DOCTOR OF PHILOSOPHY

Cognitive effects of language differences collection categories in Welsh and English

Roberts, Seren Haf

Award date:
2003

Awarding institution:
University of Wales, Bangor

University of Wales

Prifysgol Cymru

Cognitive Effects of Language Differences: Collection Categories in Welsh and English

Seren Haf Roberts

A Thesis submitted to the School of Psychology, University of Wales, Bangor, in partial fulfilment of the requirements of the Degree of Doctor of Philosophy.

1st September 2003

Summary

The ways in which collective nouns are acquired is poorly understood. The research that has been carried out to date indicates that collective nouns are extremely difficult for children to acquire; children have not yet developed the cognitive complexity required to understand their meanings. However, only English speaking participants have been investigated. Learning collective nouns may be difficult for English speakers because the structure of English emphasises individuals. In contrast, Welsh has a complex number marking system whereby the basic forms of some nouns refer to collections and are modified with a unit ending to individuate one item out of the collection (e.g. *coed* 'trees' versus *coeden* 'tree'). Thus, the structure of Welsh may allow greater conceptualisation of entities as collections. This thesis attempts to explore the relationship of such a system on the acquisition of nouns by comparing Welsh- and English-speaking children and adults on a range of cognitive tasks.

Two studies investigate the differences in the distribution of different noun types across the two languages. First noun distributions in written texts are examined followed by an exploration of the use of different noun types in the language input. A third study examines categorisation and recognition patterns for novel objects and novel nouns across the two languages. A final study explores attention to and recall of *number* changes across different noun type categories by speakers of the two languages. The results indicate that language differences do influence aspects of cognitive processing across the two language groups. Several factors, including age, object properties and syntax, play a pivotal role in the acquisition of collective categories. These findings are discussed in relation to the theories of language acquisition and the theory of linguistic relativity.

Acknowledgements

The ESRC and Unilever jointly funded the PhD studentship that has resulted in the research outlined in this thesis. I would like to take this opportunity to appreciate the guidance and supervision I received from Dr Virginia C. Gathercole throughout this learning process. I would also like to thank Professor Nick Ellis for his advice and help in accessing the Welsh Language database, Tracey Bywater for her assistance in data collection, Helen Sproson for the loan of her data from her undergraduate research project, and the staff at the School of Psychology (UWB) for providing a supportive environment and many of the resources needed to carry out this research. I would also like to thank Dr. Andrew Stewart and Francis Forman at Unilever Research Laboratory for their encouragement, assistance in data collection, and for sharing their valuable research experience.

A very special thanks to all the schools, teachers, children, parents and adults that gave up their time to participate in this research. Finally, to my family and friends, especially Jois Bailey and Emyr Wyn Jones, for their unwavering support, endless helpful suggestions, and for putting up with me throughout this process, I thank you.

Dedications

Children are fascinating learning machines. I am awe-struck at the information that children assimilate in such a short time. In particular, I am astonished at the speed at which young children understand and commandeer at least one language.

Throughout my PhD studentship I have had the good fortune to observe and enjoy the developmental accomplishments that my young bilingual niece, Mari Eirun, and four bilingual nephews, Tomos Oliver Murray, Aled Roberts, Rhys William Murray, and Iwan Michael Murray, have achieved. From them I have gained the wonder, the inspiration, and the determination to complete the work in this thesis. I dedicate this thesis to them and to all the future generations of children that will help our language flourish.

Table of Contents

Preface

According to Baker (1985, p.1) "The history of the Welsh speaking population in the 20th century is a history of decline". All Wales Census figures support this claim and show that the number of Welsh speakers has declined from 50% in 1901 to 18.6% in 1991. Moreover, the number of monolingual Welsh speakers has declined from 15% to none in the same time period. More recent figures from the 2001 census have indicated an increase in the number of Welsh speakers in Wales, with nearly 600,000 Welsh speakers – amounting to 21% of the population in Wales. Nevertheless, it is currently assumed that there are no monolingual Welsh speakers over the age of 3 years (Deuchar, in press; Thomas and Gathercole, in press).

It is also important to note that the overall figures for the Welsh-speaking population mask the regional variation in the numbers of Welsh speakers observed across Wales. As pointed out by Deuchar (in press), the percentage of Welsh speakers in some areas of Northwest Wales can reach 80%.

The decline in the number of Welsh speakers is largely due to historical political pressures (see Deuchar, in press, for discussion). For example, the Education Act of 1870, born from the belief that Welsh was a hindrance to the moral progression and commercial prosperity of the Welsh people, led to the emergence of English-medium education in which children were actively punished in schools for speaking Welsh. In recent years, educational and employment legislation promotes Welsh in Wales and has undoubtedly influenced the upsurge in the numbers of Welsh speakers (Thomas and Gathercole, in press). These include the Education Act of 1988 which ensured that in Wales, Welsh was an obligatory subject in all secondary schools, and Welsh has been compulsory in primary schools since 1996 (Davies, 1999).

Currently, the schools in Wales are classified as either designated Welsh-medium schools, natural Welsh-medium schools or English-medium schools.

Although children in both primary and secondary English-medium schools in Wales are taught Welsh as a second language, the amount of exposure is minimal when compared to the Welsh-medium schools where the primary language spoken is Welsh. Despite the strength of Welsh in Welsh-medium schools, English continues to have a dominant place in the community. Consequently, Welsh-speaking children are exposed to English in some form or other and are largely bilingual. However, the degree of bilingualism varies from one child or adult to another depending on the region, education and home language.

Therefore, although Welsh- and English-speaking children are in the same cultural and educational systems, research on Welsh language can be problematic for several reasons. First, the Welsh-speaking population is relatively small compared to many other languages. Second, Welsh speakers are largely bilingual with very few, if any, monolinguals available for research. Third, the degree of bilingualism varies from one individual to another with no standardised method of determining the degree of bilingualism in any Welsh speaker.

Thus, the research participants in this thesis were identified on the basis of the information contained in language background questionnaires. These were self-reports for adult participants and parental reports for child participants. To include participants that were as near monolingual Welsh speakers as possible in the research, only Welsh speakers that reported their home language to be over 80% Welsh, attended a Welsh-medium school and reported Welsh as their first language were included (unless otherwise stated). Note that the same criteria in English were set for English speakers, and that all English-speaking participants rated their home language as 100% English.

Introduction to thesis

There are three main objectives to this thesis:

(1) Research using cross-linguistic evidence has suggested that children acquire language in ways that reflect the structure of the language they are learning. Children acquiring different languages may conceptualise entities in ways facilitated by their language. Using cross-linguistic investigations, this thesis explores the influence of language structure on cognitive processes and language acquisition.

(2) The acquisition of collective nouns is poorly understood. Current theories of language acquisition do not adequately account for the use and acquisition of these noun types. This thesis aims to develop empirically based knowledge of the acquisition of collection categories that will contribute not only to understanding the acquisition of collective nouns but also to a greater understanding of language acquisition in general.

(3) Research conducted on language acquisition is largely based on the English-speaking population. Theories of language acquisition need to be applicable across languages. Welsh differs from English in crucial ways with regard to collective nouns. This thesis utilises a cross-linguistic approach to examine the acquisition of collection categories across these two languages.

To achieve these objectives, this thesis will be structured as follows. Chapter 1 consists of two sections. The first section introduces the notion of linguistic relativity and reviews the research on the relationship between conceptual and linguistic categories. The second section discusses the theories of word meaning acquisition with particular emphasis on the empirical evidence pertaining to the acquisition of collective nouns.

Chapter 2 outlines the number marking systems in Welsh and English paying particular attention to the collection categories unique to Welsh. This chapter aims to highlight specific differences between Welsh and English that may have

important effects on the way speakers of each language conceptualise the world and acquire language.

Chapter 3 describes a study that addresses the differences between Welsh and English. The study explores the different noun type distributions across the two languages to identify and measure the main points of divergence between them.

Chapter 4 describes a study that examines the nature of mothers' input language to infants acquiring each language. To fully understand the acquisition of collection categories it is necessary to establish the nature of the language the child hears. The noun distributions of Welsh- and English-speaking mothers' speech to their children is explored in this study.

Chapter 5 describes a large-scale study that explores the effects of differences in language structure on categorisation and recognition. The study examines the categorisation of collections of novel objects presented with novel nouns and later recognition of these novel objects by Welsh- and English-speaking children and adults. Categorisation and recognition patterns across the two languages, as indicators of participants' interpretations of the novel nouns, are discussed.

Chapter 6 describes the final study that examines the way differences in language structure affects attention and memory for number. Verbal descriptions and judgement tasks are used to explore Welsh- and English-speaking children's and adults' attention to number as well as their short-term and long-term memory for number. These are discussed with particular emphasis on the collection categories distinct to Welsh.

Chapter 7 summarises the main findings of all four studies and discusses these in relation to theories of language acquisition and linguistic relativity.

Linguistic relativity and the acquisition of collective nouns

The relationship between language and thought

The relationship between language and cognitive processes, in particular categorisation, has been of considerable interest to social scientists in recent years. A growing body of research indicates that language carries information that may guide children's categorisation patterns.

Categorisation is a basic component of perception, cognition, language and behaviour (Coley and Medin, 1997; Lakoff, 1987). Research has shown that very young infants have the capacity to categorise the world in certain ways (Quinn and Johnson, 1997; Hayes, Slater and Brown, 2001). If categorisation is a basic human cognitive ability then perhaps universally people categorise the world in similar ways. Different languages may merely allow different labels for the same pre-linguistic categories.

It has been argued, however, that categories do not exist in the objective world; rather, they are subjectively learned as a result of linguistically encoded categories in the language we speak. For example, Leach (1964) states:

> I postulate that the physical and social environment of a young child is perceived as a continuum. It does not contain any intrinsically separate 'things'. The child, in due course, is taught to impose upon this environment a kind of discriminating grid which serves to distinguish the world as being composed of a large number of separate things, each labeled with a name. This world is a representation of our language categories, not vice versa. Because my mother tongue is English, it seems self evident that *bushes* and *trees* are different kinds of things. I would not think this unless I had been taught that this was the case. (Leach, 1964, p. 34)

The notion that culture, through language, affects cognitive processes has become known as the theory of linguistic relativity. Attributed to Hulmbolt, Boas, Sapir and Whorf, linguistic relativity postulates that a language affects the way its speakers think. In particular, the language we speak influences the way we categorise the experienced world. Language, thought and culture are so closely linked that a given language may be associated with a distinctive 'world view'. Speakers of different languages experience the world differently (depending on the demands of their culture) and such experiential differences are encoded in the language they speak (Gumperz and Levinson, 1996).

A stronger view, linguistic determinism, suggests a causal relationship between language and thought. A language is not only interlocked with thought but rather it determines the way its speakers conceptualise the world.

Linguistic determinism makes three claims:

1. Different languages utilise different linguistic categories

And

2. Linguistic categories determine aspects of conceptual representations.

Therefore,

3. Conceptualisations of the world differ across linguistic communities as a consequence of the languages they speak.

According to Gumperz and Levinson (1996) this rationale would hold under the weakest conditions. They state that 'if there is *at least some* aspect of semantic structure that is not universal, and *at least some* cognitive effect of such distinctive semantic properties, then there must be *at least some* systematic cognitive variation in line with linguistic difference' (p. 24).

Linguistic determinism has inspired many, and a considerable body of

research has examined this issue. For example, the way languages encode colour categories has been the subject of much research (see Davidoff, 2001 and Taylor, 1995 for discussion). Colour terms in one language do not necessarily correspond to colour terms in other languages (Lyons, 1968). Some languages have many colour terms while others have few. For this reason, it was asserted that categorisation of colour is arbitrary (Bloomfield, 1933; Gleason 1955) and that such arbitrary discrimination could be extended to other areas of human experience and behaviour. Categorisation and conceptualisation of the world may be an artifact of linguistically determined arbitrary distinctions.

To investigate this arbitrary categorisation of colour, Berlin and Kay (1969) presented speakers of different languages with an array of colour cards. Participants were asked to select the colour cards to be classed under the basic colour terms of their language. Participants showed variability, across and within languages, in their selection of colour cards to be classified under each colour term. However, when subjects were asked to identify the best examples of the basic colour terms in their language, variability disappeared and unanimity of best examples of colours emerged within and across languages. This led to the conclusion that colour categorisation is not arbitrary but that languages seemed to select their basic colour terms from a universal inventory of eleven focal colours. Heider (1971) also found in a series of experiments that there was a high degree of agreement between best examples of colour terms across languages. Colour terminology then does not seem to support the arbitrariness of linguistic categorisation but instead lends support to the notion that cognitive components such as perceptual salience underlie category formation (Heider, 1971). Languages label colour categories differently but do not determine colour categorisation.

Other research has argued that categories are indeed language-dependent.

Many cross-linguistic studies have been conducted to examine the way speakers of different languages categorise the world.

Among these, Bowerman and colleagues (Bowerman, 1996a; 1996b; 2000; Choi and Bowerman, 1991; Choi, 1997; Choi, McDonough, Bowerman, and Mandler, 1999; Bowerman and Choi, 2001) have conducted numerous cross-linguistic studies to explore the relationship between language and categorisation. They have demonstrated that different languages encode spatial relations of objects in different ways. The spatial relations associated with English prepositions such as *in*, *on*, *over*, *under*, *out* and so on do not necessarily correspond with spatial categories found in other languages. Different languages adopt different linguistic devices to express spatial relations. For example, Korean uses specific verbs to express spatial relations that lead to differences in the way these spatial relations are categorised. Korean speakers categorise spatial relations in terms of the 'closeness of fit'. A specific set of verbs distinguishes between interlocking or close fitting items (*kkita)*, loose fitting items (*nehta*) and placing on flat surfaces (*nohta*). Thus, a distinction is made between 'putting a ring on a finger' and 'putting a cup on the table'. In English, there is no such distinction. The preposition *on* allows these spatial relations to be categorised as the same. Does this mean that speakers of different languages have different cognitive representations of these spatial relations, or do different languages merely use different linguistic devices to express the same representations?

Choi and Bowerman (1991; Choi, 1997, Bowerman, 1996a) explored this issue by examining English- and Korean- speaking 2- to 3;6-year old children's spontaneous and elicited speech. Children's descriptions of a variety of actions were elicited through play. Actions included putting objects into and out of tight or loose containers, placing objects onto other objects, opening and closing objects and so on. The results showed that the children categorised more like adult speakers of their own

language than like same-aged children acquiring different languages. These findings suggest that children do not start out with a universal set of spatial categories but that spatial categories are language specific.

A further study by Choi et al (1999) employed a preferential looking paradigm to examine whether English- and Korean-speaking infants distinguish between close fitting and loose fitting items. Infants were shown four pairs of videotaped actions of putting objects in different spatial relations to other objects. Two pairs of actions were conflated in that the two languages did not distinguish the actions. These actions involved containment and tightness of fit (e.g. 'putting a book into tight fitting case' [in + tight] versus 'putting a book on top of another' [on + loose]). The other two pairs of actions were split pairs in that containment and tightness of fit were split for each pair (e.g. 'putting a plastic ring in a basket' [on + loose] versus 'putting a plastic ring onto a pole' [in + tight]). Infants were given the target word in a sentence in each language to direct the infants' attention (i.e. *in* for English and *kkita* for Korean). For the conflated pairs it was expected that the infants from both language groups would look at the same scene, the one depicting *in* for English and *kkita* for Korean. For the split pairs, it was expected that Korean-speaking infants would look longer at the scene depicting tightness of fit and that English-speaking infants would look longer at the scene depicting containment. The results confirmed the predictions. Infants did look longer at the scenes that matched the target word of the infant's language. Young infants between 18 and 23 months were able to respond in ways consistent with their language-specific categories. English-speaking children notice that containment is necessary for *in* and Korean children notice tightness of fit is necessary for *kkita*. Children appear to quickly and easily categorise spatial relations in ways that reflect the language they are learning (Bowerman 1996, Bowerman and Choi, 1991; 2001).

Lucy (1992a, 1992b, and 1996) also addressed language effects on categorisation. He compared the way American English and Yucatec Mayan speakers attended to and categorised objects. In English, nouns generally fall into two classes, count nouns and mass nouns. Count nouns e.g. *books*, *tables*, and *hands* are used to refer to discrete countable entities. Mass nouns e.g. *water*, *food* and *air* are used to refer to continuous entities. When talking about several entities that are labelled with a particular count noun in English, it is obligatory to use a plural marker. In contrast, few nouns in Mayan require an obligatory plural marker. They do, however, require noun classifiers when quantifiers are used. Essentially, nouns (with exception of humans, animals and some artefacts) refer to material and require individuation by noun classifiers. For example, in English one can say "two candles" yet in Yucatec it would be necessary to give a measure of the material, " two long, thin wax". According to Lucy (1992a) the "basic English pattern overtly distinguishes plural from both singular and neutral in the noun phrase" (p.39) while the "basic Yucatec pattern is to disregard number, and most lexical noun phrases are neutral in number" (p.55). Given such differences between the languages, Lucy (1992a, 1992b, and 1996) attempted to examine the effects of such differences in a series of studies.

Initially, Lucy (1992a) examined the cognitive significance of number for speakers of American English and speakers of Yucatec Mayan. The aim was to explore whether cognitive sensitivity to number would follow the number marking systems used by the two language groups. The fact that English requires pluralisation of nouns may lead English speakers to notice the number of items more than Yucatec Mayan speakers. English and Yucatec both pluralise nouns for animate objects. English pluralises nouns for implements but this is optional in Yucatec. Neither language pluralises nouns for substances. Lucy (1992a) expected that English speakers would pay more attention to, and be more sensitive to changes in number

than Yucatec Mayan speakers. This would be especially noticeable for implements since this is the main point of divergence between the two languages.

In a series of 5 tasks Lucy (1992a) used sets of line drawings specifically designed to include target items varying in number. In the first task, participants were presented with 3 pictures individually and asked to describe each picture in turn while the picture was in view. The second task again involved 3 pictures; each of these were shown individually for a limited time. Following a short break after each presentation, participants were asked to describe each picture in turn without the picture in view. For the third task, participants were presented with 3 pictures individually for a limited time. Following the presentation of each picture, the four variant pictures associated with the original were presented. These variant pictures differed from the original in the numbers of target items in the pictures. In the absence of the original picture, participants were asked to determine which of the four variant pictures was most similar to the original. The fourth task involved presenting 3 pictures individually for a limited time. Following a short interval after each presentation, a set of 5 pictures was presented. Each set consisted of the original picture and 4 variant pictures. Participants were then asked to select the original from the set. For the final task, participants were shown the 3 pictures they had seen in the first task. These pictures were presented individually for a limited time (either 30 seconds or 60 seconds). Following a long interval after each presentation, they were again asked to select the original from each set of 5 pictures. Lucy (1992a) measured the instances of mention of the target items and the instances where the number of target items were mentioned. In addition, the precise differences between original pictures and those selected by the participants were observed to give a measure of sensitivity to changes in number.

The findings of these tasks showed that, overall, English speakers indicated

number much more often then their Yucatec counterparts. The results showed that both language groups frequently indicated number for animate objects, and neither group indicated number for substances very often. However, English speakers indicated number much more often than Yucatec Mayan speakers for implements. Similarly, both language groups were sensitive to changes in the number of animals and neither language group was particularly sensitive to changes in the number of substances. But, English speakers were much more sensitive than Mayan speakers to the changes in the number of implements. The differences that emerged between the two language groups suggest that attention to and memory for number is influenced by the grammatical structure of the language. Although Lucy (1992a) recognised methodological weaknesses and limitations of interpreting his study, he concluded that the overall pattern of results "strongly implicates grammatical structure as the operative factor influencing cognitive activity in these tasks"(p. 136).

A further study by Lucy (1996) involved presenting speakers of American English and Yucatec Mayan with objects that were similar in shape (e.g. hollow tubes) but made of different materials, and objects that differed in shape but were made of the same material (e.g. cardboard boxes). Participants were also given a target object that shared the shape and material of these objects (e.g. hollow tube made of cardboard). Participants were then asked to sort the objects into groups. The results showed that the target objects were predominantly grouped with similar-shaped objects by the English speakers, but grouped with objects of similar material by the Yucatec Mayan speakers. These findings indicate that material or substance has a greater significance for Yucatec Mayan speakers than for English speakers, at least for the purpose of categorisation. This is consistent with the differences between the two languages. Again, these findings are in keeping with the idea that language structure has an important relationship with the way speakers of different languages

experience and interpret the world.

Mazuka and Friedman (2000) replicated Lucy's study with Japanese and English monolinguals and bilinguals. Like Yucatec, Japanese is a noun classifier language and using the same methodology as Lucy, Mazuka and Friedman (2000) examined Lucy's hypothesis. They found that both Japanese and English speakers classified objects on the basis of shape more often than material composition when the participants were comparable on cultural and educational background. They failed to replicate Lucy's findings and suggest that the differences he identified between Yucatec and English may stem from the cultural and educational differences of the experimental groups.

Further research carried out by Slobin (1996) led to a more conservative view of linguistic relativity. He suggests that languages affect the way speakers conceptualise the world only for the purpose of communicating. That is, speakers are restricted to the particular grammaticized notions highlighted by the structure of their language. It is not that speakers of different languages conceptualise the world in different ways. Instead, they are forced to conceive of some things in a particular way so that they can express their thoughts through language. Slobin (1996) argues that it is better to think of 'world views' that are distinct for different languages as "thinking for speaking".

A cross-linguistic study carried out by Slobin (1996) examined children's narrative strategies of picture stimuli. The picture stimuli were shown to English-, Spanish-, Hebrew-, Turkish- and German-speaking children. The pictures depicted events that, when described, required different aspects of the scene to be noticed by speakers of different languages. In order to describe a scene in a given language, children are constrained by the grammatical requirements of their language. Thus, the scenes were talked about in different ways by speakers of different languages as a

result of the grammatical properties of the language in question. The children used strategies consistent with the encoding experience in their language when talking about the scenes. In conclusion, Slobin argues that languages "are not neutral coding systems of an objective reality. Rather, each one is a subjective orientation to the world of human experience, and this orientation **affects the ways in which we think while we are speaking.**" (p.91).

Kay (1996) goes further by suggesting that there are many ways of expressing a conception even within a language. Speakers of the same language use different linguistic strategies to communicate thoughts about the same event. Variation across languages does not constitute different 'world views'. Instead, the variation, even within a language, symbolises a distinct 'world view' for each individual speaker.

Research investigating the relationship between language and thought provides evidence both in favour of, and against, the notion that language influences the way speakers think about the world. Despite such a concerted effort to determine the role of language on cognitive development, very little is understood about the true nature of such a relationship or the mechanisms involved.

