acquisition

University of Edinburgh Research Project

Informed Consent Form

This experiment is designed to investigate how children acquire new words. The experiment involves playing a series of games using picture cards and a board. Firstly, we will check that your child knows the word for each of the pictures. Then, in the form of a game, the experimenter will introduce new labels for some of the pictures. We will then play a memory game and a version of snap to see how your child reacts to these new words.

You or your child may decide to withdraw from the experiment once participation has begun at any stage.

Your child will be given a small gift for participation.

An audiovisual recording of the experiment will be made. Data or summarized results will not be released in any way that could identify you or your child.

At the end of the session, you have the right to a complete explanation ("debriefing") of what this experiment was all about. If you have questions afterward, please ask Madeleine Campbell or email s0567971@sms.ed.ac.uk

I have read the above information and agree to let my child participate in this study. I have received a copy of this form.

Parent/Guardian's name (print) ___________________________________

Child's name (print) ___________________________________

Child's date of birth _________________________ **Child's Age** ___________

Parent/Guardian's signature _____________________________________

Date ______________

Thank you for participating!