Since language structure may influence categorisation for language learning, understanding the processes involved in categorisation is paramount and involves understanding the underlying structures, how they are formed, and how they interact with each other.

Categorisation is the process by which the human mind organises and associates information about the world into category structures that facilitate and integrate knowledge and experience. Categories emerge from the mental process of classification. They are "stored in our minds as concepts and signalled by the words of a language, so one might come to think that they are equivalent to the meanings of these words" (Ungerer and Schmid, 1996, p.19).

However, some members of some categories are better examples or more representative of its category. Rosch (1973) adopted the term *prototypes* to represent these 'best examples'. Boundaries between categories are fuzzy because category members can be graded by typicality of that category (Heider, 1971; Lakoff, 1987; Reeves, Hirsh-Pasek and Golinkoff, 1998; Rosch and Mervis, 1975).

There are several proposed explanations about the way prototypes are formed. Some prototype categories, such as focal colours, may emerge from inherent perceptual properties while others are formed by frequency of exposure. However, these explanations do not adequately explain most categories. Instead, prototypes may be the average of all the attributes of members of the category or emerge as a result of culturally salient attributes. All of these factors may contribute to prototype formation. The important point to note is that prototypes are cognitively efficient ways of categorising the world and may reflect the nature of the human cognitive system (Taylor, 1995).

Underlying cognitive systems may affect not only category formation but also the relations between categories. Categories are organised in a hierarchical structure with differing relations between category members. The underlying principle of this hierarchical structure is that of class inclusion (i.e. the superordinate class includes all items on the subordinate level). Basic level categories are at an intermediate level of a general-to-specific hierarchy. 'Generalisation proceeds "upwards" from the basic level and specialisation proceeds "downward"' (Lakoff, 1987: 13).

Brown (1958) observed that objects have many names from subordinate to superordinate labels and that a particular label from a particular level of categorisation had a 'superior status'. He proposed that such labels are morphologically shorter and used more frequently. When asked to describe events or objects, people invariably

used basic level terms.

Further evidence for the salience of basic level categories in natural linguistic taxonomies emerged from a study by Berlin, Breedlove and Raven (1974) on Tzeltal plant classification. They found that classification systems focus on the basic level (or "folk-generic" level) of categorisation.

To account for this centrality of basic level categories Rosch (1978) argues that categories at the basic level provide the greatest amount of information with minimal cognitive effort. Categorisation allows the endless variation in the world to be divided into manageable chunks. The degree of information provided would depend on the size of the chunks (level of categorisation). For example, if the world were cut up into very small chunks (low level subordinate categories including individual instances such as FIDO) this would put considerable strain on cognitive systems such as memory. Large chunks (high level superordinate categories such as *animate*) would not provide enough information to interact with the physical world. Basic level categories then, reflect a cognitive economy in organising information about the world, by achieving a balance between the cost of providing maximum information and the benefits of minimal cognitive effort (Craig, 1986; Lakoff, 1987; Rosch, 1975; Taylor, 1995).

The different levels of categorisation are expected, for the most part at least, to be universal across cultures because all humans share the same general cognitive capacities. But what factors allow category members to form a cohesive class? Murphy and Medin (1985) criticise current accounts of categorisation because they do not adequately account for why objects are grouped together to form a category. They propose that people's theories and knowledge of the world make a major contribution to conceptual coherence. Knowledge often imposes category cohesion even when similarities between members are low. Relying on similarity relations

alone is insufficient to provide a theory of concepts.

This is especially pertinent to collection categories. What category relations allow collections of objects to be conceptualised as one cohesive group? Categories such as TREES are basic level categories while FOREST is a collection category that is similar to a superordinate category. But the structural relations between basic level categories within collections differ from those of class inclusion.

Markman (1978) argues that collection categories form more natural and more psychologically viable wholes than classes. Collections have a more literal part/whole relation whereas the part/whole relations of classes are more abstract. According to Markman and Siebert (1976) classes differ from collections in three fundamental ways. Firstly, collections are organised into part/whole relations (e.g. trees are parts of forest), classes are organised into class inclusion relations (e.g. roses are examples of flowers). Secondly, members of collections must be related to each other (e.g. exist together in spatial proximity), members of classes do not need to relate to each other. Thirdly, structural relations between members allow collections to have a greater coherence between members than classes. Therefore, Markman and Siebert (1976) argue that collections easier to conceive as organised wholes than classes.

Markman, Horton and McLanahan (1986) investigated the way children organise hierarchical relations between individuals and their associated class or collection. Children aged 6 to 17 years were taught novel class inclusion hierarchies. In one condition, children were given ostensive definitions (e.g. "these are trees" and "these are oaks"). In the other condition, they were given additional information (e.g. "oaks and pines are kinds of trees"). Given this additional information, even the youngest children were able to correctly interpret the relation as class inclusion. However, when only ostensive definitions were given, children up to 14 years would

incorrectly impose a collection structure instead of a class inclusion structure on the hierarchy.

Smith and Rizzo (1982) also examined 4 and 5-year-old children's understanding the distinct referential properties of collective and class nouns using part-whole comparison tasks. They demonstrated that children understood the hierarchical structure of both collective and class nouns. The children seemed to know that a superordinate class noun correctly labelled a set and any subset, and that a collective noun correctly labelled a set but did not correctly label a part of that set. Smith and Rizzo (1982) argue that children's knowledge of the hierarchical relationships between nouns and the sets to which they refer plays an important role in children's failure with class-inclusion task and their success when superordinate nouns are replaced by collective nouns.

If it were easier for children to form collection categories than classes, then learning labels for collection categories should be easier than classes. It follows that children should acquire collective nouns more easily than for class nouns because collections are easier to conceptualise. Research has indicated, however, that learning the names for collection categories is difficult for children. This suggests that children find conceptualising objects as collections difficult.

Learning the meanings of words or learning the names for categories

Words are the names for conceptual categories. Learning the meanings of words then involves categorisation. If language influences categorisation, then it follows that language influences the way children learn new words. Children's initial interpretation of a new word may be influenced by the structure of the language they are acquiring.

Children learn what a word means despite the infinite amount of environmental information available to them. For this reason, a number of theories have emerged to explain the success children have at overcoming the induction problem. Many theories have proposed innate constraints or biases. Children are born with specific capacities or tendencies that guide the way they learn new words. Markman (1994) argues that children are equipped with three types of constraints that facilitate and guide early language acquisition. These constraints are: the whole object assumption (directs the child to interpret new words as referring to objects as a whole rather than parts, substance, or other properties (e.g. colour, smell, and texture)); the taxonomic assumption (directs children to extend words to objects/entities of the same kind); and the mutual exclusivity assumption (directs the child to avoid having more than one label for the same object). Markman (1994) reports that all three constraints are available to infants by the time the naming explosion emerges, and that constraints function as an entering tool into language acquisition. These constraints however, do not explain how children acquire names for collections. A whole object bias would allow a child to interpret a novel name as referring to the individuals of a collection while the mutual exclusivity assumption would not allow names for collections (e.g. forest) to refer to the same objects as plural names for the individuals (e.g. trees).

Golinkoff, Mervis and Hirsh-Pasek (1994) propose additional principles (e.g. categorical scope, novel name – nameless category (N3C) and conventionality) that children use to overcome the induction problem. These constraints or principles are believed to allow the language learner to narrow the possibilities of what a new word might mean, thereby functioning as a channel into the semantics of the language. Landau, Smith and Jones (1988) found that children extend novel nouns to objects of similar shape, suggesting that children use object-shape to guide their word learning.

Bloom (1994; 1996; 2001; Bloom and Markson, 1998; Diesendruck and Bloom, 2003) criticises the notions of constraints or biases for word learning because children rapidly acquire words that violate the principles of these constraints. In particular, some nouns can refer to non-individuals such as collections of objects (e.g. 'forest') or to parts of objects (e.g. 'foot'). Instead, Bloom (1994; 1996; 2001) posits that people are innately equipped with syntax-semantics mappings that act to constrain inferences about the meaning of a new word. He argues that people possess general conceptual categories such as INDIVIDUALS, KINDS OF INDIVIDUALS and KINDS OF NON-INDIVIDUATED ENTITIES (stuff) which correspond to syntactic categories. Children use the syntactic category of a word as a cue to what a word means. Syntactic categories include count nouns (e.g. dog) that refer to discrete countable entities and mass nouns (e.g. sand) that refer to continuous, non-countable entities. Thus, according to Bloom (1994), noun phrases (e.g. 'a dog') are construed as referring to individuals, count nouns (e.g. 'dog') are construed as referring to kinds of individuals, and mass nouns (e.g. 'sand') are construed as referring to kinds of substances.

Biases arise as a result of the relationship between grammatical categories (e.g. noun phrases, count nouns and mass nouns) and abstract semantic categories. Bloom (2001; Diesendruck and Bloom, 2003) claims that a shape bias may emerge as

a consequence of count nouns referring to kinds of individuals and whole object biases result from noun phrases. Recent research has indeed indicated that knowledge of lexical form class does contribute to the process of learning the reference of a new word (Hall and Graham, 1999).

Bloom and colleagues (Bloom, 1994; 1996; 2001; Bloom and Keleman, 1995; and Bloom, Keleman, Fountain and Courtney, 1995) argue that by 3 years of age, children use count noun syntax to learn new words, and at 2 years of age, children begin to show sensitivity to the grammatical distinction between count nouns and mass nouns. However, collective nouns are count nouns that refer (in singular form) to collections. Bloom and colleagues conducted a series of experiments to determine what it is about collection categories that make them possible individuals so that children can learn the count nouns that refer to them.

Initially, 4- and 5-year-olds and adults were shown three piles of four objects in a row. The name for each pile were presented in either a singular (e.g. "this is a fendle..this is a fendle.. and this is a fendle") consistent with collective nouns (e.g. *forest*) or plural context (e.g. "these are fendles...these are fendles.. and these are fendles") consistent with object names (e.g. *trees*). Subjects were then tested on what they thought the meaning of the novel word was. The results showed that adults and 5-year-olds were highly sensitive to syntax (i.e. view 'fendle' as collection) but the 4-year-olds were not. With increasing age, more collective responses were given only for the singular condition.

Bloom et al (1995) explain the poor performance of the 4-year olds by claiming that the nature of the stimuli (i.e. unnatural or poor candidates for collections) did not motivate young children to treat the groups of objects as individuals. It was proposed that one reason for this is that collective nouns refer to superordinate categories and require a higher level of abstraction. To learn the

meaning of a superordinate level name, children must have already learned the meanings of the names for the basic level categories that make up the superordinate category. For example, to learn the meaning of *army*, children must already know what *soldier* means, and that *army* refers to a group of *soldiers*. Adults, however, can infer from the syntax that the novel nouns are intended to be interpreted as referring to a collection.

What are the properties of collective nouns that allow them to be learned by children and adults? Bloom et al (1995) proposed that the Gestalt principles of grouping play a role. It was assumed that objects in close proximity (such as found with referents of collective nouns e.g. 'packs' or 'flocks') would allow collections to be thought of as individuals. A study by Bloom (1998) was designed to test the Gestalt principles of grouping incorporated the proximity of the objects in the test groups. Adults were shown 12 novel objects in 3 different piles. The subjects were told 'these are fendles" without pointing to any of the piles. Here the syntax did not cue the subjects to interpret the objects in any particular way. The findings indicated that natural groupings were not sufficient to cause people to treat collections of objects as individuals. Although subjects perceived the objects as falling into discrete groups, this did not lead them to treat the groups as individuals. A possible explanation for this was that the objects in each group had no relation to each other that would distinguish them for the other groups.

This led to the idea that to interpret groups of objects as individuals requires some form of explanatory motivation to do so. Bloom (1998) tested this idea in a third experiment. Groups of stimuli were presented to adult subjects on a computer screen whereby each group was shown as a single moving unit. Again, the scene was described as "these are fendles" which did not cue the subjects to think of the stimuli in any particular way. They found that 73% of responses with moving groups were

collective responses. When the same stimuli were presented as static groups only 16% of responses were collective. These results suggest that objects viewed as having common movement does lead people to see the collections as individuals. Causal factors do seem to motivate people to treat groups of objects as individuals.

These findings led to an experiment to test the Gestalt principles of common fate. In this study, adult subjects were presented with the same groups of objects but a circle surrounded each group. The scene was described as "these are fendles on plates" (where plates explain the common movement of objects). It was expected that subjects would continue to interpret the novel nouns as referring to individual groups if the motivation to do so was common fate (i.e. moving as single unit). However, only 13% collective responses were given in this case. Thus, common fate is not sufficient to construe groups as individuals. However, is common fate necessary to construe groups as individuals?

To investigate this Bloom (1998) presented static groups of objects depicted as targets of other entities (machines). The aim was to make groups of objects salient as individuals. In this instance the count nouns were interpreted by the participants as collective nouns. This suggests that even static groups can be thought of as individuals under some conditions.

Bloom (1998) conducted a final experiment to determine the effects of intentions of the experimenter on the way subjects interpreted novel nouns. Here groups of five objects were placed next to each other. Adults, 4- and 5-year-olds were told either "the name for this is fendle" (neutral syntax) or "this is a fendle" (singular). The first statement can be construed as an object name, a collective noun or a proper noun. The second statement can only be construed as a collective noun. For half of the subjects, the objects were placed haphazardly into groups, but for the other half, the objects were carefully and slowly placed on the table to give the

impression that the groups had a precise structure. It was expected that more collective responses would be given when the syntax could only be construed as collective and when the intention of the experimenter was that the group should be treated as an individual.

The adults confirmed the prediction giving more collective responses for the singular syntax, especially when the objects were carefully laid out before them. However, for the children, neither the syntax nor intentionality had much effect on their interpretation of novel nouns as collective. Bloom et al (1995) suggest that one possible explanation for why children did poorly on this task is that children have a weaker conception of INDIVIDUAL in that the idea of INDIVIDUAL does not consist of items that are individuals by the intention of others. How do children learn collective nouns? In summary, Bloom and colleagues advocate the role of increasing sensitivity to syntax, perceptual salience of groups and intentionality in the acquisition of collective nouns.

Further studies by Huntley-Fenner (1995) also investigated the way children acquired collective nouns. He asked children to count objects they saw in pictures, for example, a picture depicting three people holding two balloons each. Children were asked either 'how many balloons do the people have?' or 'how many balloons does the family have?' There are two possible answers to the former question, a distributive response e.g. 'two balloons each' or a collective response e.g. 'six balloons in all'. For the latter question, only one possible answer is correct, a collective response e.g. 'the family has six balloons'. The results showed that only adults knew that the collective noun (e.g. 'family') required a collective response. In a second experiment, children were asked to choose a picture from an array of 3 choices to give to a puppet. Each set of 3 pictures included 2 pictures of different single items and a picture of a group of one of the single items (e.g. a single tree, a

group of trees and an anchor). The results indicated that when probed with a collective noun (e.g. 'can you give the puppet one forest?') subjects preferred to select a single familiar object than a group. This suggests that subjects were unwilling to choose a group as a possible candidate for the meaning of novel nouns. In conclusion, Huntley-Fenner (1995) argues that 'collective nouns are not easy to conceptualise or learn' (p.153). And that 'mere sensitivity to syntax does not give the word learner enough information about the kinds of things the speaker might be referring to.' (p.153). The findings support the idea that object kinds are especially salient as candidates for the meaning of new count nouns.

A fundamental weakness with the experimental evidence outlined above is that only English-speaking individuals were tested. To argue that learning collective nouns is difficult for children and that interpreting novel nouns as collectives require specific conditions may apply only to the English-speaking populations and not universally across languages. This poses considerable problems for Bloom's (1994) account of innate syntax-semantic mappings. If children are equipped with particular categories and that children learn the meanings of new words from the syntax-semantics mappings then the patterns found for English speakers should hold across speakers of other languages. A major criticism arises over the lack of cross-linguistic evidence to support Bloom's (1994) ideas of innate links between syntax and semantics.

Ravid and Hayek (2003) investigated the development of collective forms in Palestinian Arabic. Palestinian Arabic has singular, dual, plural, and collective forms. Ravid and Hayek (2003) examined the language production of these different forms by children aged between 4 – 8 years. Following a training session, participants were shown pictures with different numbers of objects to elicit verbal responses. The experimenter showed the picture with one object (e.g. orange) and said "this is one

orange (*burda:ne*) ". Then showing a picture with two, three, or a collection of objects (i.e. oranges) the experimenter said "when there are three we say...". They found that there was no effect of age on correctly producing the collective form. Collective forms were used inappropriately more often when a plural response was expected than when a dual response was expected but this was not affected by age. Although children appeared to show a strong learning curve with regard to dual and plural forms, reaching 85% success, this was not the case with collective forms. Even the oldest children did not achieve greater than a 50% success rate in producing a correct collective form. Collective nouns in Palestinian Arabic then seem to be difficult for children to acquire.

Ravid and Hayek (2003) suggest that because forming a collective noun requires dropping the final vowel of the singular form and that collective nouns are restricted to a small class of nouns, then collective nouns need to be learned lexically. Moreover, the results showed that plural forms were often used in collective contexts suggesting a tendency for regularisation of reference to multiple entities. Ravid and Hayek (2003) suggest one possible explanation for why collective nouns were difficult for children to acquire on the one hand, yet seemed to be viewed as belonging to a common set of plural forms on other, is that Palestinian Arabic collectives are semantically distinct from the collective nouns examined in the literature. For example, collective nouns such as *family* and *army* refer to single bound entities, and these nouns can be pluralised to refer to several collections. Palestinian Arabic collective nouns, which are not pluralised, also refer to bound wholes but the whole consists of units with the same name, albeit morphologically modified, and therefore refer to the same units as the singular forms. Thus, Palestinian Arabic speakers may view collective nouns as alternative plural forms. In

conclusion, Ravid and Hayek (2003) suggest that different languages may have semantically different collection categories.

Other recent cross-linguistic research (Gathercole, 1997; Gathercole and Min, 1997; Gathercole, Thomas and Evans, 1999; Gathercole, Thomas and Kim, 1999; Imai, 1999; Martinez and Shatz, 1996) has indicated that children acquiring different languages show patterns of learning novel nouns that are consistent with differences in the languages being learned.

Gathercole, Thomas and Evans (1999) report that English-, Welsh- and Spanish-speaking children respond differently to new nouns in ways that are consistent with differences in the language being learned. The structure of English and Spanish emphasises individuation whereby nouns often refer to whole objects with clear singular/plural reference. In Welsh, however, the singular/plural distinction is not as clear-cut and many nouns refer to collections. Given that the structure of Welsh highlights collections more than English or Spanish, it was expected that Welsh-speaking children would be more likely than English- and Spanish- speaking children to interpret the meaning of a new noun as referring to collections. In a series of experiments, Gathercole et al (1999) presented children between 2 and 4 years of age with novel nouns in a storybook context. The character in the book, and her bear, are looking for objects given a novel name (e.g. 'Ellen is looking for her blicket'). The children were asked to select an object from an array to give to the bear (e.g. 'can you give the bear his blicket?'). The arrays of objects the child could chose from took the form of one versus many objects, a part versus many parts, and a small part versus a big part. The results showed that Welsh-speaking children were more inclined to interpret a new word as referring to collections of objects than either the English- or Spanish-speaking children while Spanish-speaking children were more inclined to interpret new words as parts of objects than either

English- or Welsh-speakers. These findings were consistent with the predictions that the structures of the languages being learned would affect the ways in which children acquired novel nouns.

These findings conflict with the position proposed by Bloom and colleagues that children have difficulty with interpreting new words as collective nouns because they have low sensitivity to syntactic cues. Since a collection is a kind of individual, the syntax-semantics mapping should allow children to interpret collections as whole individuals. The children in the studies conducted by Bloom et al (1995) did poorly on interpreting a new noun as a collective noun. Children acquiring Welsh show a greater tendency to interpret new nouns as collections than English-speaking children do. This finding highlights that the difficulty experienced by the children in Bloom's studies may reflect the structure of English. The structure of Welsh might allow Welsh-speaking children to accept new words as referring to collections more than English-speaking children.

In contrast to Bloom's (1994) ideas of innate syntax-semantics mappings many researchers lean toward a learning account of language acquisition and provide evidence that challenges notions of innate language faculty. Gathercole et al (1999) postulate that word meaning acquisition arises from learning to coordinate multiple cues to meaning. They argue that information in the input guides children to make certain inferences about word meaning and that language acquisition is a matter of learning to coordinate the numerous sources of information about word meaning. Languages with distinct characteristics, such as Welsh, can provide valuable information about the way language affects categorisation and language acquisition. Theories of language acquisition must account for the way children learn different languages, which may be directly influenced by the structure of the language being learned.

Summary

A growing body of evidence supports the notion that language structure does indeed guide the cognitive processes necessary for language learning. Numerous cross-linguistic studies have suggested that children acquiring different languages show patterns of cognitive and linguistic behaviour that are consistent with the structure of the language they are learning (Bowerman, 1996a; 1996b; 2000; Bowerman and Choi, 2001; Choi and Bowerman, 1991; Choi, 1997; Choi et al, 1999; Lucy, 1992a; 1992b; 1996; Gathercole, 1997; Gathercole and Min, 1997; Gathercole et al, 1999; Imai, 1999).

A key cognitive process involved in learning new words is categorisation. That is, to know what a word means it is necessary to determine what the word refers to. Given that Welsh speakers have more options about the reference of a word, does this influence the process of categorisation for word learning? Do Welsh and English speakers categorise entities in ways governed by the way each language refers to entities?

Differences in language structure may have important effects on cognitive processes, and in turn, language acquisition. Some research has suggested that children find learning words for collection categories especially difficult. However, the research focuses mainly on English-speaking children and adults and the difficulty in forming collection categories may merely reflect the grammatical characteristics of English. Given that Welsh has a distinct system that allows for the conceptualisation of collections, Welsh-speaking children may learn names for collection categories more easily than English speakers do.

Language Structure: Number marking in Welsh and English

One of the main purposes of this thesis is to explore whether the grammatical properties of a language being learned influences the way children and adults think about objects and ultimately learn the meanings of new words. More specifically, Welsh has a system for expressing number that differs from English and many other languages. This chapter aims to identify and discuss the grammatical characteristics of Welsh that differ from English, in particular collection categories. Since Welsh grammar relies heavily on inflectional morphemes, a brief introduction to inflectional morphology and number marking will be presented, followed by a more detailed description of the inflectional morphology of number marking in Welsh. An account of the syntactic properties of the Welsh number system and how this differs from English will conclude this chapter.

Inflectional morphology and number marking

This thesis is primarily concerned with inflectional morphemes used for number marking in Welsh and English. Morphemes are the smallest linguistic units that carry meaning (Bloomfield, 1933; Aronoff, 1994). Inflectional morphemes are affixes that carry linguistic information such as number, case, tense and aspect. They have meanings beyond the words to which they are bound and although inflections change the meaning of the root word they do not change the grammatical class of the word (Aronoff, 1994; Jensen, 1990; Spencer, 1991).

Jackobson (1968) applied the notion of a universal set of features to morphology whereby the meanings of morphemes, at least in part, are represented by a universal set of features. The values assigned to features are related to markedness in that positive values are marked while negative values are unmarked. Markedness in this sense refers to the frequency and rarity of the item in different languages,

marked being least frequent (Jensen, 1990). Thus, universally across languages, the most natural forms are the least marked. This often leads to the assumption that "certain morphosyntactic categories are 'simpler' than, or prior to, others. For example, 'singular' is a basic category, while 'plural' is in some sense derived. Hence, the natural way of signalling the plural is to take the form which conveys the singular and do something extra to it" Spencer (1991, p. 224). This process is iconic in that morphological additions represent semantic additions.

Carstairs-McCarthy (1992) also uses inflectional properties of singular and plural to highlight markedness or, conversely, naturalness. "Plural is more marked than Singular because it is textually less frequent and because in many languages Singular may be used freely in reference to sets or collectives whereas Plural is never freely used to refer to individuals" (Carstairs-McCarthy, 1992, p. 218) with the exception of pluralia tantum (e.g. *scissors*). It follows then that natural coding will lead to plurals being the overtly marked forms and singular being unmarked. The English language relies heavily on such a system (typically using plural *-s* to convey plurality). Although there are some irregular forms (e.g. *man > men, fish > fish)* they are still consistent with the singular/plural system (Spencer, 1991).

Moravcsik and Wirth (1983, p. 2) argue that "singular has a morphologically simpler expression and perhaps in no language is it morphologically more complex than the plural". But, "if the more complex category is represented by a simpler form, then we have a countericonic process" (Spencer, 1991, p. 224). In such cases "nouns which are most typically used to refer to a collection of objects may display additive inflection for a singular, or 'Singulative', form." Carstairs-McCarthy (1992, p. 219).

This last option is precisely the case in Welsh. Some nouns, in their unmarked form, refer to collections of objects. To refer to a single item from the collection, a unit inflection is used (e.g. *dail/deilen* 'leaves/(a) leaf', *offer/offeryn* 'instruments/(an) instrument') (Thomas, 1996). This can also be observed in other

languages such as Russian and Palestinian Arabic. In Russian, mass nouns that refer to collections of discrete entities such as *grass* or *grain* can take a singulative suffix *–in- (-ink-)* to refer to those minimal discrete entities (Wierzbicka, 1988). Collective nouns in Arabic are formed by dropping the final vowel of singular forms (Ravid and Hayek, 2003).

Traditionally, Welsh grammarians have treated the uninflected form as a 'plural' form and the inflected form as a 'singular' form (see discussion in King, 1993). If this were indeed the correct analysis, this would be an example of a countericonic language. The semantically more complex category of plurality is represented by a simpler morphological form than the semantically simpler category of singularity represented by a morphologically more complex form. Languages such as Welsh, Russian and Arabic pose a problem for theories of markedness because the 'plural' form, not the 'singular' form, should be treated as unmarked or more natural. Although accepting the notion of natural morphology, Wurzel (1984) acknowledges the oversimplification of inflectional morphology because of anomalies in some languages. Nevertheless, he argues that there is an assumption that inflectional systems are pressured to be congruent or regular.

Although it is tempting to view Welsh collection nouns as plurals and units as singulars, King (1993) warns against this for two reasons. First, to form singular from plural (as in *mochyn* from *moch*, 'pig' from 'pigs') contradicts the principles of a singular/plural system. And second "the relationship between the group and its individual components is neatly expressed only in the c/u [collection/unit] system" (p.49). This second point refers to the meaning or 'quality' of information provided by a collective term that is not conveyed in a plural form.

Thus, an alternative and more useful way to treat such forms is as members of a distinct system that is separate from the singular/plural system. King (1993) argues that much confusion arises over the Welsh number system because it is

misconstrued as operating on the same two-way system as English. Instead, he stresses that "Welsh has mutually exclusive twin systems" (p.48), a singular/plural system and a collective/unit system. These systems should be treated as having distinct properties and operations, as will be seen below.

Inflectional morphology and number marking in Welsh

Languages use many different linguistic devices to express number. Like many other languages, English adopts a singular/plural opposition with regard to nouns. The singular is the base form and the plural is the marked form. For the most part, English uses the plural *-s* inflection to denote plurality although irregular forms do occur such as those forms that require an internal vowel change (e.g. *goose/geese*) or no change (e.g. *sheep/sheep*).

The singular/plural system in Welsh operates semantically in a similar way to English and other languages. Plural forms in Welsh are formed by one of the following ways (Thomas, 1996; King, 1993; and Thorn, 1993):

(1) Singular nouns can take one of thirteen plural inflections (e.g. *llyfr* > *llyfrau* 'book > books', *hoel* > *hoelion* 'nail > nails', *bisced* > *biscedi* 'biscuit >biscuits').

(2) Internal vowel changes (e.g. *car* > *ceir* 'car > cars').

(3) Internal vowel changes in addition to one of the thirteen plural inflections (e.g. *mab* > *meibion* 'son >sons', *awr* > *oriau* 'hour >hours').

There are nouns in Welsh within the singular/plural system that occur as bound forms; the singular form takes a unit inflection while the plural form takes a plural inflection (e.g. *malw**en*** 'snail' > *malw**od*** 'snails', *oedol**yn*** 'elder' > *oedol**ion*** 'elders'). Some nouns can optionally add a unit inflection to the singular form (e.g. *llyfr* or *llyfr**yn*** 'book' > *llyfr**au*** 'books', **pel** or *pel**en*** 'ball' > *pel**i*** 'balls'). There are also nouns that have more than one plural form (e.g. *tref* /*trefydd* or *trefi*

'town/towns') and nouns that have more than one singular form (e.g. *dant* or *daint/dannedd* 'tooth/teeth'). See Appendix I for a more detailed description and examples of the number morphology of Welsh nouns.

Within the singular/plural system of English, some nouns in their basic form denote groups or collections of objects. These are collective nouns such as *army*, *family* and *forest*. Although the singular form refers to groups of entities, collective nouns have plural forms that denote several groups such as *armies*, *families* and *forests*. This is also true in Welsh (e.g. *byddin/byddinoedd* 'army/armies', *teulu/teuluoedd* 'family/families' and *tyrfa/tyrfaoedd* 'crowd/crowds') (Williams, 1980; Thomas, 1996).

As noted above, Welsh has an addition distinct collection/unit system. This will be referred to as the "cluster/unit" system (c/u) to distinguish these nouns from the collective nouns mentioned above. In the cluster/unit system, the unmarked forms refer to collections of entities. The majority of these cluster nouns refer to basic level categories (e.g. *coed* 'trees', *ser* 'stars', *dail* 'leaves') but some do refer to superordinate categories (e.g. *dillad* 'clothes', *offer* 'instruments/equipment' and *dodrefn* 'furniture').

Unit forms refer to individuals and are derived from the cluster forms by the addition of one of two unit inflections *–yn* (masculine) and *–en* (feminine), and in some cases an additional internal vowel change. For example, *moch* > *mochyn* 'pigs > pig', *llygod* > *llygoden* 'mice > mouse', *dail* > *deilen* 'leaves' > '(a) leaf', *dillad* > *dilledyn* 'clothes > (an item of) clothing', *cnau* > *cneuen* '(a) nut' > 'nuts', *blew* > *blewyn* 'hair/fur' > '(a strand of) hair/fur' (Thomas, 1996). Most cluster/unit nouns are feminine (King, 1993).

The semantic properties of cluster nouns in Welsh are similar to the semantic properties associated with mass nouns in English. A cluster noun in Welsh has a sense of a continuous, undifferentiated entity or substance. For example, *coed*

conveys a 'sense of homogeneousness' that cannot be neatly expressed by the English plural translation 'trees' but would be better translated as 'wood'. Likewise, 'foliage' is a closer approximation to the sense of homogeneity expressed by *dail* than 'leaves' is, (King, 1993).

Nouns that fall into the c/u system refer to entities that generally occur in large clusters (e.g. STARS, FUR, and HAY). These nouns frequently refer to living things that are predominantly associated with being in groups (e.g. TREES, PLANTS, FISH, and ANIMALS (particularly those living in groups and swarming/colonising insects), as well as other entities associated with these categories, (King, 1993; Thomas, 1996). Welsh treats the referents of cluster nouns as substance-like, which allows speakers to talk about them in ways that are relevant and meaningful. However, since the items that make up the collections are individuals rather than portions of substance, it is useful to have a way to individuate if needed.

Grammatical properties of different noun types in Welsh and English

In Welsh and English, different types of nouns can be identified by their grammatical properties. In particular, the way nouns are quantified shows distinctive patterns in the two languages. The main semantic and syntactic differences between the different noun types in the two languages are summarised in Appendix II.

There are different ways in which linguists distinguish different forms. There are two main approaches to understanding the form-meaning relations in grammar, notional definitions and distributional method.

Notional definitions of parts of speech rely primarily on semantic classes rather than syntactic behaviour. For example, nouns denote people, objects and places, while verbs denote actions or events. However, words of a given semantic class often fall into different categories. For example, the word *movement* refers to an action, yet it can be found as a noun (Croft, 2000).

The alternative approach, the distributional method, allows syntactic categories or parts of speech to be identified from the constructions in which they are used. There are several weaknesses with this approach to grammar. For example, the distributional method uses constructions (e.g. Noun phrases) to define categories in a language (e.g. nouns) but these categories, as basic elements of syntactic representations, are used to define constructions. This method is circular. Furthermore, constructions used in one language for defining categories may not exist or may differ in other languages (Croft, 2000). Although the distributional method is favoured by Croft (2000; 2001a; 2001b), he offers a more comprehensive theory of grammar (Radical Construction Grammar) whereby grammatical categories are derived from constructions but do not exist as universal syntactic categories in their own right. Rather, constructions, not syntactic categories and relations, are the basic elements of syntactic representations.

Wierzbicka (1988) has a more complex analysis of the meanings of different form classes. She identifies several classes based on the meaning that the noun class has within the intended message. Words that are traditionally classified in the same form class may show distinctive patterns that indicate the meaning a speaker wish to express. For example, "I ate too much cake" has a different meaning to "I ate too many cakes".

Sub-classification of nouns, in English, into count nouns (e.g. *books)* and non-count or mass nouns (e.g. *water)* has often been used in linguistic analysis because it is a useful way to 'explain the DISTRIBUTION of nouns in relation to the use of such ITEMS as ARTICLES and QUANTIFIERS' (Crystal, 1997, p.97). Count nouns denote what the language treats as separable, bound entities and mass nouns denote what the language treats as continuous entities. This is because the way in which a count noun is quantified differs from the way in which a mass noun is quantified (Crystal, 1997).

Count nouns usually have a singular and plural form. They can be preceded by the indefinite article *a/an*, and can usually take the plural *–s*. Plural forms can co-occur with numerals (e.g. *three books*) and co-occur with specific quantifiers (e.g. *many* and *few*). Mass nouns usually occur as singular forms, are not preceded by the indefinite article *a/an*, and do not take the plural *–s*. Mass nouns are not quantified directly by numerals but require a nominal measure noun (e.g. *three cups of water*) and occur with specific quantifiers (e.g. *much* and *little*). Some nouns can be used as both mass nouns and count nouns (e.g. *I had too much beer* or *too many beers*).

In Welsh, the way in which nouns are quantified is different: (1) there is no indefinite article in Welsh; (2) a single set of quantifiers is used with most noun types; (3) numerals occur with singular forms (e.g. *wyth llyfr* 'eight book')[1]; and (4) Welsh utilises a partitive system to quantify nouns (King, 1993; Thomas, 1996).

The partitive system allows plural forms to occur with numerals (e.g. wyth *o* lyfrau ' eight of books') and quantifiers such as *ychydig o* 'a bit of' and *llawer o* 'lots of' (e.g. *llawer o lyfrau* ' a lot of books'). Thus, numerals occur directly with singular nouns but occur with plural forms only if the partitive *o* 'of' is used (e.g. *dwy gadair* (two chair), *chwech o gadeiriau* (six of chairs) (King, 1993; Thomas, 1996).

Some nouns in Welsh only occur in singular form (e.g. *pren* 'wood' and *dwr* 'water'). These nouns occur with the same quantifiers as mentioned above and require the partitive *o* 'of' (e.g. *ychydig o ddwr* 'a bit of water'). Despite being singular, these nouns do not occur directly with numerals. Instead, these nouns require singular nominal measure nouns (e.g. *wyth tamaid o fara* 'eight piece of bread'). This is similar to what is found with English mass nouns except that English requires a plural nominal measure noun to quantify mass nouns with numerals (e.g. eight pieces of wood). A distinction between these noun types in Welsh can be made

[1] There are exceptions in English that adopt a similar pattern (e.g. 'He weighed ten *stone*')

on the basis of whether the nouns occurring with the partitive *o* 'of' are singular or plural (King, 1993; Thomas, 1996).

The syntactic properties of collective nouns (e.g. *family*) are identical to those of other singular/plural nouns in each language and their quantification is as specified above for singular/plural nouns.

Although cluster nouns in Welsh refer to collections, they are distinct from the collective nouns mentioned above. Examples include *adar* 'birds'/ *aderyn* '(a) bird', *ser* 'stars'/ *seren* '(a) star', *dillad* 'clothes'/ *dilledyn* '(an item of) clothing', *offer* 'instruments' / *offeryn* '(an) instrument'. These nouns seem superficially like collective nouns because they refer to collections, yet they do not share the same syntactic characteristics.

First, cluster nouns do not have a plural form, do not occur with numerals without the use of the partitive *o* 'of', and require a singular nominal measure noun for quantification. Note that these characteristics are similar to those of the singular only type in Welsh (e.g. *pren* 'wood') mentioned above. Compare (a) with (b)

(a) dau aderyn (unit noun) 'two bird'

dau o adar (cluster noun) 'two of birds'

*dau o adaroedd 'two of birds-Pl'

*dau adar 'two birds'

(b) dwy fyddin (collective noun) 'two army'

dwy o fyddinoedd 'two of armies'

*dwy fyddinoedd 'two armies'

Second, these cluster/unit nouns take a unit inflection to individuate an item from the collection. Again, examples include *adar* 'birds'/ *aderyn* '(a) bird', *ser* 'stars'/ *seren* '(a) star', *dillad* 'clothes'/ *dilledyn* '(an item of) clothing', *offer* 'instruments' / *offeryn* '(an) instrument'. Yet, like plural nouns, the quantifiers *ychydig o* 'a bit of' and *llawer o* 'lots of' occur with the form that refers to the

collection of entities, not with the form that refers to an individual entity. Similarly, the inflected unit form can occur with numerals just like singular nouns (e.g. *chwech mochyn* 'six pig'). The uninflected form can occur with numerals only in the presence of the partitive *o* 'of' (e.g. *chwech o foch* 'six of pigs').

According to King (1993) there are no 'hard and fast' grammatical rules as to the way a speaker chooses to quantify but suggests that numerals occur directly with singular/unit forms for lower numbers but with the partitive and plural/cluster forms for higher numbers (i.e. over 10).

These grammatical patterns in Welsh suggest that it may not always be clear from the form of a noun, or from the construction in which the noun occurs, whether the noun refers to an individual or a collection.

(1) An uninflected form could refer to either a single entity (singular nouns), multiple entities (plural nouns with internal vowel changes), or a collection (cluster nouns). For example, compare *ychydig o geir* (plural noun) 'a bit (few) of cars', *ychydig o foch* (cluster noun) 'a bit (few) of pigs' and *ychydig o ddwr* (singular only noun) 'a bit of water'.

(2) Different noun types can even have the similar basic form. For example, the noun *pysgod* 'fish' refers to a collection and is inflected for individuation (*pysgod<u>yn</u>* '(a) fish') while the noun *cysgod* 'shadow' is a singular noun and is inflected for plurality (*cysgod<u>ion</u>* 'shadows'). Similarly, the noun *gwenyn* 'bees' is a cluster noun and is unitised to *gwenynen* 'bee' while *menyn* 'butter' is a singular only noun.

The construct *ychydig o N* does not reveal whether the noun is singular or plural, or whether the noun denotes a substance or a group. A Welsh-speaking child hearing *ychydig o bysgod* ' a bit of fish' for the first time cannot necessarily tell from the construct whether the noun is a plural noun denoting many individuals, a singular

noun denoting substance or a cluster noun denoting a collection. In this instance the noun refers to a collection.

To gain a better understanding of the way these two languages allow speakers to categorise entities and ultimately learn new words, it is important to first establish the extent to which the two languages differ. The following chapter describes a study on the frequency of noun types and forms that was carried out to address this issue. The objective was to identify and highlight the main differences and similarities between Welsh and English.

Study 1: Frequency analysis of Welsh and English forms

The aim of this study is to determine the kinds of differences, and the strength of these differences, across Welsh and English. Since this thesis is primarily concerned with the differences in the way the two languages refer to collection categories, this study will focus on differences in the distribution of nouns only. The study involves a frequency analysis of the different noun classes occurring in the two languages. To achieve this a distribution of the most frequent nouns from each language was examined.

Nouns were classified into different noun classes to give a noun **type** frequency for each language. The number of times a noun of a given type occurred gives a **token** frequency. Type and token frequencies for singular, plural, cluster and unit forms were identified.

The goal was to identify the frequency with which distinct noun types, and their forms, occur in Welsh and English. A classification system was developed to categorise different noun types in a comparable way across the two languages.

Method

Sample

A sample of English nouns was obtained from the MRC Linguistic database of written texts (Francis and Kuçera, 1982). To ensure the sample was of a manageable size, the criteria of selecting only nouns that had a Francis and Kuçera frequency count greater than 10 was set. First, all obsolete and archaic nouns were excluded. This produced a list of 3143 nouns. Then proper nouns, numerals, and ordinals were eliminated from the sample because they were not considered relevant to the issues being address in this study. This left a sample of 2567 nouns to be examined.

A sample of Welsh nouns was obtained from a Welsh Language database of written text, Crondeb Electroneg o Gymraeg (Ellis, O'Dochartaigh, Hicks, Morgan, and Laporte, 2001). Initially, a list of all the nouns (26,591 in total) was established. Again, only nouns with a frequency count greater than 10 were selected. This produced a sample that included nouns designated in the database as noun person and noun place (i.e. proper nouns). These were eliminated from the sample leaving a sample of 2751 nouns to be examined.

Procedure

The frequencies of each form (singular, unit, plural or cluster) for each noun were recorded for the two samples. The Welsh language database did not distinguish the cluster/unit nouns (c/u) from the singular/plural nouns and consequently had frequencies for singular forms and plural forms. Once the nouns of the c/u type had been identified, the singular forms of this type were re-coded as unit forms and the plural forms were re-coded as cluster forms.

The nouns and their corresponding form frequencies from the Welsh and English samples were inserted into a database package. Each noun was systematically classified into one (or more) of six noun type classifications.

The noun type classes and their operational definitions were as follows:

a. Singular/Plural Nouns

Nouns were defined as singular/plural nouns if they satisfied two conditions:

(1) The noun occurs in both a singular and plural form. It takes a plural *–s* in English, and in Welsh, takes one of the 13 plural inflections, an internal vowel change or both. Irregular plural forms were included in this class if they satisfied criterion (2).

AND

(2) The nouns co-occur with a specific set of quantifier nouns in the two languages. In English the singular form can be preceded by *a /an*; and plural forms could be preceded by *many, few,* and numerals such as *two* or *three.* In Welsh, the singular form can be preceded by numerals such as *dau/dwy* 'two', *tri/tair* 'three'; and, plural forms can be preceded by partitives such as *dau/dwy o* 'two of', *tri/tair o* 'three of', and the quantifiers *ychydig o* 'a bit of', *llawer o* ' lots of'.

b. Singular only Nouns

Nouns were classified as singular only nouns if they satisfied three conditions:

(1) The noun occurs only in a singular unmarked form. The noun does not take a plural *–s* in English, or take one of the 13 plural inflections, an internal vowel change or both in Welsh.

(2) The noun does not normally occur directly quantified by numerals or ordinals. The nouns cannot be preceded by *two, three, first, second* in English and *dau/dwy* 'two', *tri/tair* 'three', *cyntaf* 'first', or *ail* 'second' in Welsh.

AND

(3) The noun co-occurs with specific sets of nominal measure nouns in the two languages. In English the noun can be preceded by *much, little, some*. In Welsh, they can be preceded by *ychydig o* 'a bit of', *llawer o* 'lots of'.

c. Plural only Nouns

Nouns were classified as plural only nouns if they satisfied two conditions:

(1) The noun occurs only in a plural marked form. The noun has a plural *–s* in English or any of the 13 plural inflections, internal vowel change or both, in Welsh. For example, the nouns *economics, mathematics* and *statistics* were classified as plural only.

(2) The noun does not normally occur directly quantified by numerals or ordinals. The nouns cannot be preceded by *two, three, first, second* in English and *dau/dwy* 'two', *tri/tair* 'three', *cyntaf* 'first', *ail* 'second' in Welsh.

d. Measure Nouns

Nouns were classified as measure nouns if they satisfied two conditions:

(1) The noun meets the conditions for classification as a singular/plural noun (see definition above).

AND

(2) The noun refers to a measure or an amount. The noun acts to quantify other singular/plural or substance nouns. In English, "a *pile* of books", " a *piece* of bread" and in Welsh, "*swp* o lyfrau" '(a) pile of books', "*darn* o fara" '(a) piece of bread'.

e. Collective Nouns

A noun was classified as a collective noun if it satisfied 2 conditions:

(1) The noun meets the conditions for classification as a singular/plural noun (see conditions above).

AND

(2) The referent of the unmarked (singular) form consists of multiple entities; the referent of the marked (plural) form is multiple sets of those multiple entities.

f. Cluster/unit nouns

A noun was classified as a cluster/unit noun if it satisfied 3 conditions:

(1) The referent of the unmarked (cluster) form consists of multiple entities.

(2) The noun can take one of two unit inflections (*–yn* [masculine] or *–en* [feminine]).

AND

(3) The noun cannot take a plural inflection.

Some nouns overlapped noun classes by satisfying some or all the conditions for more than one class. Those nouns that satisfied some or all the conditions for more than one noun type were classified into all the relevant classes. For example, nouns such as 'noise', 'cake' and 'plaster' satisfied the conditions for classification as singular/plural nouns and singular only nouns. Similarly, nouns such as 'pack' 'set' and 'bunch' satisfied the conditions for classification as collective nouns and measure nouns. The data for those nouns that were classified in more than one class are included in the results.

Results

Noun **type** frequencies were calculated as the number of nouns that were classified into each noun class. Type frequencies for Welsh and English are shown in Table 3.1.

Table 3.1 Noun Type frequencies in Welsh and English

Noun Types	**Welsh Type Frequency**	**English Type Frequency**	**Overall Type Frequency**
Singular(sg)/Plural(pl) Nouns	1966	1860	3826
e.g. *chairs/s*	(71.46%	(72.46%)	(71.94%)
Dual Function (singular/plural and singular only) Nouns	58	325	383
	(2.11 %)	(12.66%)	(7.2%)
Singular Only Nouns	577	293	870
e.g. *milk*	(20.97%)	(11.41%)	(16.36%)
Plural Only Nouns	--	10	10
e.g. *trousers*		(0.35%)	(0.19%)
Measure Nouns	38	30	68
e.g. *pile*	(1.38%)	(1.17%)	(1.28%)
Dual Function (measure and collective) Nouns	12	12	24
	(0.44%)	(0.47%)	(0.45%)
Collective Nouns	31	37	68
e.g. *family*	(1.34%)	(1.44%)	(1.38%)
Cluster/Unit Nouns	69	--	69
e.g. *moch/mochyn* 'pigs/pig'	(2.51%)		(1.3%)
TOTAL	2751	2567	5318
	(100%)	(100%)	(100%)

Since the noun **type** data are nominal, a chi-square analysis was carried out, that indicated a significant difference (χ^2 (20, n=5354) = 894.64, p<.000) in type distribution across the two languages. Similarities and differences are observed

across Welsh and English noun type patterns. As can be seen in Table 3.1, the main points of divergence between the two languages are: (1) no cluster/unit types were found in English; (2) no plural only types were observed in Welsh; (3) fewer singular only types were found in English than in Welsh; and, (4) fewer dual function (singular/plural - singular only) types were found in Welsh than in English. Table 3.1 also highlights the similarities between Welsh and English noun type patterns; similar proportions of other noun types were observed across the two languages. Specifically, singular/plural nouns, measure nouns, collective nouns and dual function (collective-measure) nouns were comparable across Welsh and English.

Token frequencies of noun occurrence as singular (sg), unit (u), plural (pl) or cluster (c) forms for Welsh and English are shown in Table 3.2.

Mean token frequencies were calculated by dividing the token frequency counts by the type frequency counts. Mean token frequencies of singular, plural, cluster and unit forms for all noun types in Welsh and English are shown in Table 3.3.

For the purpose of statistical analysis, it was necessary to collapse the token frequency data of the different forms by referent for comparison. Consequently, forms denoting a single referent (i.e. singular and unit forms) were distinguished from forms denoting multiple referents (i.e. plural and cluster forms). This was because English does not have cluster/unit forms equivalent to Welsh forms for comparison. Noun types were defined as those in Table 3.3.

A language x type x referent form (single versus multiple) ANOVA was carried out on the token frequency data. Significant main effects of type, $F(8,5290) = 7.01, p<.000$, and referent form, $F(1,5290) = 72.80, p<.000$, were found. Significant interactions for type x referent form, $F(8,5290) = 2.87, p<.003$, and language x referent form, $F(1,5290) = 14.56, p<.000$, were also found.

Table 3.2 Token frequencies for the noun types in Welsh and English

Noun Types	Welsh Token Frequency	English Token Frequency	Overall Token Frequency
Singular(sg)/Plural(pl) Nouns	Sg = 172725	Sg = 92262	Sg =112887
e.g. *chairs/s*	Pl = 2693	Pl = 36356	Pl = 39049
Dual Function (singular/plural	Sg = 4146	Sg = 17968	Sg = 22114
and singular only) Nouns	Pl = 104	Pl = 3817	Pl = 3921
e.g. *cake/s*			
Singular Only Nouns	Sg = 31497	Sg = 8759	Sg = 40256
e.g. *milk*			
Plural Only Nouns	--	Pl = 392	Pl = 392
e.g. *trousers*			
Measure Nouns	Sg = 4171	Sg = 2615	Sg = 6786
e.g. *pile*	Pl = 54	Pl = 790	Pl = 844
Dual Function (measure and	Sg = 909	Sg = 1040	Sg = 1949
collective) Nouns	Pl = 9	Pl = 334	Pl = 343
e.g. *bunch*			
Collective Nouns	Sg = 4229	Sg = 3009	Sg = 7238
e.g. *family*	Pl = 1	Pl = 1059	Pl = 1060
Cluster/Unit Nouns	Unit = 5067	--	Unit = 5067
e.g. *moch/mochyn* 'pigs/pig'	Cluster = 4095		Cluster = 4095
Total	Sg = 213952	Sg = 125653	Sg = 339605
	Unit = 5067	Pl = 42748	Unit = 5067
	Pl = 2861		Pl = 45609
	Cluster = 4095		Cluster =4095

Post hoc analysis (LSD) of type shows that token frequency of the cluster/unit type was significantly higher than the singular/plural type (MD = 37.55, $p<.000$), the singular only type (MD = 53.96, $p<.000$), the dual function (singular/plural - singular only) type (MD = 39.27, $p<.000$), and the plural only type (MD = 57.31, $p<.026$). The token frequency of the collective type was also significantly higher than the singular/plural type (MD = 20.57, $p<.027$), the singular

only type (MD = 36.98, $p<.000$), and the dual function (singular/plural - singular only) type (MD = 22.28, $p<.037$). The singular only type had a token frequency significantly lower than the singular/plural type (MD = -16.41, $p<.000$), the measure type (MD = -32.93, $p<.001$), and the dual function (singular/plural - singular only) type (MD = -14.69, $p<.015$). These results show that token frequencies differ across noun types for both languages.

Table 3.3 Mean token frequencies of forms for each noun type in Welsh and English

Noun Types	**Welsh Mean Token Frequency**	**English Mean Token Frequency**	**Total Mean Token Frequency**
Singular(sg)/Plural(pl) Nouns	Sg = 87.85	Sg = 49.67	Sg =137.52
e.g. *chairs/s*	Pl = 1.36	Pl = 19.67	Pl = 21.03
Dual Function (singular/plural	Sg = 85.17	Sg = 55.46	Sg = 140.63
and singular only) Nouns	Pl = 1.80	Pl = 11.82	Pl = 13.62
e.g. *cake/s*			
Singular Only Nouns	Sg = 65.17	Sg = 29.89	Sg = 95.06
e.g. *milk*			
Plural Only Nouns	--	Pl = 39.20	Pl = 39.20
e.g. *trousers*			
Measure Nouns	Sg = 112.39	Sg = 87.17	Sg = 199.56
e.g. *pile*	Pl = 1.42	Pl = 26.33	Pl = 27.75
Dual Function (measure and	Sg = 75.75	Sg = 86.67	Sg = 162.42
collective) Nouns	Pl = 0.75	Pl = 27.83	Pl = 28.58
e.g. *bunch*			
Collective Nouns	Sg = 136.42	Sg = 79.18	Sg = 215.60
e.g. *family*	Pl = 0.03	Pl = 27.87	Pl = 27.90
Cluster/Unit Nouns	Unit = 96.38	--	Unit = 96.38
e.g. *moch/mochyn* 'pigs/pig'	Cluster = 59.35		Cluster = 59.35
Overall mean frequency	Sg = 474.83	Sg = 388.04	Sg = 862.87
	Unit = 96.38		Unit = 96.38
	Pl = 5.36	Pl = 152.72	Pl = 158.08
	Cluster = 59.35		Cluster = 59.35

Token frequencies also differ across referent form in the two languages. That is, forms referring to individuals occur significantly more often than forms referring to multiple entities in both languages.

The referent form x type interaction shows that token frequencies for each form (i.e. forms denoting a single referent versus multiple referents) differ across the different noun types.

Post hoc analysis (LSD) indicates that the form referring to single items has a significantly higher token frequency than the form referring to multiple items for each noun type apart from the plural only type where the reverse is true. The unit form has a higher token frequency than cluster forms (MD = 37.03, p.021) for the c/u type. Singular forms have a higher token frequency than the plural forms for the collective type (MD = 89.54, p<.000), the dual collective and measure type (MD = 66.92, p<.000), the measure type (MD = 88.85, p<.000), the singular only type (MD = 53.65, p<.000), the singular/plural type (MD = 60.45, p<.000) and the dual singular/plural singular only type (MD = 49.81, p<.000). The plural form token frequency is significantly higher than the singular form for the plural only type (MD = 37.30, p<.000).

The language x referent form interaction demonstrates that the token frequency of the referent forms differ across the two languages. Welsh has higher token frequency for forms referring to individuals (MD = 37.313, p<.000) than English, yet lower token frequencies for forms referring to multiples (MD = 14.209, p<.000). This is illustrated in Figure 3.1.

Not only were the mean token frequencies distinguishable in the two languages, but also the relative proportions of the nouns referring to individuals versus multiples. In Welsh, 96.97% of the all forms referred to single entities, with only 3.03% referring to multiples. In English, 74.62% of forms referred to individuals and 25.38% to multiples.

A large proportion (58.87%) of all the forms referring to multiple entities in Welsh was made up of the cluster forms; only 41.13% were morphologically marked plurals. Thus, cluster nouns make up the largest portion of the forms in Welsh that have multiple referents.

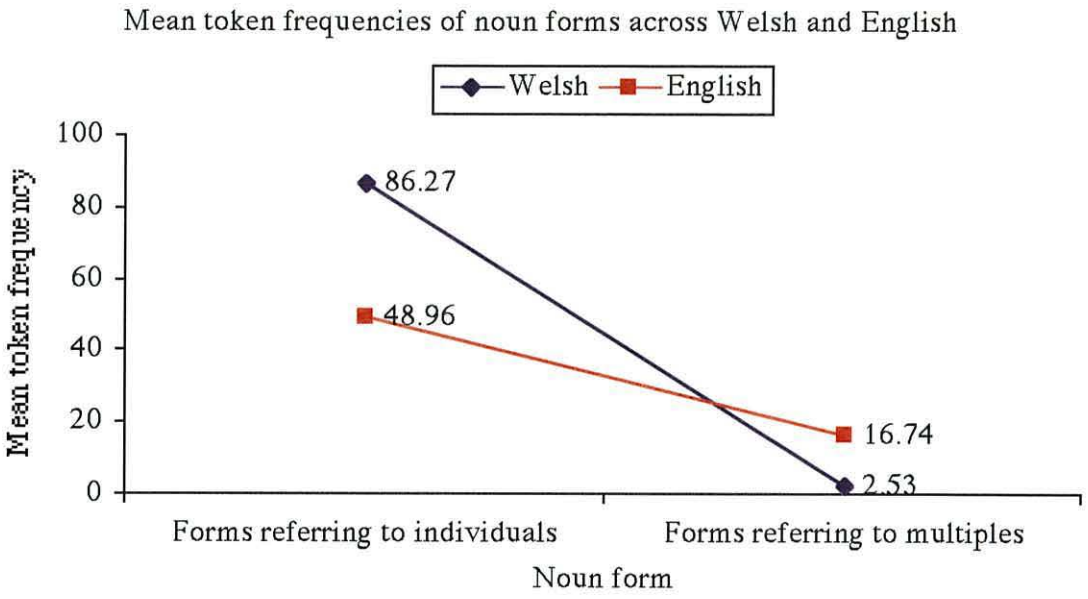


Figure 3.1 Mean token frequencies of forms denoting individuals and forms denoting multiples in Welsh and English.

Further analysis of the frequency counts for the cluster/unit types in Welsh indicates that the cluster forms (denoting multiples) occur almost as frequently (44.7%) as the unit forms (denoting individuals) (55.3%). Compare this to the singular and plural forms of the other noun types in both languages. As mentioned above, only a quarter of English nouns referred to multiple entities and even less in Welsh. For all other noun types in Welsh, 98.7% occur as singular forms and only 1.3% occurs as plural forms. This indicates that the cluster forms are the most frequently used forms to refer to multiple entities in Welsh.

Summary

The results identified both similarities and differences in noun distributions in Welsh and English. For type frequencies, Welsh and English had similar numbers of singular/plural nouns; measure nouns; and, collective nouns. Within both languages, certain noun types are more frequent (e.g. singular/plural nouns) than others (e.g. collective nouns). Forms that refer to individuals occur more often than forms that refer to multiples within and across both languages, with exception of the plural only nouns in English.

The differences between the two languages are highlighted by the fact that Welsh has cluster/unit nouns; no plural only nouns; more singular only nouns; and fewer dual function (singular/plural and singular only) type nouns. Forms that refer to individuals occur more often in Welsh than in English whilst the forms that refer to multiples occur more often in English than in Welsh. Most of the forms that refer to multiples in Welsh are cluster nouns, and nouns of the cluster/unit type occur as cluster forms more often than as unit forms.

Discussion

The results show that although there are similarities between Welsh and English, the two languages differ in crucial ways with regard to the way each language encodes *number* on nouns.

Nouns of the singular/plural type made up the largest portion of noun types in both languages. This suggests that most noun referents are treated as countable entities by the two languages. Semantically, the singular/plural system provides a clear distinction between individuals and multiples. The fact that nouns referring to individuals occur more often than nouns that refer to multiples in both languages suggests that there is greater emphasis on individuals. This is especially noticeable in Welsh, with very few plural forms being used.

One possible explanation for the strong disparity between forms referring to individuals and forms referring to multiples in Welsh is the way nouns are quantified. As detailed in chapter 2, numerals in Welsh occur directly with singular or unit forms (e.g. *wyth llyfr* 'eight book'; *chwech mochyn* 'six pig') or with plural or cluster forms if the partitive *o* 'of' is used (e.g. *wyth o lyfrau* ' eight of books', *chwech o foch* 'six of pigs'). This means that despite using a singular or unit form, the *number* reference is not necessarily clear. The singular and unit forms in Welsh are in some sense non-specific to *number* reference since they can be used with numerals to refer to several entities. Welsh has several different ways to express *number* when quantifying nouns. The frequency of the forms used (i.e. singular, unit, plural and cluster) may reflect the way different noun types are quantified.

Moreover, singular only nouns in both languages generally refer to substances or at least allow the referents to be thought of as substance-like. Given that Welsh has more singular only nouns than English does, it appears that Welsh, compared to English, provides a stronger emphasis on substance-likeness.

The relative scarceness of collective nouns in the two languages suggests that neither language emphasise collection categories. In Welsh, though, 2.5% of nouns were of the cluster/unit type. This then may increase the emphasis in Welsh to think about entities as collections. The c/u system in Welsh is in opposition to the singular/plural system. Not only is the semantic nature of cluster nouns (substance-like meaning) distinct from plural forms (meaning many) but also these nouns occur as clusters almost as often as they occur as unit forms. In no other noun type category, in Welsh or English, do the forms referring to multiples occur nearly as often as the forms referring to single entities[1].

Cluster nouns then distinguish Welsh from English in very important ways. First, cluster/unit nouns provide an additional way for Welsh speakers to conceptualise entities by the way they encode *number*. That is, the c/u system in Welsh allows for a greater scope for Welsh speakers to conceptualise entities in the world as collections. Second, cluster forms are the main source of describing multiple entities in Welsh.

These patterns reflect differences in the way Welsh and English encode reference to entities in the world. Given the differences in the way Welsh, compared to English, encodes *number*, it is not always clear from a noun whether it refers to a single entity, several entities or a collection. As discussed in chapter 2, an uninflected form could refer to either a single entity (singular nouns), multiple entities (plural nouns with internal vowel changes), or a collection (cluster nouns). For example, compare *cwch* (singular noun) 'boat', *ceir* (plural noun) 'cars', and *moch* (cluster noun) 'pigs'. There is no clear indication from these nouns about whether they refer to individuals, multiples or a collection. How do children learn to make these distinctions?

---

[1] With exception of the plural only nouns in English.

To explore the role of language differences between Welsh and English on language acquisition and cognitive processes it is necessary to first establish what children are hearing in the input language. Research has suggested that child language acquisition correlates highly with the language used by the mother (Rowland & Pine, 2000; Theakston, Lieven, Pine & Rowland, 2002; Naigles & Hoff-Ginsberg, 1998; Winjen, Kempen & Gillis, 2001; Crago, Allen & Pesco, 1998). Are the differences noted between Welsh and English written texts maintained in mothers' speech to their children? Do Welsh-speaking mothers use the different noun types in Welsh in their speech to their children to facilitate language learning?

The following chapter describes a semi-naturalistic study that examines mothers' speech to infants. The aim is to explore any differences in the input language across Welsh and English with regard to the noun types identified in the above study.

Study 2: Input Characteristics of Noun Types in Mothers' Speech to infants

The typological study discussed in the previous chapter examined the distribution of noun types and their token frequencies in both Welsh and English text material. The aim was to identify, and establish an empirical measure of, the differences and similarities between the two languages. Although the study provides a quantitative account of the distributional differences between the two languages with regard to written texts, it does not indicate the usage of different noun types in child directed speech. One of the main questions being addressed by this thesis is the ways in which children acquire collective nouns. To address this question, it is useful not only to understand the differences between the languages in general, but also to examine the language patterns that the child hears in the input.

It is widely recognised that mothers modify their speech to infants. Research into child directed speech has indicated that language acquisition correlates highly with, and is consistent with, the language the child hears. Children's early words and grammatical behaviours reflect those in their mother's speech to them (Bloom and Wynn, 1997; Huttenlocher et al, 2002; Rowland & Pine, 2000; Mintz, Newport & Bever, 2002; Theakston, Lieven, Pine & Rowland, 2001; Naigles & Hoff-Ginsberg, 1998; Winjen, Kempen & Gillis, 2001; Crago, Allen & Pesco, 1998; Rohde & Plaut, 1999). Moreover, the context of play (e.g. book reading versus toy play) also seems to influence the mother's and child's language use (Yont, Snow & Vernon-Feagans, 2002). To determine the effects of differences in language structure on language acquisition it is necessary to explore the channels through which this information is passed on. That is, mothers' use of particular grammatical categories, in this case collection categories, may have important implications for the way their

children acquire them.

In this chapter, the aim is to identify and establish the differences and similarities in the use of noun types by Welsh- and English-speaking mothers in their speech to infants. Mothers' speech to children will be examined in two different play conditions. This is to explore whether mothers' use of nouns referring to collections can be elicited with the use of specific toys. This study adopts a coding strategy formulated in study 1 to identify particular noun types in mothers' speech and is mainly concerned with the collection categories identified in chapter 3.

Method

Design

This study was designed to examine Welsh- and English-speaking children's exposure to different noun types and noun forms in the mothers' input[1]. To determine whether specific play conditions would elicit particular noun usage by mothers from each language group, the study involved two sessions, each with a different play condition. Each session was one-hour long. Two sessions were video recorded for all 10 mother-child dyads. Each session was either a 'freeplay' session or a 'structured' session. The 'freeplay' session involved the mother and child playing with the child's own toys. The 'structured' session involved the mother and child playing with specifically chosen toys to examine whether such toys would elicit particular noun usage by the mother. The order in which participants undertook the play conditions was counterbalanced across participants for each language group. For each mother-child pair, the two sessions were recorded back to back with the exception of one Welsh mother-child dyad where the 'structured' session was recorded a week after the 'freeplay' session.

Participants

Five Welsh-speaking and five English-speaking infants and their mothers took part in this study. The mean age of the Welsh-speaking infants was 1;9.3 (range: 1;8.2 – 1;11.0); the mean age of the English-speaking infants was 1;7.3 (range: 1;5.0 – 1;10.0). The data for the English-speaking children were collected by a final-year project student and are reported in Sproson (unpublished final year project, 2001). The author collected the data for the Welsh-speaking children.

Participants were recruited from the local community subject panel. Recruitment of Welsh-speaking infants and mothers were based on the parental reports of the child's, and their own, language background via a questionnaire [Appendix III]. The criteria for inclusion into the Welsh group were based on parental reports indicating that the mother's own and child's first language was Welsh and that the primary language spoken at home was Welsh (80-100% Welsh at home). Each mother and child pair participated in both the 'freeplay' and 'structured' sessions.

Apparatus

A video recorder was used to record mother-child interactions during each session. A specific selection of toys was used for the 'structured' session only.

Stimuli

The toys selected for the 'structured' session were objects that fell into one of three noun type categories identified in study 1. These are the singular/plural nouns, cluster/unit nouns and collective nouns. A set of toys was chosen to represent examples of each noun type and are detailed in Table 4.1.

Procedure

All sessions involving the Welsh mother-child dyads were recorded at the Child Language Laboratory. Two of five English mother-child dyads had both sessions recorded at their home. For the 'freeplay' sessions at the laboratory, mothers were asked to bring along the child's own toys to play with because the general purpose toys available at the laboratory were limited. For the 'structured' session, mothers were asked to play with their child using only the toys specified in Table 4.1.

---

[1]The design of this study is based on the work of Sproson (2001, unpublished final year project).

Initially, mothers were not told the precise nature of the study to avoid mothers interacting with their children in an unnatural or contrived manner. Instead, mothers were told that the aim of the experiment was to observe the way children played and used language in two different play conditions. Mothers were asked to play with their child in a natural way during both sessions. The video recorder was switched on at the start of each session. The experimenter remained in the room recording but did not engage in the play unless spoken to directly by the mother or child. This was to encourage naturalistic mother-child interactions. Two one-hour play sessions, one 'freeplay' and one 'structured', were recorded for each mother-child dyad. Participants were given a break between sessions when the sessions were undertaken successively. Following the final session, mothers were given details of the precise nature of the study and debriefed.

Table 4.1 Toys used for each noun type category in the experimental condition.

Noun Type	**Toy**	**Welsh Nouns**	**English Nouns**
Singular/plural nouns in Welsh and English	5 balls (different colours and sizes)	Pêl	Ball
	3 teddies (different shapes, colours and sizes)	Tedi	Teddy
	5 cars (different shapes, colours, sizes)	Car	Car
Cluster/unit nouns in Welsh and singular/plural nouns in English	3 star shaped cushions	Seren	Star
	5 plastic pigs	Mochyn	Pig
	5 plastic fish	Pysgodyn	Fish
Collective nouns in Welsh and English	4 dolls (male, female, small female and baby)	Teulu	Family
	7 plastic cows	Gwartheg	Cattle
	9 plastic trees	Coedwig	Forest

Results

Each video was transcribed. Only the nouns in the transcripts were analysed. The nouns were classified into the noun **type** classes identified in study 1 using the same classification criteria. The frequency of each form (i.e. singular, plural, cluster or unit) was also calculated for each noun.

Noun type frequencies were calculated as in study 1 to determine the number of nouns that were of a given type (e.g. collective or singular only). Type frequencies for Welsh and English are shown in Table 4.2.

The table clearly shows the striking similarity between the patterns of noun type distribution in mothers' input to children to those found on written texts across both languages. Again, since noun type data was nominal a chi-square analysis was carried out. A significant difference in noun type distributions across the two languages was found (χ^2 (6, n=1364) = 78.15, p<.000).

The differences and similarities across Welsh and English noun types in these data are consistent with those found in written texts (study 1). The data show that the main points of divergence between the two languages are: (1) no cluster/unit forms were found in English; (2) no plural only forms were observed in Welsh; (3) fewer singular only nouns were found in English than in Welsh; and, (4) fewer dual function (singular/plural - singular only) nouns were found in Welsh than in English. The data also show that the proportions of specific noun types were comparable across the two languages. Specifically, singular/plural nouns were the most frequent noun types across the two languages, measure nouns were relatively scarce across the two languages, and collective nouns were a very small proportion of the noun types distribution across the two languages.

Table 4.2 Number of noun types for Welsh and English occurring for each condition

Noun type	Structured session		Freeplay session		Total	
	Welsh Type Frequency	English Type Frequency	Welsh Type Frequency	English Type Frequency	Welsh	English
Singular (sg)/plural (pl) e.g. *chair/s*	113 (73.86%)	127 (89.44%)	169 (80.09%)	158 (89.77%)	282 (77.47%)	285 (89.62%)
Dual function (singular/plural and singular only) e.g. *cake/s*	2 (1.31%)	4 (2.82%)	3 (1.42%)	8 (4.55%)	5 (1.37%)	12 (3.79%)
Singular only e.g. *milk*	26 (16.99%)	8 (5.63%)	28 (13.27%)	8 (4.55%)	54 (14.84%)	16 (5.03%)
Plural only e.g. *trousers*	--	1 (0.70%)	--	--	--	1 (0.31%)
Measure nouns e.g. *pile*	1 (0.65%)	1 (0.70%)	3 (1.42%)	2 (1.14%)	4 (1.10%)	3 (0.94%)
Collective nouns e.g. *forest*	--	1 (0.70%)	2 (0.95%)	--	2 (0.55%)	1 (0.44%)
Cluster/unit e.g. *moch/mochyn* 'pigs/pig	11 (7.19%)	--	6 (2.84%)	--	17 (4.67%)	--
Total	153 (100%)	142 (100%)	211 (100%)	176 (100%)	364 (100%)	318 (100%)

A chi-square of noun type x play condition for each language revealed a significant differences between the two conditions for Welsh (χ^2 (5, n=1364) = 13.59, p<.018) but not for English (χ^2 (5, n=1364) = 6.89, p<.229). Welsh mothers used cluster/unit nouns almost twice as often in the 'structured' session as in the 'freeplay'

session, as can be seen in Table 4.2. This suggests that toys specifically chosen to provide a context for noun use do elicit cluster/unit noun use by Welsh mothers.

The **token** frequency counts were the total number of nouns in each noun type class. Token frequencies of singular, unit, plural and cluster forms for each of the noun types for each language and each play condition are shown in Table 4.3.

Table 4.3 Token frequencies of singular/unit and plural/cluster forms for the noun types in Welsh and English for each play condition

	Structured session		**Freeplay session**		**Total**	
Noun type	Welsh Type Frequency	English Type Frequency	Welsh Type Frequency	English Type Frequency	Welsh Type Frequency	English Type Frequency
Singular (sg)/plural (pl) e.g.*chair/s*	Sg = 783 Pl = 42	Sg = 503 Pl = 160	Sg = 731 Pl = 45	Sg = 627 Pl = 158	Sg = 1514 pl = 87	Sg = 1130 pl = 318
Dual function (singular/plural and singular only) e.g. *cake/s*	Sg = 23 Pl = 0	Sg = 7 Pl = 6	Sg = 34 Pl = 0	Sg = 25 Pl = 3	Sg = 57 Pl = 0	Sg = 32 Pl = 9
Singular only e.g. *milk*	Sg = 71	Sg = 23	Sg = 84	Sg = 25	Sg = 155	Sg = 48
Plural only e.g. *trousers*	--	Pl = 1	--	--	--	Pl = 1
Measure nouns e.g. *pile*	Sg = 3 Pl = 0	Sg = 0 Pl = 3	Sg = 7 Pl = 0	Sg = 4 Pl = 6	Sg = 10 Pl = 0	Sg = 4 Pl = 9
Collective nouns e.g. *forest*	--	Sg = 1 Pl = 0	Sg = 2 Pl = 2	--	Sg = 2 Pl = 2	Sg = 1 Pl = 0
Cluster/unit e.g. *moch/mochyn* 'pigs/pig	U = 83 C = 13	--	U = 12 C = 4	--	U = 95 C = 17	--
Total type frequency	Sg = 880 U = 83 Pl = 42 C = 13	Sg = 534 Pl = 170	Sg = 858 U = 12 Pl = 47 C = 4	Sg = 681 Pl = 167	Sg = 1738 U = 95 Pl = 89 C = 17	Sg = 1215 Pl = 337

Mean token frequencies were calculated by dividing the token frequency counts by the type frequency counts giving an indication of the mean number of

occurrences of a given noun type in the mothers' speech. Again, since English does not have equivalent cluster/unit forms, it was necessary to collapse the token frequency data of the forms by referent for comparison. Consequently, forms denoting a single referent (i.e. singular/unit forms) were distinguished from multiple referents (i.e. plural/cluster forms) for the purpose of statistical analysis. Types were defined as those in Table 4.3. Mean token frequencies and standard deviations of forms referring to individuals and forms referring to multiples for Welsh and English are shown in Table 4.4 for the 'freeplay' session and Table 4.5 for the 'structured' session.

A language x noun type x referent form ANOVA revealed a significant main effect of referent form, $F(1,1360) = 149.34, p<.018$, and a significant interaction for language x referent form, $F(1,1360) = 10.62, p<.042$). No significant effects or interactions were found for language x noun type or language x play condition for the token frequency data.

The main effect of form suggests that, overall, nouns referring to individuals (i.e. singular/unit nouns) (M = 4.44, SD = 7.81) occur significantly more often than nouns that refer to multiple entities (i.e. plural/cluster) (M = 0.68, SD = 1.89).

The interaction demonstrates that the token frequencies of the noun forms differ across the two languages. Post hoc analysis (LSD) shows that Welsh had a significantly higher frequency of forms referring to individuals (MD = 1.231, $p<.040$), and significantly lower frequency of forms referring to multiples (MD = -.772, $p<.000$) than English. This is illustrated in Figure 4.1. These data are consistent with the noun distribution patterns in written texts (study 1).

Table 4.4 Means (standard deviations) of token frequencies of forms referring to single items and multiple items for each noun type in Welsh and English for the freeplay session

Noun type	Forms referring to single items		Forms referring to multiple items			
	Welsh Mean Token Frequency	English Mean Token Frequency	Welsh Mean Token Frequency	English Mean Token Frequency	Total Mean Token Frequency	
	M(SD)	M(SD)	M(SD)	M(SD)	Welsh	English
Singular (sg)/plural (pl) e.g.*chair/s*	4.36 (7.50)	3.97 (5.49)	0.26 (1.13)	1.01 (2.08)	4.62 (8.63)	4.98 (7.57)
Dual function (singular/plural and singular only) e.g. *cake*	11.33 (8.63)	3.13 (2.03)	--	0.38 (1.06)	11.33 (8.63)	3.51 (3.09)
Singular only e.g. *milk*	3.00 (3.98)	3.13 (2.85)	0.04 (0.19)	--	3.04 (4.17)	3.13 (2.85)
Plural only e.g. *trousers*	--	--	--	--	--	--
Measure nouns e.g. *pile*	2.33 (2.31)	2.00 (2.83)	--	3.00 (2.83)	2.33 (2.31)	5.00 (8.49)
Collective nouns e.g. *forest*	1.00 (0.00)	--	1.00 (1.41)	--	2.00 (1.41)	--
Cluster/unit e.g. *moch/mochyn* 'pigs/pig	2.00 (2.10)	--	0.67 (1.63)	--	2.67 (3.73)	--
Total mean token frequencies	24.02 (25.09)	12.23 (13.20)	1.97 (5.51)	4.39 (5.97)	25.99 (30.60)	16.62 (19.17)

Table 4.5 Means (standard deviations) of token frequencies of forms referring to single items and multiple items for each noun type in Welsh and English for the structured session

Noun type	Forms referring to single items		Forms referring to multiple items			
	Welsh Mean Token Frequency	English Mean Token Frequency	Welsh Mean Token Frequency	English Mean Token Frequency	Total Mean Token Frequency	
	M(SD)	M(SD)	M(SD)	M(SD)	Welsh	English
Singular(sg)/plural (pl) e.g.*chair/s*	6.93 (13.2)	3.96 (5.96)	1.69 (0.56)	1.26 (3.05)	8.62 (13.76)	5.22 (9.01)
Dual function (singular/plural and singular only) e.g. *cake*	11.50 (4.95)	1.75 (1.71)	--	1.50 (1.91)	11.50 (4.95)	3.25 (3.62)
Singular only e.g. *milk*	2.88 (3.12)	2.88 (2.47)	0.04 (0.20)	--	2.92 (3.32)	2.88 (2.47)
Plural only e.g. *trousers*	--	--	--	1.00 (0.00)	--	1.00 (0.00)
Measure nouns e.g. *pile*	3.00 (0.00)	--	--	3.00 (0.00)	3.00 (0.00)	3.00 (0.00)
Collective nouns e.g. *forest*	--	1.00 (0.00)	--	--	--	1.00 (0.00)
Cluster/unit e.g. *moch/mochyn* 'pigs/pig	7.55 (9.35)	--	1.18 (1.33)	--	8.73 (10.68)	--
Total mean token frequencies	31.86 (30.59)	9.59 (10.14)	2.91 (2.09)	6.76 (4.96)	34.69 (32.68)	16.35 (15.10)

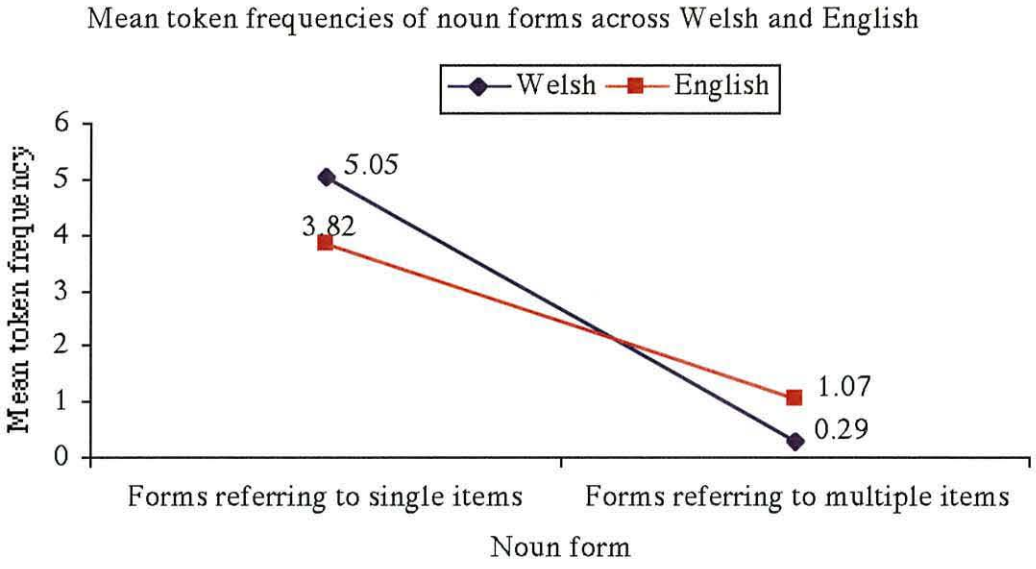


Figure 4.1 Mean token frequencies for each noun form in each language

Since there was no effect of play condition on **token** frequency distribution the data were collapsed across the two conditions for the remaining analysis.

To further explore the interaction between the nouns that refer to individuals (singular/unit) and the nouns that refer to multiple individuals (plural/cluster) in Welsh and English, the proportions of singular/unit and plural/cluster forms were observed. Of the nouns used by the Welsh mothers 94.53% referred to individuals compared to only 5.47% that referred to multiple referents. Of the nouns used by English mothers 78.29% referred to individuals with 21.71% referring to multiple referents. These data are consistent with those found for written texts. Again, forms that refer to individuals occur much more often in Welsh than in English.

Of those few nouns used by Welsh mothers to refer to multiple referents only 16.04% were cluster forms whilst 83.96% were plural forms. The number of cluster nouns used by Welsh mothers was much lower than was expected given the data from study 1. Recall that in written texts, 58.87% of the nouns referring to

multiple referents were cluster forms yet in mothers' speech only 16.04% were cluster forms.

In addition, by examining the cluster/unit forms in Welsh it is apparent that, unlike the findings of study 1, these nouns occur as unit forms far more often than as cluster forms. In the mothers' speech, 84.82% of cluster/unit nouns were unit forms and only 15.18% were cluster forms. Compare these to those in written texts (study 1) where 44.70% were cluster forms.

To explore the high ratio of singular/unit forms to plural/cluster forms in Welsh further, the context in which the singular/unit forms used by the mother were examined. Primarily, this was to determine the frequency at which the mothers used numerals with the singular or unit forms.

There were 8 (1.02%) instances of numerals co-occurring with singular forms and 3 (3.61%) instances of numerals co-occurring with unit forms in the 'structured' session. There were also 3 (0.41%) instances of numerals co-occurring with singular forms in the 'freeplay' session but none with unit forms. This indicates that singular and unit forms do not always refer to individuals in Welsh. On occasion these forms were used when mothers were referring to several entities.

Summary

As with study 1, the results identified both similarities and differences in noun distributions in Welsh and English. The similarities across the two languages are highlighted by the similarity in frequency of the different types. In particular, both languages had similar numbers of singular/plural nouns; measure nouns; and, collective nouns as shown in study 1. Again, certain noun types (e.g. singular /plural nouns) were more frequent than others (e.g. collective nouns). Forms referring to individuals occur more often than forms that refer to multiples. The differences between the two languages are shown by the fact that Welsh has cluster/unit nouns; no plural only nouns; more singular only nouns; and fewer dual function (singular/plural and singular only) type nouns. Forms that refer to individuals occur more often in Welsh than in English whilst the forms that refer to multiples occur more often in English than in Welsh. In contrast to the findings in written texts, most of the forms that refer to multiples in Welsh input were plural forms not cluster forms, and of the cluster/unit type, unit forms were more frequent than the cluster forms. Moreover, a proportion of the singular and unit forms co-occurred with numerals in the Welsh mothers' speech. Overall, the data from mothers' language input are consistent with those found in written texts.

Discussion

The noun patterns used by mothers in their speech to children for the most part reflect the noun patterns of the language. The distribution of noun types in the input to children is similar to that found in written texts across the two languages. For example, most nouns used are of the singular/plural type. Welsh- and English-speaking mothers rarely used collective nouns in their speech to infants. Since collective nouns make up only 1.5% of the different noun types in both languages (see study 1), it is not surprising that mothers rarely use them. With few collective nouns in the input, children will not have sufficient exemplars from which to acquire collection categories.

Study 1 demonstrated that the additional system in Welsh, the cluster/unit (c/u) system, was an important difference between Welsh and English. Welsh children may have a greater possibility of exposure to nouns that refer to collections than English-speaking children. The data from this study show that Welsh-speaking mothers use nouns of the c/u type in a similar proportion to that expected from the data in Study 1 on written texts. And given particular contexts to use the c/u types such as the 'structured' play condition, Welsh-speaking mothers did use more of these noun types. Interestingly, however, few of the nouns of this type used were cluster forms. Essentially, mothers used the unit forms more often than the cluster forms despite these forms being morphologically more complex.

Recall the way Welsh quantifies entities as discussed in chapter 2. To quantify entities with a numeral in Welsh, singular or unit forms are used (e.g. *wyth llyfr* 'eight book', *chwech mochyn* 'six pig'). Numerals can only occur with plural or cluster forms with the use of the partitive *o* 'of' (e.g. wyth *o* lyfrau ' eight [of] books' *chwech o foch* 'six [of] pigs).

Indeed, mothers did use numerals with the singular and unit forms a large proportion of the time and very few plural forms were used at all. This means that for a child learning Welsh, it is possible that singular and unit forms are non-specific with regard to number. Given that so few plural forms were used by Welsh mothers, and that singular and unit forms occurred with numerals, then Welsh speaking children do not have a clear guide to the *number* reference of the noun. This is compounded by the fact that there are several ways to pluralise in Welsh. Compare this to English where numerals usually occur with plural forms. Here, there is a much clearer indication of *number* from the noun.

It is probable then that singular and unit forms were particularly frequent when compared to the plural/cluster forms in Welsh because of the way Welsh-speaking mothers quantified nouns when interacting with their child. This then complicates the process for children acquiring Welsh because it is not always clear from the noun to what the noun refers.

The two play conditions elicited different noun distributions from the Welsh mothers but not from the English mothers. Welsh mothers used more nouns of the c/u type in the 'structured' session than in the 'freeplay' session. The 'structured' session involved a specific set of toys chosen to provide a context for use of these noun types. No differences were found between the play condition for the token frequency data. Welsh mothers used more different types of the c/u nouns in the 'structured' session, but did not use them more frequently. Neither did the ratio of each form (i.e. singular, plural, unit and cluster) differ across the two play conditions. This suggests that context does play a role in providing Welsh mothers with the opportunity to use more c/u nouns but does not affect the frequency with which these nouns are used. The 'structured' play condition did not elicit greater collective noun

use, as was expected, for either the Welsh or English-speaking mothers. Mothers across the two languages and the two play conditions used very few collective nouns.

The main issue highlighted by these results is the complexity of the number marking system in Welsh. Given that (1) Welsh has the additional c/u system, (2) plural forms in Welsh were scarcely used, (3) there are several ways to pluralise in Welsh, and (4) singular and unit forms are used with numerals, then children are exposed to a complex system that does not allow them to make easy distinctions between references to an individual, several individuals or collections.

The complexity of the *number* marking system in Welsh and the lack of sufficient exemplars may have important cognitive implications for Welsh-speaking children. For example, Welsh-speaking children may categorise entities for word learning in a different way to English speakers. Or Welsh speakers may pay less attention to *number* because their language does make clear *number* distinctions.

Welsh differs from English in important ways with regard to the way each language refers to entities and many questions arise from these findings. To address some of these questions, the following chapter describes a large study that explores the effects of differences in language structure on categorisation and recognition. The aim was to determine whether Welsh speakers more readily categorise novel objects as collection than English speakers do. And, if so, does this effect recognition of the individuals that make up the collections? The study attempts to explore the role of object properties (such as number of items in a group) and syntactic cues in the categorisation of novel stimuli. Categorisation and recognition patterns across the two languages were analysed as indicators of participants' interpretations of the novel nouns.

Study 3: Categorisation and recognition of novel stimuli

Although children learning both Welsh and English rarely heard collective nouns in their mothers' speech to them (Study 2), Welsh-speaking children are exposed to the additional cluster/unit (c/u) system that allows basic noun forms to refer to collections. Children acquiring Welsh may take advantage of the options that their language gives them when interpreting new words. These options may allow Welsh speakers to think differently about the meanings of new nouns.

This study was designed to investigate three main questions: (1) Does language structure guide the way people categorise entities? And if so, does this impact on word learning? (2) What factors are important for representing collections as individuals? (3) If collections are represented as 'individuals', does this affect memory for the individuals that make up the collection?

Does language structure affect the way people categorise entities and learn new words?

Existing literature (Bloom, 1994; 1996; Bloom, Kelemen, Fountain and Courtney, 1995; Huntley-Fenner, 1995) indicates that collective nouns are more difficult to acquire because they involve a higher level of abstraction (i.e. generally collective nouns refer to superordinate level categories e.g. *crowd*). Since the literature focuses primarily on English-speaking children and adults, the findings may reflect the influence of the structure of English rather than the degree of difficulty in acquiring collective nouns. The structure of Welsh differs from English in that more basic level nouns referring to collections are used in Welsh than in English.

Given that the structure of Welsh provides the learner with a complex number marking system that differs from English, does this influence, and if so to what extent, the way children categorise novel objects and learn the meanings of

new words? Unlike English, Welsh has a c/u system that highlights groupness. In contrast to collective nouns, cluster nouns in Welsh generally refer to basic level categories. Such a system may provide Welsh speakers with a greater willingness to conceptualise groups of objects as individuals and to interpret novel words as referring to collections than their English-speaking counterparts. Thus, are child- and adult-speakers of Welsh more likely to interpret a new word as referring to collections than child- and adult-speakers of English as a result of the competing singular/plural (sg/pl) and c/u systems of their language? It is expected that Welsh-speaking children will find collective nouns easier to learn than English-speaking children do, and that adult Welsh speakers will be more inclined than English-speaking adults to represent arrays of stimuli as collections.

What factors are important for representing collections as individuals?

There are many factors or cues that may influence categorisation of objects as collections. Bloom (1994; 1996; Bloom and Verses, 1999) identified age, perceptual salience of groups, intentionality, and the syntax of nouns as key factors that influence interpreting novel words as collective nouns.

Factors such as the number of items in a group or the number of groups in a scene may also influence categorisation of collections as individuals. These factors may influence Welsh and English speakers in different ways. Since c/u nouns in Welsh refer to entities that exist in large groups (King, 1993), Welsh speakers may be more likely to interpret novel nouns as referring to collections especially when the groups are made up of large numbers.

Similarly, if syntax plays a vital role in interpreting novel nouns as referring to collections, as suggested by Bloom (1994; 1996), syntax may influence Welsh and English speakers in different ways.

As pointed out by Bloom (1994; 1996), singular syntax may facilitate a collective interpretation of novel nouns in the presence of groups of stimuli. For this reason, it was necessary to explore the influence of singular syntax across the

two languages. Singular forms are unmarked forms in both Welsh and English but in Welsh there is no indefinite article that would give a clear indication of singular reference. And singular forms can occur with numerals, rendering unmarked forms non-specific with regard to number reference. Unmarked forms then are especially ambiguous in Welsh. It was hypothesised that unmarked forms may be interpreted as referring to collections more often by Welsh speakers than by English speakers.

Moreover, in English, nouns can be preceded by the indefinite article *a/an* to highlight singularity. This is not the case in Welsh; therefore only unit forms in Welsh would clearly highlight singularity. This would also ensure that Welsh speakers would be tapping into both the c/u system as well as the sg/pl system. It was hypothesised that unit forms in Welsh, and singular forms with an indefinite article in English, would guide both Welsh and English speakers to interpret the meanings of these nouns as referring to individuals.

A plural form, in Welsh and English, refers to several individuals or several collections. Since there are few collective nouns (i.e. nouns that refer to collections and can be pluralised) in the either language, it was expected that novel plural forms would be interpreted as referring to several individuals by speakers of both languages.

This study then aimed to explore whether object properties (e.g. number of groups, number of items per group) or syntactic cues (e.g. singular, unmarked and plural forms) facilitate representations of groups and whether these factors have similar effects for Welsh and English speakers.

Do representations of groups affect memory for the individuals?

Markman and Siebert (1976) point out that collections are organized into part/whole relations (e.g. trees are parts of forest). Thus, individuals that make up a collection are conceived of as parts of the whole object. Studies on memory for object parts have revealed that participants have a poorer memory for

object parts than for whole objects. Akrum and Palmer (1991) used the same-different paradigm to study memory for object parts. They compared performances on whole object comparisons and part/whole comparisons. They found that whole object comparisons were more accurate and faster than the part/whole comparisons. These effects were found for presentations of single whole objects and for presentations of two whole objects. This supports the idea that hierarchical relations between wholes and parts are retained in memory representations of objects. This allows whole objects and object parts to be available in memory for comparison.

Although Akrum and Palmer (1991) found a whole object advantage for object recognition, this was not found for unconnected figures. They suggest that the parts are themselves represented as whole objects and not object parts, when depicted as unconnected figures. This leads to the idea that clusters of objects are represented as many whole objects rather than parts. Collections are not represented as groups of individual whole objects.

This study also aims to address this issue by exploring the possibility that categorising objects as collections leads to poorer memory for the individual items than if the groups of objects are categorised as discrete individuals. With this aim, Welsh- and English-speaking children and adults were presented with novel objects arranged into groups and named with novel nouns. Since the majority of cluster nouns in Welsh refer to large clusters of entities, the study controlled for the number of groups available and the number of items in each group. Similarly, the study controlled for syntax to determine the extent to which syntax plays a role in the two languages when conceptualising collections. Recognition of the novel object was then tested to explore whether context of first exposure to the stimuli would influence representation and ultimately memory for individuals and collections.

It was hypothesised that Welsh speakers will be more willing to represent collections of entities as individuals and to interpret novel nouns as

referring to collections than their English-speaking counterparts as a result of the structure of their language. It was also expected that factors such as number and syntax act as important cues that facilitate interpretation of novel nouns as collections. These cues were expected to be more robust for Welsh speakers in that higher number of items and unmarked contexts would increase collective interpretations. In essence, differences between the two language groups will be more pronounced when there are more items in the groups, when there are more groups available and when the syntax is unmarked. In addition, it was hypothesised that Welsh speakers would have a poorer memory for the individual objects that make up a group as a consequence of categorising and representing groups as whole individuals.

Method

General Design

This study involves three tasks run in a pre-specified order and administered in a single session. A collective categorisation task was undertaken first, followed by an arithmetic task and, finally, a recognition task. The collective categorisation task was designed to examine whether language differences affect Welsh- and English- speaking participants' reactions when faced with learning novel nouns for novel objects, given that the structure of Welsh differs from English in crucial ways with regard to collection categories. The arithmetic task was a basic arithmetic test and was designed to ensure that any differences between the two language groups on the collective categorisation task were not a consequence of differences in mathematical abilities. The recognition task was designed to examine whether participants' memory for the novel objects, introduced in the collective categorisation task, was affected by their initial interpretation of the novel nouns associated with those objects.

The tasks were administered in the same order across subjects (collective categorisation task, then arithmetic task, and then recognition task) for the following reasons. The recognition task required that participants had previously seen images in the collective categorisation task and thus, necessarily, occurred later. The nature of the arithmetic task could have confounded the responses for the collective categorisation task, if administered first, by priming the participants for a mathematical task. Therefore, the arithmetic task was administered after the collective categorisation task. If some participants had undertaken the recognition task immediately after the collective categorisation task, there may have been an advantage for recalling the most recently seen items. For this reason, all participants undertook the arithmetic task before the recognition task ensuring that all participants were given some distraction prior to undertaking the recognition task.

Participants

For all three tasks, there were 136 participants, 54 males and 82 females. They were recruited from the local community, the University of Wales, Bangor, subject panel, Unilever Research Laboratory subject panel, four English medium schools, and seven Welsh medium schools.

Each participant was allocated to one of two language groups, the Welsh group or the English group. Allocation to language group was in accordance to language background based on self-reports for adult participants and parental reports for child participants, obtained via questionnaires [Appendix III].

As previously stated, recruiting Welsh monolinguals is problematic. Welsh speakers are generally proficient English speakers too. For this reason, strict criteria were established for participant allocation to maximize the probability of tapping into the cognitive effects of language structure, if any, for speakers of each language.

Adult participants were allocated to the Welsh language group if they reported that their first language was Welsh and that the primary language spoken at home, throughout childhood, was Welsh (80 - 100% Welsh at home). Child participants were allocated to the Welsh language group if parental reports indicated that the child's first language was Welsh, the primary language spoken at home was Welsh (80-100% Welsh at home), and the child attended a Welsh medium school. Similarly, adult participants were allocated to the English language group if their first language was English and the primary language spoken at home throughout childhood was English (80 -100% English at home). Child participants were allocated to the English language group if parental reports indicated that the child's first language was English, the primary language spoken at home was English (80-100% English at home), and the child attended an English medium school.[1]

[1] All English-speaking participants in this study reported 100% English at home.

Participants fell into 4 age groups: Adults (mean age: 28;8, range:18;3 – 72;6), 10-year-olds (mean age:11;0, range:10;5 – 11;4), 8-year-olds (mean age:8;0, range: 7;1 – 8;10), and 4-year-olds (mean age: 4;9, range: 4;4 – 5;7).

The adult group consisted of 33 participants: 15 Welsh speakers (mean age: 32:4, range: 18;7 – 72;6, 13 females, 3 males); and 18 English speakers (mean age: 25;8, range: 18;15 – 40;2, 15 females and 3 males).

The group of 10-year-old children consisted of 35 participants: 16 Welsh speakers (mean age: 10;9, range: 10;6 – 11;4, 6 females and 10 males); and 19 English speakers (mean age 11;1, range: 10;5 – 11;7, 6 females and 5 males).

The group of 8-year-old children consisted of 40 participants: 20 Welsh speakers (mean age: 8;1, range: 7;4 – 8;10, 11 females and 9 males); and 20 English speakers (mean age: 7;9, range 7;1 – 8;7, 9 females and 11 males).

The group of 4-year-olds consisted of 28 participants: 13 Welsh speakers (mean age: 4;9, range: 4;6 – 5;7, 7 females and 6 males); and 15 English speakers (mean age 4;9, range 4;4 – 5;6, 9 females and 6 males).

Apparatus

A Macintosh PowerBook G3 was used to run a Psyscope 6.1 program, described below. The software recorded participants' typed responses and reaction times for all three tasks.

Collective categorisation task

Design

The aim of this task was to determine whether Welsh-speaking participants interpreted novel nouns as referring to collections more often than their English-speaking counterparts. Participants were presented with arrays of stimuli and asked to answer questions about the objects in a way that would indicate their interpretations of novel nouns and novel objects.

As previously stated, this task was designed to explore the factors that may influence participants' interpretations of novel nouns as referring to collections. It was necessary to control for the number of groups and number of items in each group to determine any effects of *number* on participants' responses. Similarly, the task controlled for syntax to determine any effects of syntactic cues on categorisation of novel objects.

To examine the issues outlined above, this task adopted a mixed factorial design. There were two between-subject variables, language and age. As mentioned earlier, there were two language groups (Welsh and English) and four age groups (adults, 10-year-olds, 8-year-olds, and 4-year-olds).

There were three within subject variables: the number of groups in an array ('groups'), the number of items per group ('items'), and syntax. The numbers of groups in an array were two or three. For Items, there were two test conditions and one control condition. The test conditions had five or fifteen items per group and the control condition had a single item per group. There were three syntax conditions, a singular condition, a plural condition, and an unmarked condition.

Non-linguistic Stimuli

Novel objects were created using Clarisworks drawing package. These novel objects were then used to create picture stimuli in accordance with the number of groups and number of items per group conditions. Sample stimuli are

shown in Figure 5.1. For each groups x items condition there were 18 different pictures with a total of 108 trials. There were 54 target trials and 54 distracter trials.

Linguistic Stimuli

Novel nouns

Each trial of the collective categorisation task required a novel noun. Two lists of 108 novel word forms were created - one consistent with the phonotactic characteristics of Welsh[2] and the other consistent with the phonotactic properties of English. Both novel word lists varied in word length and onset consonants [see Appendix IV for the list of novel word forms].

Syntax

The words in each list were randomly allocated to one of the three syntax conditions -- the singular condition, the plural condition, and the unmarked condition. This was done twice, thereby creating two modified versions of the original list for each of the two languages. For each of the two versions of word lists in each language, 36 of the forms were assigned to the "singular condition", 36 to the "plural condition" and 36 to the "unmarked condition".

1. Singular condition

To create forms that had a clear singular reference in the two languages, the following modifications were made to the 36 word forms assigned to the singular condition.

Welsh: For the Welsh novel forms assigned to the singular condition, the unitizer suffixes *–yn* and *–en* were added to the forms. Half of the 36 novel forms assigned to this condition were given the *–yn* ending and half the *–en* ending (e.g. nwl -> nwl*yn*, cedur -> cedur*en*).

English: The indefinite article *a/an* was added preceding the English

[2] Novel Welsh forms were obtained from Thomas (2002) and Gathercole et al (1999).

novel forms assigned to the singular condition (e.g. clopic -> *a* clopic, orbink -> *an* orbink).

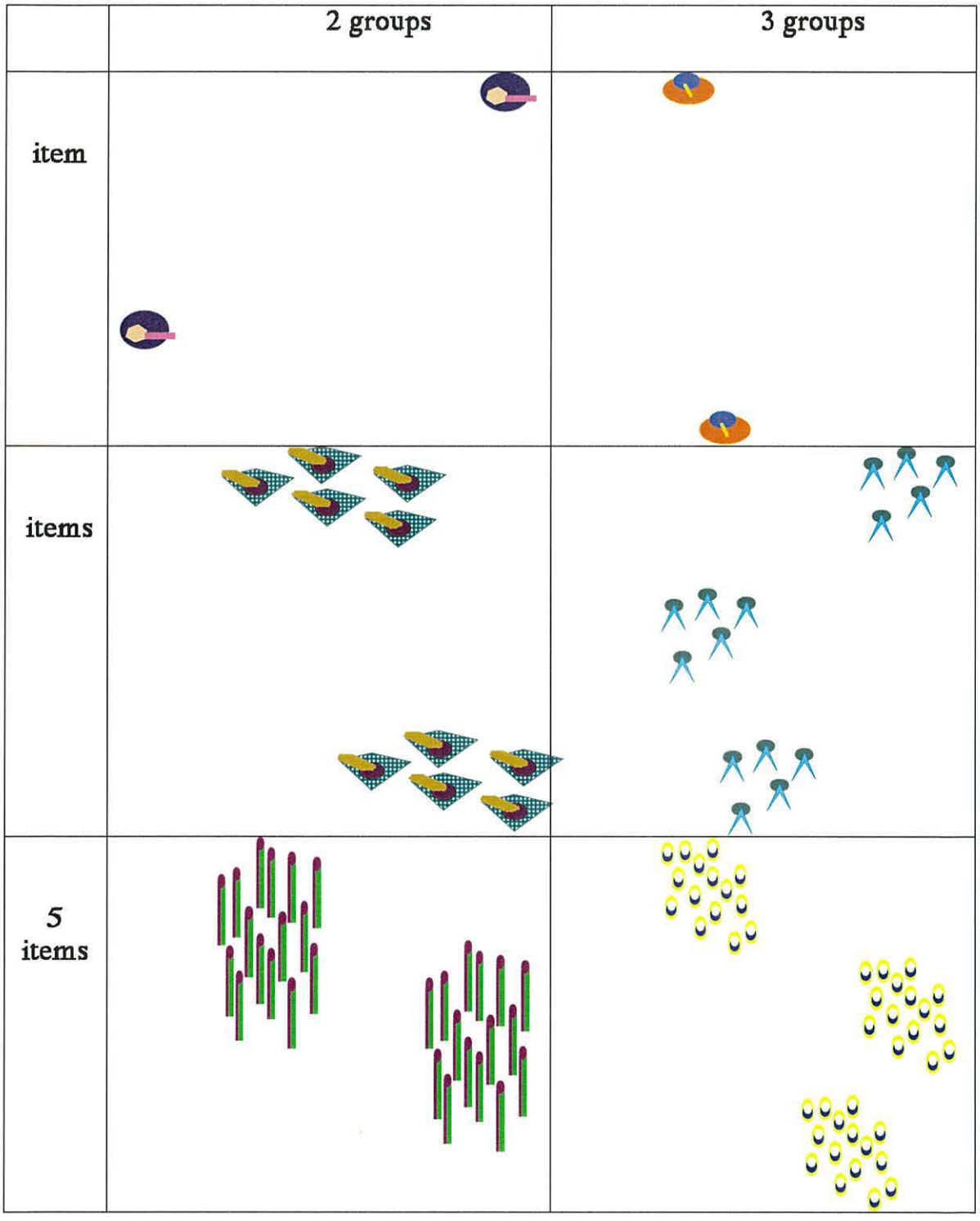


Figure 5.1 Examples of non-linguistic stimuli for the collective categorisation task.

2. Unmarked conditions

The novel forms were not changed for the unmarked condition for the Welsh and English word forms. [See Appendix IV for modified noun lists with singular and plural endings].

3. Plural Condition

To create word forms that had a clear plural reference in the two languages, plural inflections were added.

Welsh: One of the 13 plural inflections found in Welsh was added to the Welsh word forms assigned to the plural condition. Each of the 13 plural endings (King, 1993; Thomas, 1996; Thorne, 1993; Williams, 1980) was used at least once and none was used more than 5 times (e.g. pibot -> pibot*iau*, balan -> balan*od*, gleidd -> gleidd*ion*, dolyn -> dolyn*edd*, pall -> pall*i*).

English: the plural –s was added to all English novel forms in this condition (e.g. lig -> lig*s*, neaf -> neaf*s*, torkly -> torkl*ies*).

To ensure novelty of the novel noun lists, five Welsh speakers rated the Welsh lists and five English speakers rated the English lists on familiarity and likelihood of being possible names. Both lists were rated 100% as lists of possible names and 100% unfamiliar to the raters.

With each non-linguistic stimulus that participants saw on the screen, they heard a novel noun form. Two versions of a randomly selected order of presentation were created, each with a different version of the noun lists detailed above. Nouns within a given syntax condition were randomly assigned to visual stimuli within a given non-linguistic stimulus condition. That is, for each stimulus type as depicted in Figure 5.1, 6 singular, 6 plural and 6 unmarked nouns were assigned to a given set of stimuli. Two such randomly assigned matches of nouns with stimuli were prepared. Half the participants from each language and each age group were randomly assigned to each of the two versions.

Questions

For each trial, participants saw a stimulus and heard the novel noun associated with it as described below under Procedure. Once the stimulus disappeared, they heard a question about what they had seen. For the target trials participants were asked a target question, and for the distracter trials, participants heard one of three distracter questions. The questions were as follows:

Target Question:

English:

"How many were there?"

Welsh:

"Faint oedd yna?"

Distracter Questions:

English:

"What was the colour?"

"Was there one in the centre?"

"Was there one at the bottom?"

Welsh:

"Beth oedd y lliw?"

"Oedd 'na un yn y canol?"

" Oedd 'na un ar y gwaelod?"

The novel nouns and the questions were presented on the computer via digital sound files. This ensured consistent auditory presentation across subjects.

Procedure

Initially, participants were given general instructions regarding all three tasks. They were then given detailed instructions for this first task only. Instructions for the task were displayed on the screen for adults, 10-year-olds and 8-year-olds. The 4-year-olds received verbal instructions only. The adults, 10-year-olds and 8-year-olds were offered verbal instructions in addition to the written instruction to ensure understanding of the procedure. The instructions

were as follows:

English:

"This experiment involves 3 tasks.
For each task please look at the cross at the center of the screen to ensure that your gaze is central.
For this first task, after looking at the cross for a short time, images will appear for a few seconds.
At the same time you will hear a new name.
At the end of each image you will be asked a question about it.
Please type your response into the computer as accurately and as quickly as possible, then press the 'space bar'.
Instructions for the remaining tasks will appear as needed.
Press any key to continue."

Welsh:

"Mae yna dair rhan i'r arbrawf yma.
Ar gyfer bob rhan, gofynnir i chi edrych ar y groes yng nghanol y sgrîn er mwyn sicrhau edrychiad canolog.
Ar gyfer y rhan gyntaf, ar ôl edrych ar y groes, mi fydd lluniau newydd yn cael eu dangos am ychydig eiliadau.
Ar yr un adeg, mi glywch chi enw newydd.
Ar ddiwedd bob llun mi glywch chi gwestiwn amdano.
Teipiwch eich ateb i fewn i'r cyfrifiadur mor gywir â chyflym a sy'n phosib os gwelwch yn dda, yna pwyswch y 'space bar'.
Mi fydd hyfforddiant ar gyfer y rhannau eraill yn cael eu dangos fel sydd angen.
Pwyswch unrhyw fotwm i fynd ymlaen."

In addition, participants were verbally informed that if they had any questions they could ask the experimenter at any time. To commence the trials, participants pressed any key.

There were 54 target trials and 54 distracter trials in the task. Each involved the visual presentation of a picture stimulus, from one of the non-linguistic stimulus conditions, on the computer screen for 3500 milliseconds. Simultaneously, an audible presentation of the novel noun associated with the visual stimulus (e.g. "a clopic", "clopics" or "clopic") was presented at the onset of the picture presentation. On completion of the presentation, a blank screen would appear and a target or distracter question was presented audibly (e.g. "How many were there?"). Participants then typed their responses into the computer using the computer keyboard. For the next trial to commence, participants were required to press the spacebar on the keyboard.

For the 4-year-olds group the 108 trials were divided into 4 sets with a short break between the sets. This was to reduce fatigue and maintain the child's interest. Some child participants preferred not to type their responses into the computer. Consequently, the experimenter typed in the verbal responses of the child as soon as a response was made.

On occasions when participants told the experimenter that they had forgotten an answer they were asked to make a guess. Similarly, when participants asked the experimenter if their responses were correct, the experimenter told the participants to respond in a way that was comfortable for them and that the way they chose to interpret and answer the questions was entirely up to them.

Results

Two sets of data were obtained from this task, participants' responses and their reaction times.

Responses

Data for the target trials only were used for analysis. No significant differences were observed across the two language groups for the single item control conditions. Only the analyses carried out on the test conditions will be discussed further.

Participants' responses of "2" or "3" in the test conditions only (i.e. presentations of 2 or 3 groups of 5 or 15 items) were coded as a collective response. All other responses in the test conditions were coded as unit responses. A score of 1 was assigned to each collective response and a score of 0 was assigned to each unit response.

The coding of data was designed to ensure a conservative approach by avoiding incorrectly coding responses as collective responses. The reason for this is that low responses following presentations of 30 or 45 items did not always indicate collective responses. Some child participants overtly counted the individual objects in a single group and responded with "4" or "5". Since participants may have reported the number of items they had been able to count, in a relatively short presentation time, it was necessary to apply a strict coding criteria.

A language x age x syntax x items x groups ANOVA revealed a significant main effect of items, $F(1,129) = 6.03$, $p<.015$, and significant interactions of groups x syntax x age $F(3,129) = 4.73$, $p<.004$; items x syntax x age $F(3,129) = 2.79$, $p<.043$; and items x groups x syntax $F(1,129) = 7.00$, $p<.009$. A near significant interaction of language x age x groups x items, $F(3,129) = 2.36$, $p<.075$, was found.

The main effect of items illustrates that significantly more collective

responses were given when fifteen items (M = .500, SD = .401) were presented in a group than for five items (M = .470, SD = .470) per group.

The mean number of collection responses and standard deviations for the interactions are shown in Tables 5.1, 5.2, and 5.3 respectively. Post hoc analysis was not feasible due the large number of conditions within each variable. The experimentwise alpha level would be too great to perform multiple pairwise comparisons on the 3-way interactions thereby increasing the risk of making a type I error. For this reason, the presumed loci of the interactions are discussed and highlighted in the table of means for each interaction.

Table 5.1 Means and standard deviations for groups x syntax x age

Age	Singular				Unmarked				Plural			
	2 groups		3 groups		2 groups		3 groups		2 groups		3 groups	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
4yrs	.389	.450	.453	.611	.414	.416	.356	.416	.385	.426	.389	.453
8yrs	.410	.473	.450	.495	.415	.477	.417	.470	.415	.478	.413	.470
10yrs	.492	.491	.531	.494	.525	.489	.539	.490	.521	.500	.532	.494
Adults	.613	.438	.561	.484	.631	.440	.584	.480	.553	.484	.654	.437
Total	.477	.467	.498	.518	.496	.464	.477	.471	.470	.475	.498	.473

As can be seen from the Table 5.1, there are general trends indicating that more collective responses are given (1) with increasing age group, (2) for unmarked condition over the singular and plural condition and (3) for 3 groups over 2 groups. However, adults gave more collective responses for 2 groups over 3 groups for the singular and unmarked conditions but the reverse for the plural condition. And, the four-year-olds also gave more collective responses for 2 groups over 3 groups in the unmarked condition. These are highlighted in Table 5.1.

Similar trends are shown in Table 5.2. More collective responses are given (1) with increasing age group, (2) for unmarked over singular and plural

conditions, and (3) for 15 items over 5 items per group. However, as can be seen from the highlighted cell in Table 5.2, the 4-year-olds gave more collective responses for 5 items over 15 items in the unmarked and plural conditions. These data indicate that 4-year-olds more readily accept five items over fifteen items per group as collections in an unmarked or plural context. The adults also gave more collective responses for the 5 items over 15 items in the singular condition although this difference is negligible.

Table 5.2 Means and standard deviations for items x syntax x age

Age	**Singular**				**unmarked**				**Plural**			
	5 items		15 items		5 items		15 items		5 items		15 items	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
4yrs	.384	.467	.458	.594	[.409]	.412	[.362]	.420	[.405]	.444	[.370]	.435
8yrs	.415	.482	.446	482	.406	.465	.425	.480	.375	.460	.452	.488
10yrs	.491	.491	.532	.494	.506	.487	.558	.491	.505	.488	.547	.500
Adults	[.591]	.455	[.584]	.467	.568	.465	.647	.455	.588	.456	.619	.465
Total	.470	.477	.505	.508	.472	.462	.498	.473	468	.469	.497	.480

Table 5.3 Means and standard deviations for items x groups x syntax

Syntax	**2 groups**				**3 groups**			
	5 items		15 items		5 items		15 items	
	M	SD	M	SD	M	SD	M	SD
Singular	[.474]	.467	.479	.467	[.468]	.487	.528	.550
Unmarked	[.482]	.458	[.510]	.470	[.462]	.465	[.491]	.476
Plural	.446	.471	.494	.479	.488	.466	.507	.479
Total	.469	.465	.492	.476	.472	.473	.508	.502

General trends can be identified in Table 5.3: More collective responses are given for (1) 15 items over 5 items per group; (2) 3 groups over 2;

and (3) unmarked over singular and plural conditions. However, as highlighted in Table 5.3 more collective responses were given for 2 groups over 3 groups when 5 items were in a group and the context was singular or unmarked. Similarly, more collective responses were given for 2 groups over 3 groups when there were 15 items in a group and the context was unmarked.

The means and standard deviations for the near language x age x groups x items interaction are shown in Table 5.4.

Table 5.4 Means and standard deviations for language x age x groups x items

Language	**Age**	**2 groups**				**3 groups**			
		5 items		15 items		5 items		15 items	
		M	SD	M	SD	M	SD	M	SD
	4's	.4206	.4400	.3507	.4421	.3934	.4810	.4444	.6167
	8's	.3806	.4494	.4389	.4866	.4111	.4755	.4500	.4903
Welsh	10's	.5625	.5024	.6042	.5041	.5625	.4923	.6023	.4978
	adults	.4930	.4735	.5555	.4701	.5101	.4845	.5406	.4829
	4's	.3961	.4369	.4107	.4322	.3814	.4388	.3794	.4688
	8's	.4084	.4979	.4248	.4797	.3945	.4768	.4500	4947
English	10's	.4152	.4852	.4678	.4914	.4620	.4956	.4794	.5000
	adults	.6660	.4321	.6819	.4439	.6599	.4461	.6506	.4626

The table shows that the English-speaking adults gave more collective responses across the group and items condition than the Welsh-speaking adults; however, the Welsh-speaking 10-year olds gave more collective responses across the items and groups conditions than the English speakers of the same age.

Summary

The results show that generally, more collective responses were given with increasing age, for unmarked condition over the singular and plural conditions, for 3 groups over 2 groups, and for 15 items over 5 items per group. However, adults gave more collective responses for 2 groups over 3 groups for the

singular and unmarked conditions and the 4-yr olds gave more collective responses for 5 items over 15 items in the unmarked and plural conditions. More collective responses were given for 2 groups over 3 groups for the singular and plural conditions when only 5 items were presented in a group. The results also suggest that Welsh 10-year olds gave more collective responses than English 10-year-olds but the Welsh adults gave fewer collective responses than the English adults.

Reaction Times

The reaction times for the 4-year old children, whose responses were typed by the experimenter, were eliminated from the analysis to ensure that the data were not contaminated by the reaction times of the experimenter.

A language x age x syntax x items x groups ANOVA revealed significant main effects of language $F(1,102) = 5.54$, $p<.021$, age, $F(2,102) = 25.53$, $p<.000$, items, $F(1,102) = 76.74$, $p<.000$, groups $F(1,102) = 8.19$, $p<.005$ and syntax $F(1,102) = 4.51$, $p<.036$.

Significant interactions were found of items x age, $F\,2,102 = 9.16, p<.000$; groups x syntax, $F(1,102) = 27.04, p<.000$; items x syntax, $F\,1,102 = 4.05, p<.000$; groups x items x syntax, $F(1,102) = 14.10, p<.000$; groups x syntax x language x age, $F(2,102) = 4.05, p<.019$; groups x items x syntax x age, $F(2,102) = 7.07, p<.024$; and, groups x items x syntax x language x age, $F(2,102) = 7.07, p<.001$

The significant main effect of language showed that Welsh-speaking participants (M = 4197.61, SD = 5315.63) took significantly longer to categorise novel objects than the English-speaking participants (M = 3430.00, SD = 2152.17).

The main effect of age indicated that response reaction times are faster with increasing age. Post hoc analysis (LSD) shows that the adults (M = 2585.23, SD = 1240.18) were significantly faster than the 10-year-olds (M = 3507.45, SD =

1247.75, MD = -915.69, p<.014) and 8-year-olds (M = 5059.64, SD = 1907.92, MD = -2452.19, p<.000). The 10-year-olds were also significantly faster than the 8-year-olds (MD = -1536.49 p<.000).

The main effect of group showed that reaction times are faster for 2 groups (M = 3677.52, SD = 1696.26) than 3 groups (M = 3906.49, SD = 2050.21).

Post hoc analysis (LSD) for items showed that reaction times are significantly faster for single items (M = 2589.51, SD = 1027.44) than for five items (M = 3792.77, SD = 1973.51, MD = -164.70,p<.000) and fifteen items (M = 4993.75, SD = 3221.05, MD = -2339.88,p<.000). Reaction times were also significantly faster for five items than fifteen items (MD = -1175.18,p<.000).

Post hoc analysis of syntax indicated that participants were significantly faster to respond in the singular condition (M = 3575.62, SD = 2009.92) than the unmarked (M = 4022.3974, SD = 1935.96, MD = -435.28,p<.000) and plural condition (M = 3778.01, SD = 1815.49, MD = -195.70,p<.036), and significantly faster in the plural condition over the unmarked condition (MD = -239.58,p<.002).

The items x age interaction, shown in Figure 5.2, indicates that the speed of categorisation increases cumulatively with age and number of items. Post hoc analysis (LSD) shows that for single items, 8-year-olds were significantly slower than the adults (MD = 1100.44, p<.000) and the 10-year-olds (MD = 648.36, p<.003). The 10-year-olds were also significantly slower than the adults (MD = 452.08, p<.047). For five items, the 8-year-olds were significantly slower than the adults (MD = 2442.14, p<.000) and the 10-year-olds (MD = 1848.77, p<.000). For 15 items, 8-year-olds were significantly slower than the adults (MD = 3880.65, p<.000) and 10-year-olds (MD = 2159.42, p<.001), and the 10-year-olds were also significantly slower than the adults (MD = 1721.22, p<.013).

Figure 5.3 shows the group x syntax interaction and indicates that the speed of categorisation for 2 or 3 groups depended on the syntactic context of the

novel noun. Post hoc analysis (LSD) of the groups x syntax interaction indicated that for 2 groups, categorisation in the unmarked condition is significantly longer than for the singular (MD = 359.39, $p<.008$) and plural (MD = 523.77, $p<.000$) conditions. Similarly, for 3 groups, the unmarked condition is significantly longer than for the singular (MD = 815.56, $p<.000$) and plural (MD = 666.03, $p<.000$) conditions.

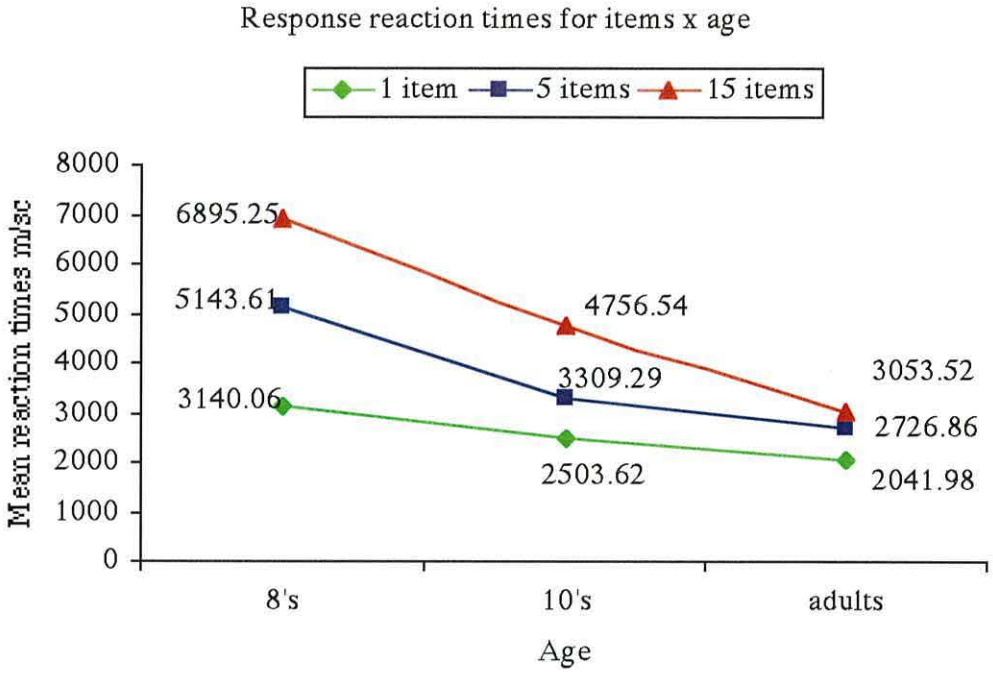


Figure 5.2 Mean reaction times for single, five, and fifteen items across each age.

The interaction of items x syntax showed that speed of categorisation across the syntax conditions depended on the number of items presented. This is shown in Figure 5.4. Post hoc analysis (LSD) indicated that categorisation of single items in the plural condition was significantly faster than in the unmarked condition (MD = -176.58, $p<.031$). Categorisation of 5 items was significantly faster in the singular condition than the unmarked (MD = -679.73, $p<.000$) and plural (MD = -382.81, $p<.036$) condition, and significantly faster in the plural condition than the unmarked condition (MD = -296.92, $p<.009$). For 15 items, categorisation was significantly faster in the singular condition than the unmarked

condition (MD = -520.04, p<.024).

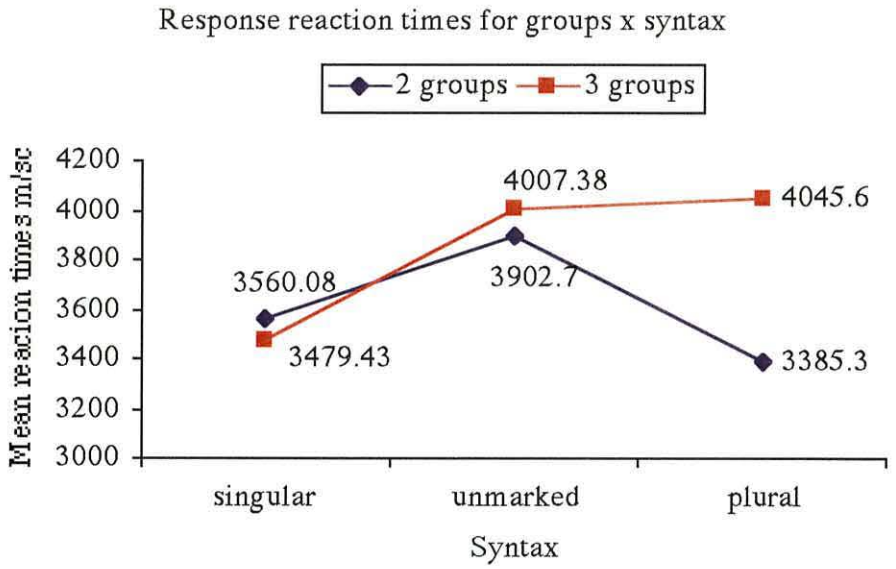


Figure 5.3 Mean reaction times for 2 and 3 groups across the syntax conditions

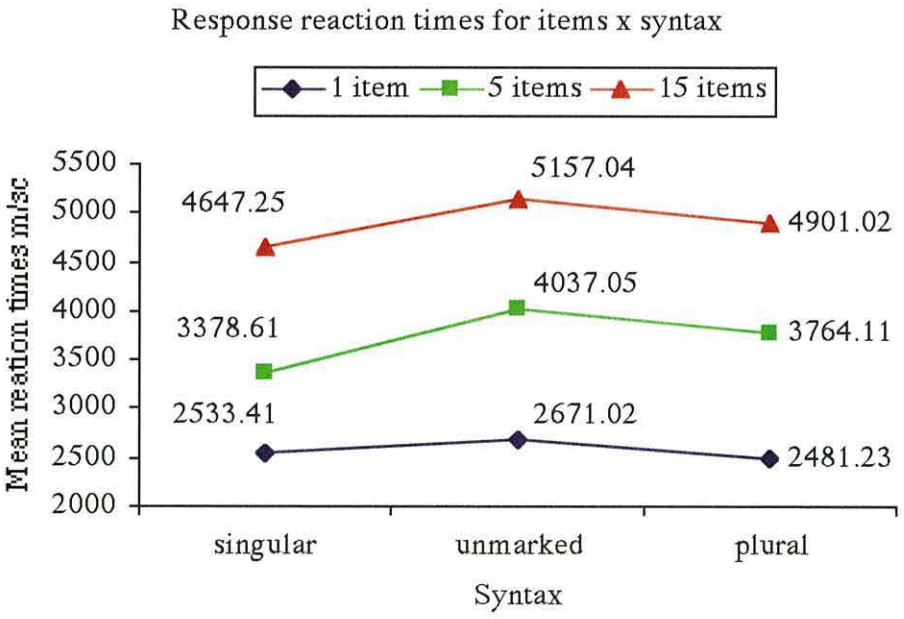


Figure 5.4 Mean reaction times for single, five and fifteen items for each syntax

condition

Post hoc analysis was not carried out on the 3-way, 4-way and 5-way interactions due to the increased experimentwise alpha level that would result from performing multiple pairwise comparisons. The presumed loci of the interactions are discussed and highlighted in the relevant Tables.

The means and standard deviations for the groups x items x syntax interaction can be seen in Table 5.6. A general pattern emerges from the means for this interaction that indicate (1) reaction times increase with increasing number of items (2) reaction times increase with increasing number of groups, and (3) reaction times are generally faster in the singular and plural conditions than in the unmarked condition. Participants took longer to categorise objects presented in the unmarked condition except for 3 groups of 15 items where participants took longer in the plural condition.

Table 5.6 Means and standard deviations for groups x items x syntax

	2 Groups					
	1 items		5 items		15 items	
Syntax	M	SD	M	SD	M	SD
Singular	2517.77	1341.52	3373.22	1957.15	4946.56	3606.37
Unmarked	2601.04	1271.44	3426.59	1600.03	5888.68	4156.29
Plural	2603.42	1116.45	3163.52	1797.26	4577.49	3365.01
	3 groups					
	1 items		5 items		15 items	
Syntax	M	SD	M	SD	M	SD
Singular	2591.52	1592.5	3503.97	3190.19	4520.65	3546.83
Unmarked	2789.42	1401.75	4810.64	3344.06	4618.60	3244.66
Plural	2433.88	1084.45	4479.27	2544.98	5410.49	4147.53

Means and standard deviations for the groups x syntax x language x age interaction are shown in Table 5.7. The general trends observed from the means of this interaction show that (1) Welsh speakers took longer to categorise objects than English speakers, (2) reaction times for categorisation are faster with increasing age, (3) categorising 3 groups over 2 groups took longer, and (4) categorising objects presented in the unmarked condition took longer than in the singular and plural conditions.

Table 5.7 Means and standard deviations for groups x syntax x language x age

Language	**Age**	**2 groups**					
		singular		unmarked		plural	
		M	SD	M	SD	M	SD
Welsh	8's	7429.81	3154.93	5389.88	2192.96	5361.12	3033.93
	10's	3673.47	1843.76	3940.33	1812.22	3287.07	1538.25
	Adults	2768.11	1160.98	2892.56	1490.00	2527.16	1015.87
English	8's	4202.68	1805.67	5408.93	3013.41	3871.44	164570
	10's	3858.63	4864.88	3483.46	1406.95	3239.05	1246.47
	Adults	2179.62	1575.83	2301.03	1369.28	2026.97	1127.37
Language	**Age**	**3 groups**					
		singular		unmarked		Plural	
		M	SD	M	SD	M	SD
Welsh	8's	5503.48	4358.78	4782.89	2792.80	5072.31	2899.31
	10's	3193.79	1832.48	3588.89	1659.77	3628.11	1872.46
	Adults	3034.15	1497.62	2349.99	1423.49	2468.9	1552.54
English	8's	3692.72	2523.53	6089.82	3283.95	5803.32	2992.05
	10's	3140.37	1571.50	3793.81	1893.87	3883.46	2018.33
	Adults	2051.61	1273.70	3438.85	2304.31	3250.18	1726.24

The highlighted cells in Table 5.7, however, indicate that the Welsh 8-

year-olds took longer to categorise 2 groups over 3 groups in the singular, unmarked and plural conditions whilst the Welsh 10-year-olds did so for the singular and unmarked conditions only. Similarly, the English 8-year-olds and 10-year-olds took longer to categorise 2 groups than 3 groups in the singular condition only.

Again, the trends observed from the means of this interaction are similar to those found in the previous interactions. These are (1) Welsh speakers took longer to categorise overall than the English speakers, (2) reaction times decreased with age, (3) reaction times decreased with increasing number of groups and items, and (4) reaction times were longer in the unmarked condition than in the singular and plural conditions.

Means and standard deviations for the groups x items x syntax x language x age interaction are shown in Table 5.9 for Welsh and Table 5.10 for English. The trends illustrated by this interaction reflect those previously discussed. However, as can be seen from the highlighted means in the two Tables, Welsh and English 8-year-olds took longer to categorise 3 groups of 5 items over 3 groups of 15 items in the unmarked condition.

Table 5.8 Means and standard deviations for groups x items x syntax x age

Syntax	Age	2 groups					
		1 item		5 items		15 items	
		M	SD	M	SD	M	SD
Singular	8's	2814.23	1024.64	4515.08	2396.16	6891.66	4529.94
	10's	2513.30	1045.19	3017.93	1180.01	4793.71	2537.15
	Adults	2173.47	1809.27	2385.15	1235.19	2811.06	1474.35
Unmarked	8's	3177.33	1499.86	4326.92	1746.05	8693.96	4716.30
	10's	2569.51	1055.61	3321.80	1145.65	5163.75	2620.95
	Adults	1954.59	796.58	2470.30	1203.92	3313.29	2437.43
Plural	8's	3429.53	1145.78	4154.27	2277.15	6263.54	4200.49
	10's	2355.84	728.79	2905.87	803.20	4518.99	2552.44
	Adults	1879.08	726.60	2255.60	958.58	2652.38	1442.28
Syntax	**Age**	**3 groups**					
		1 item		5 items		15 items	
		M	SD	M	SD	M	SD
Singular	8's	3169.75	246.71	4907.32	4641.06	6107.43	4273.32
	10's	2426.52	879.40	2717.29	1032.30	4348.01	3133.27
	Adults	2076.26	1138.66	2639.64	1638.10	2826.50	1830.56
Unmarked	8's	3365.57	1688.63	6988.48	4127.91	5955.03	3532.63
	10's	2656.07	981.56	3942.70	1666.36	4439.13	2590.15
	Adults	2244.94	1145.24	3116.42	1965.22	3225.81	2912.30
Plural	8's	2883.93	1057.95	5969.63	2853.50	7459.88	5190.01
	10's	2428.92	1026.93	3863.46	1704.63	5151.37	3101.50
	Adults	1909.37	952.31	3341.74	2013.95	3258.58	2095.71

Table 5.9 Means and standard deviations for groups x items x syntax x language x age for Welsh

Language	Syntax	Age	2 groups					
			1 item		5 items		15 items	
			M	SD	M	SD	M	SD
	Singular	8's	3148.23	1174.72	5558.00	2863.23	7127.65	5426.84
		10's	2620.13	1003.54	3264.82	1483.53	5135.47	3044.21
		Adults	2093.62	628.35	2872.54	1352.19	3338.17	1502.41
	Unmarked	8's	3307.32	937.91	4741.10	1824.82	8121.22	3816.15
		10's	2959.60	1312.81	3282.67	1169.97	5578.73	2953.88
		Adults	2005.79	607.37	2619.65	947.74	4052.23	2914.89
	Plural	8's	3675.23	1093.15	4962.02	2772.93	7443.10	5157.28
		10's	2317.88	644.24	2900.00	873.88	4557.29	2895.96
		Adults	1894.27	921.89	2433.48	958.58	3286.00	1644.38
Welsh	**Syntax**	**Age**	**3 groups**					
			1 item		5 items		15 items	
			M	SD	M	SD	M	SD
	Singular	8's	3215.48	1990.56	5908.27	6113.10	7386.70	4972.69
		10's	2571.38	797.82	2649.18	943.13	4360.82	3756.48
		Adults	2163.81	583.59	3419.65	1996.81	3519.96	1912.45
	Unmarked	8's	3862.03	2113.50	7556.15	4360.24	6851.28	3378.11
		10's	2724.31	935.03	4130.40	1896.32	4526.73	2850.25
		Adults	2371.50	759.19	3687.50	2481.79	4257.54	3671.94
	Plural	8's	3176.10	1277.21	6869.25	3324.62	7364.60	4374.32
		10's	2350.56	799.50	4365.44	1742.30	5436.38	3513.20
		Adults	1996.08	613.67	3717.46	2232.08	3837.00	2332.98

Table 5.10 Means and standard deviations for groups x items x syntax x language x age for English

Language	Syntax	Age	2 groups					
			1 item		5 items		15 items	
			M	SD	M	SD	M	SD
	Singular	8's	2480.23	734.99	3472.15	1139.01	6655.67	3543.01
		10's	2428.96	1096.56	2823.02	866.08	4523.89	2102.23
		Adults	2244.44	2448.40	1951.91	962.30	2342.52	1316.78
	Unmarked	8's	3047.35	1924.13	3912.73	1602.10	9266.70	5513.99
		10's	2261.54	689.58	3352.70	1157.30	4836.12	2355.97
		Adults	1909.07	949.52	2337.55	1407.61	2656.46	1750.70
	Plural	8's	3183.83	1093.15	3346.52	1257.39	5083.98	2586.55
		10's	2317.88	644.24	2910.51	767.29	4488.75	2327.87
		Adults	1894.27	921.89	2097.48	1491.50	2089.17	968.71
English	Syntax	Age	3 groups					
			1 item		5 items		15 items	
			M	SD	M	SD	M	SD
	Singular	8's	3124.02	2343.62	3906.38	2175.61	4828.17	3051.35
		10's	2313.16	944.16	2771.07	1120.24	4337.89	2650.09
		Adults	1998.43	1484.15	1946.29	780.45	2210.09	1556.51
	Unmarked	8's	2869.10	931.26	6420.80	3909.74	5058.77	3537.38
		10's	2602.19	1038.94	3794.51	1497.09	4369.96	2443.27
		Adults	2132.44	1471.38	2608.80	1217.61	2308.71	1635.49
	Plural	8's	2591.77	697.45	5070.00	1989.11	7555.15	6011.36
		10's	2490.79	1194.61	3467.16	1609.49	4926.37	2813.27
		Adults	1832.30	1189.45	2829.98	1699.44	2744.43	1768.73

Summary

The results of the reaction time data show four main trends. First, Welsh speakers, at each age, took significantly longer to categorise than the English speakers. Note that this finding is especially robust for 5 and 15 items when compared to the single item control condition. Second, reaction times decreased with age. Third, reaction times increased overall with increasing number of groups and with increasing number of items. Last, reaction times overall were longer in the unmarked condition than in the singular and plural conditions.

For 2 or 3 groups of 5 items and 2 groups of 15 items, participants took longer to categorise objects presented in the unmarked condition than the singular and plural conditions. But for 3 groups of 15 items participants took longer in the plural condition than the unmarked and singular conditions.

Welsh 8-year-olds took longer to categorise 2 groups over 3 groups across all syntax conditions and Welsh 10-year-olds also took longer to categorise 2 groups over 3 groups in the singular and unmarked conditions. English 8-year-olds and 10-year-olds took longer to categorise 2 groups than 3 groups in the singular condition only. Both Welsh and English 8-year-olds took longer to categorise 3 groups of 5 items than 3 groups of 15 items in the unmarked condition.

Discussion

The results indicate that several factors play a role in the categorisation of novel objects and that often multiple cues are used to interpret novel nouns.

There was no significant effect of language on categorisation responses suggesting that, overall, Welsh- and English-speaking participants overtly categorised novel objects in a similar way. Welsh and English speakers interpret novel nouns as referring to collections about equally often. This result does not support the hypothesis that differences in language structure would effect the way that participants thought about novel objects or the meanings of the new words.

On the other hand, a clear language effect was found from the reaction times data, suggesting that although overt categorisations of novel objects are similar across the two languages, language structure may indeed influences the categorisation process. The categorisation process took significantly longer for Welsh speakers. Welsh speakers may be taking longer to decide how they will interpret novel words and categorise novel objects because they have more choices available to them about the possible referents of a new word than English speakers do.

Furthermore, a near significant interaction for the categorisation responses also suggests that language structure does have some role to play in the categorisation process. Primarily, the influence of language structure on categorisation behaviour emerges at different ages in the two language groups.

A possible explanation for the weak language effects on overt categorisation is that Welsh speakers are predominantly proficient English speakers too. The knowledge that Welsh speakers have of the structure of English may interfere with their categorisation behaviour. Therefore, the absence of robust differences between the two languages on collective categorisation does not mean that language structure has no role in categorisation for language learning. On the contrary, the reaction time data clearly show that Welsh speakers are

influenced by the options that their language gives them when interpreting novel words and categorising novel objects.

Age also seems to be a key factor in categorisation behaviours. The general trend that emerges across the two language groups is that with increasing age, participants were more likely to categorise novel objects as collections. This shows that the youngest participants were less willing to interpret novel nouns as referring to collections than the older participants were. Again, this may be a consequence of increasing knowledge of the structure of their language.

This is also supported by the reaction time data. With increasing age, participants were faster at categorising objects and deciding how to interpret novel nouns. This may be due to greater experience with language. The more the participants know about their language, the quicker they will be at learning new words. If the structure of the language being learned guides the way that speakers categorise objects and learn new words, it would be expected that the more experience people have with the structure of their language, then the quicker they will be at making decisions about novel nouns.

The categorisation responses and reaction time data also indicate that object properties and syntactic cues are important factors in interpreting novel nouns as collections. The general pattern indicates that novel objects are categorised as collections more often with increasing number of items and with increasing number of groups across both languages. The results are consistent with the hypothesis that increasing the numbers of items and groups available for categorisation facilitates collective categorisation. These cues, particularly alongside additional syntactic cues, act to facilitate categorisation of objects as collections and ultimately allow children and adults to interpret novel nouns as referring to collections. Object properties also seem to affect the speed at which participants categorise novel objects. Participants took longer to categorise novel objects when more objects were available. Given that participants are more willing to consider a collective interpretation of the novel nouns but took longer to

categorise when more items are available, it would seem that increasing the number of items offers greater choices about how to think about the objects and novel nouns. Again, this supports the idea that the number of items available does influence categorisation.

Syntax was also shown from the categorisation responses and reaction time data to be an important cue in learning the meanings of new words. The data are consistent with the hypothesis that unmarked syntax would elicit greater collective responses than the singular and plural syntax. Since an unmarked noun is ambiguous as to the referent, participants were more willing to adopt a collective interpretation in this case. For the singular and plural syntax conditions, additional cues such as object properties appeared to be used to aid the categorisation process. The reaction time data support the idea that unmarked syntax offers more options about possible referents when interpreting new words. Participants took longer to categorise items in the unmarked condition than the singular and plural conditions, especially when more items were available. This suggests that participants are more willing to accept a collective interpretation of a noun when there is no clear indication from the noun as to its referent. Unmarked syntax, together with increasing number of items, seems to facilitate collective interpretation. This added option allowed participants to think about the objects as collections and slowed down the speed of categorisation.

Many factors then play a role in categorising novel objects when acquiring the meanings of new words. And often multiple cues are required. Although it seems that language differences do not directly affect overt categorisation, the structure of the language being learned may contribute, as one of the many factors, to the overall categorisation process for language learning.

Arithmetic Task

Design

During a pilot study, data suggested that Welsh-speaking participants, in comparison to English speaking participants might give lower estimates of the number of items they had seen when presented with groups of stimuli. For example, when presented with three groups of fifteen items, English speakers may give an estimate of 60 items while Welsh speakers may give estimates of 30. An arithmetic task was designed to eliminate the possibility that differences between the two language groups on the collective categorisation task were attributable to differences in basic arithmetic skills. Adults, 10-year-olds and 8-year-olds were asked to solve twenty basic arithmetic problems. The 4-year-olds were asked to count five sets of shapes on the screen.

Stimuli

Two lists of twenty arithmetic problems were created, one oriented towards adults and the other towards 8- and 10-year-olds. Each list contained eleven multiplication problems alongside three addition problems, three division problems, and three subtraction problems. The emphasis was on multiplication to establish that differences in performance on the collective categorisation task was not a result of an inability to multiply the number of groups by the number of items in each group. The test for the adult participants is shown in Table 5.11. The test for the 8- and 10- year-olds is shown in Table 5.12.

The counting task used for the children in the 4-year-olds group was a random selection of 5 counting tests out of a possible 10. Each test involved a picture of shapes with 1–10 items per picture. All child participants in the 4-year-olds group were asked to count a minimum of 6 items at least once during the task. Example stimuli for the counting test are shown in Figure 5.5.

Table 5.11. Arithmetic task stimuli for the adults

3 x 15 =	5 x 3 =	2 x 3 =	10 + 20 =
2 x 10 =	2 x 5 =	30 / 2 =	3 + 2 =
2 x 25 =	20 x 3 =	10 / 5 =	45 – 15 =
10 x 3 =	3 x 5 =	45 / 3 =	30 – 5 =
15 x 2 =	2 x 15 =	15 + 5 =	20 – 3 =

Procedure

On completion of the final trial in the collective categorisation task, instructions for the arithmetic task appeared on the screen for the adults, the 10-year-olds and the 8-year-olds as follows:

English:

"You may take a break now.

For this second task you will see brief presentations of basic arithmetic.

Please type your answer into the computer and press the 'space' bar.

Press any key to continue"

Welsh:

"Cymerwch rest am funud os mynnwch chi.

Ar gyfer yr ail ran mi welwch chi symiau syml am ychydig eiliadau.

Teipiwch eich ateb i fewn i'r cyfrifiadur ac, yna pwyswch y 'space bar'.

Pwyswch unrhyw fotwm i fynd ymlaen."

Table 5.12. Arithmetic task stimuli for 8- and 10-year-olds

2 x 10 =	5 x 3 =	2 x 3 =	10 + 20 =
4 x 2 =	2 x 5 =	8 / 2 =	3 + 2 =
2 x 7 =	6 x 2 =	10 / 5 =	15 – 5 =
10 x 3 =	3 x 5 =	6 / 3 =	10 – 8 =
3 x 3 =	4 x 4 =	15 + 5 =	6 – 2 =

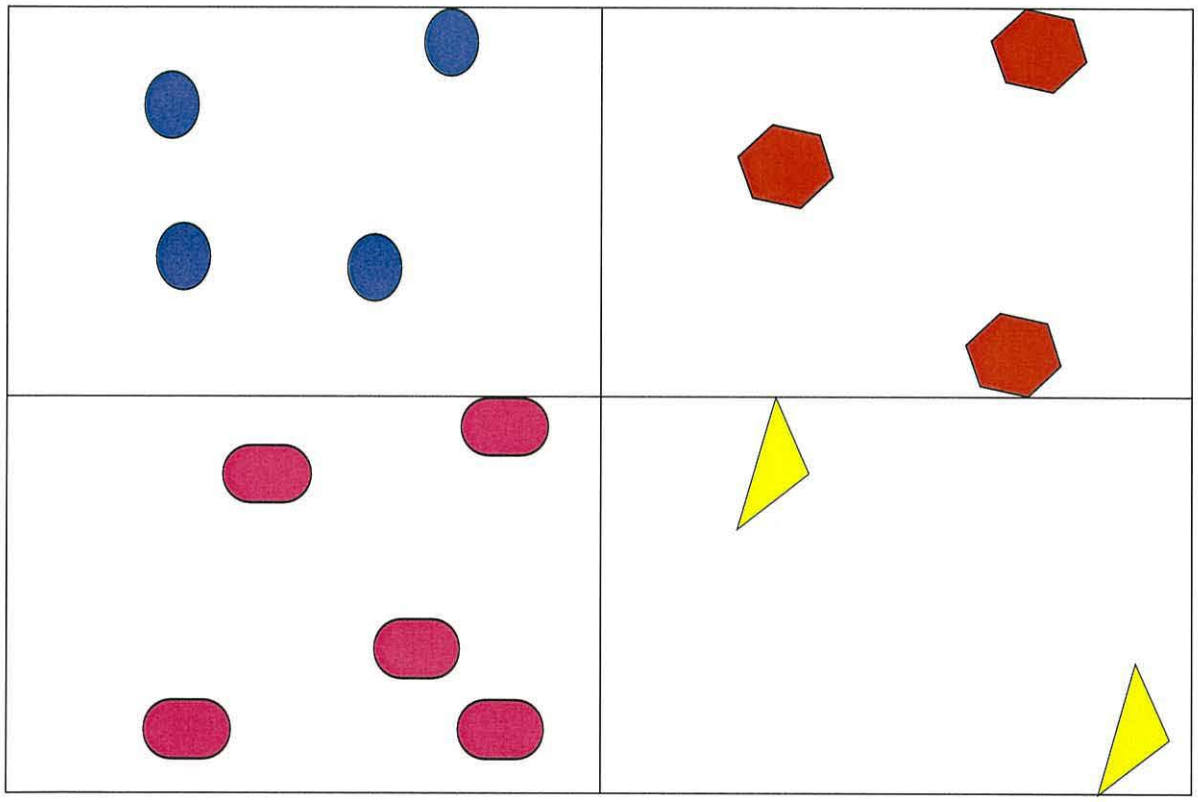


Figure 5.5 Example stimuli used for the 4-year-olds

Participants pressed a key to initiate the trials. They would see a brief presentation of the arithmetic problems in the center of the screen. The problems were presented in a random order and remained on the screen for 1000 milliseconds. Participants typed their responses into the computer and pressed the

space bar. As soon as the space bar was pressed the next trial would commence. This continued until all 20 trials were completed.

For the 4-year-olds, instructions were given verbally as follows:

English:

" Now you will see some shapes on the screen. What I would like you to do is point to each shape and count them. They will stay on the screen until you have finished counting them".

Welsh:

"Rwan mi 'nei di weld siapau ar y sgrîn. Be dw' i isio i ti 'neud ydi pwyntio at y sgrîn a cyfri nhw. Mae'r siapiau yn aros tan i ti orffen cyfri."

Participants pressed a key to start the trials. They would see a number of identical coloured shapes on the screen and would proceed to point at the shapes and count them out aloud. Responses were typed into the computer and a press of the space bar would initiate the next trial. This continued until a random selection of 5 trials had been completed.

Again, if child participants preferred not to type their responses into the computer, the experimenter typed in the verbal responses of the child as soon as a response was made.

Results

Only the accuracy of responses was analyzed for this task. Reaction times were considered superfluous to the question being asked. Responses were coded so that correct responses were given a score of 1 and incorrect responses were given a score of 0. The 4-year-olds participants had a counting task with scores out of 5 while the 8- and 10-year-olds together with the adults had an arithmetic test with scores out of 20. To compare arithmetic abilities across the two language groups a one-way ANOVA on the raw scores were carried out for each age group. No significant differences were found between the two language groups at any age.

Discussion

These data are consistent with the hypothesis that Welsh- and English-speaking participants do not perform differently on a basic arithmetic task. Since the tests were designed specifically for the different age groups no effect of age was expected. This confirms that any differences between the two language groups on the categorisation task and the recognition task are not attributable to differences in performances of the two language groups on basic arithmetic skills.

Recognition Task

Design

The recognition task was designed to examine whether participants' memory for the objects, introduced in the collective categorisation task, was affected by their initial interpretation of the novel nouns associated with those objects. Participants that interpret a novel noun as referring to a collection may subsequently have poorer memory for the individual objects that form part of that collection. Participants that interpret novel nouns as referring to individual objects contained within the collection may have a better memory for those same objects.

The recognition task consisted of 108 randomly presented trials. Each trial involved the presentation of a single item on the screen. Half of the items were previously seen target items from the collective categorisation task and half were novel items.

Non-linguistic Stimuli

This task required 54 novel objects, created on ClarisWorks drawing package, and 54 previously seen objects from the collective categorisation task. The previously seen items were all those used in the target trials of the collective categorisation task. Thus, the items had been seen previously as 2 or 3 single control items, or in 2 or 3 groups of 5 or 15 items. Examples of previously seen and unseen items are shown in Figure 5.6.

Previously seen items	Unseen items
	
	

Figure 5.6. Example stimuli from the recognition task.

Procedure

On completion of the arithmetic task, participants were given a short break to minimize fatigue. Instructions for the recognition task appeared on the screen for adults, the 10-year-olds and the 8-year-olds on completion of the arithmetic task. The 4-year-olds were given the instructions verbally. The instructions were as follows:

English:

"You may take a break now.

For this last task you will again see brief presentations of images.

You are asked to determine whether you have previously seen the image in the first task.

At the end of each trial press the "M" button for a YES response and the "Z" for a NO response.

If you are left handed, please use the "Z" button for YES and the "M" button for NO.

Press any key to continue."

Welsh:

"Cymerwch rest am funud os mynnwch chi.

Ar gyfer y rhan yma mi welwch luniau am ychydig eiliadau un waith eto.

Gofynnir i chi benderfynu os ydych wedi gweld y lluniau o'r blaen yn y rhan gyntaf.

Ar ddiwedd bob llun pwyswch fotwm 'M" am ateb DO a botwm 'Z" am ateb NADDO.

Os ydych yn lawchwith, pwswch "Z" am ateb DO a "M" am ateb NADDO.

Pwyswch unrhyw fotwm i fynd ymlaen".

Participants pressed a key to commence the first of the108 trials. Participants would then see a single item presented in the center of the screen for 1000 milliseconds. Participants then pressed a key for a "yes" or "no" response and the next trial would begin. This process continued until all the trials were completed.

For the 4-year-old child participants that preferred not to type their responses into the computer, the experimenter typed in the verbal responses of the child as soon as a response was made.

Results

The recognition task yielded two data sets, participants' 'yes' or 'no' responses and their reaction times. There were no differences between the two language groups or age groups for previously unseen items. Only the data for the previously seen items were analyzed to compare the effects of presentation status of the items (e.g. groups of items v single items) on the speed and accuracy of object recognition. Two 4-year-old English-speaking children did not complete this task due to fatigue. Thus, their responses were eliminated from the analysis.

Responses

Responses were coded for accuracy. The proportions of correct responses were calculated for each condition.

A language x age x syntax x items x groups ANOVA revealed significant main effects for age F (3,127) = 3.75 $p<$. 013, items F (1,127) = 3.75 $p<$. 002, and groups F (1,127) = 28.82, $p<$ .000.

Significant interactions were found for language x syntax, F (1,127) = 5.32,$p<$.023, age x items F (3, 127) = 2.71, $p<$.048, age x groups, F (3,127) = 10.32,$p<$.000, items x groups, F (1,127) = 26.89, $p<$ .000, groups x syntax, F (1,127) = 15.60,$p<$.000, age x items x groups F (3,127) = 11.02, $p<$.000, and age x groups x syntax F (3,127) = 5.16, $p<$. 002.

Post hoc analysis (LSD) of age shows that the 10-year olds (M= .5623, SD = .2938) recognized previously seen objects significantly more often than the 4-year-olds (M= .4277, SD = .3330, MD = .135, $p<$.006) and 8-year-olds (M= .4562, SD = .3204, MD = .108, $p<$.028). The adults (M= .5351, SD = .2945) also recognized objects more often than the 4-year-olds (MD = .106, $p<$.015).

Post hoc analysis (LSD) of items shows that items previously seen in groups of 15 items (M = .4631, SD = .3224) were recognized significantly less often than items seen as single items (M= .5165, SD = .3170, MD = .049, $p<$.015) and five items (M= .5142, SD = .3197, MD = .049, $p<$.014).

The effect of groups shows that items previously seen in 3 groups were recognized significantly more often than items previously seen in 2 groups.

Post hoc analysis of language x syntax interaction showed no significant differences between the two language groups for singular, unmarked or plural conditions. This interaction is shown in Figure 5.7.

For the age x items interaction, post hoc means comparisons (LSD) indicated that the 10-year-olds recognized the items previously seen as single items significantly more often than the 4-year-olds (MD = .1618, $p<.005$) and 8-year-olds (MD = .1119, $p<.028$). The 10-year-olds also recognized the items previously seen in groups of 5 items significantly more often than the 4-year-olds (MD = .1506, $p<.007$) and 8-year-olds (MD = .1163, $p<.020$). The adults recognized the items previously seen in groups of 5 items significantly more often than 4-year-olds (MD = .1627, $p<.004$) and 8-year-olds (MD = .1283, $p<.011$). The adults also recognized the items previously seen in groups of 15 items significantly more often than the 8-year-olds (MD = .097, $p<.039$). This interaction is shown in Figure 5.8.

Post hoc analysis (LSD) of the age x groups interaction indicated that the adults recognized the items previously seen in 3 groups significantly more often than 4-year-olds (MD = .1899, $p<.000$) and 8-year-olds (MD = .1328, $p<.003$). The 10-year-olds also recognized the items previously seen in 3 groups significantly more often than the 4-year-olds (MD = .1768, $p<.000$) and 8-year-olds (MD = .1196, $p<.008$). No significant differences were found across age for items previously seen in 2 groups. This interaction can be seen in Figure 5.9.

The items x groups interaction is shown in Figure 5.10. Post hoc analysis (LSD) showed no significant difference for items presented as 2 and 3 single items. But the items presented in 3 groups were significantly more accurately recognised than 2 groups when there were 5 items per group (MD = .067, $p<.000$) and 15 items per group (MD = .124, $p<.000$).

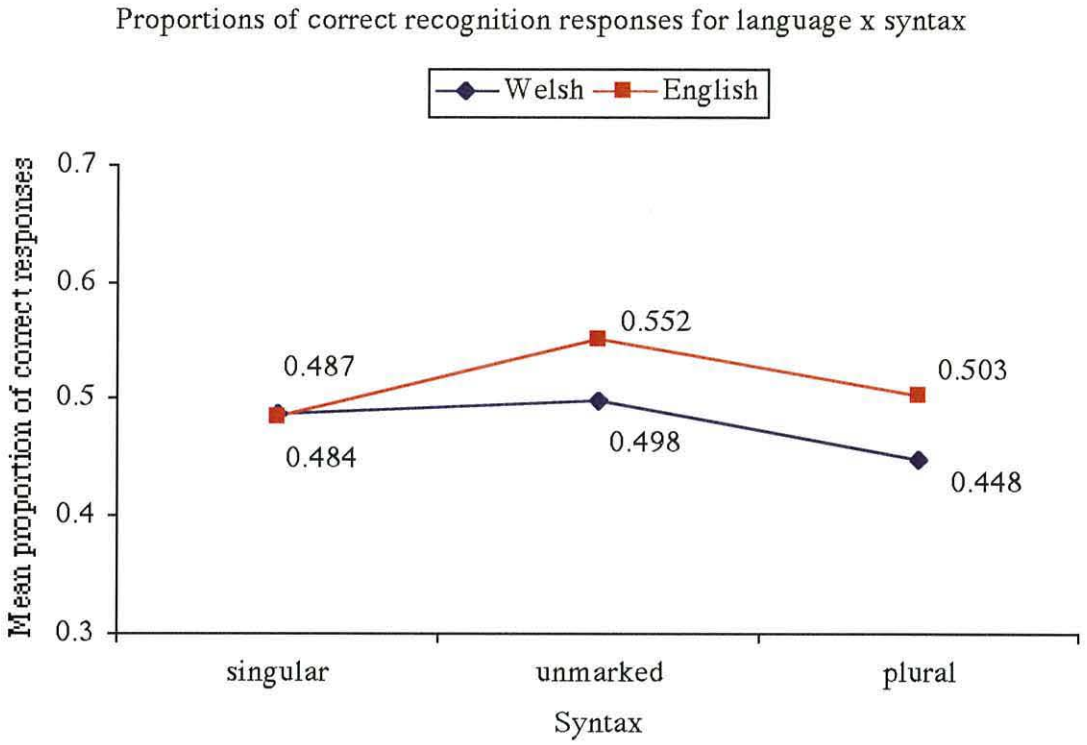


Figure 5.7 Accuracy of object recognition for each language at each syntax condition

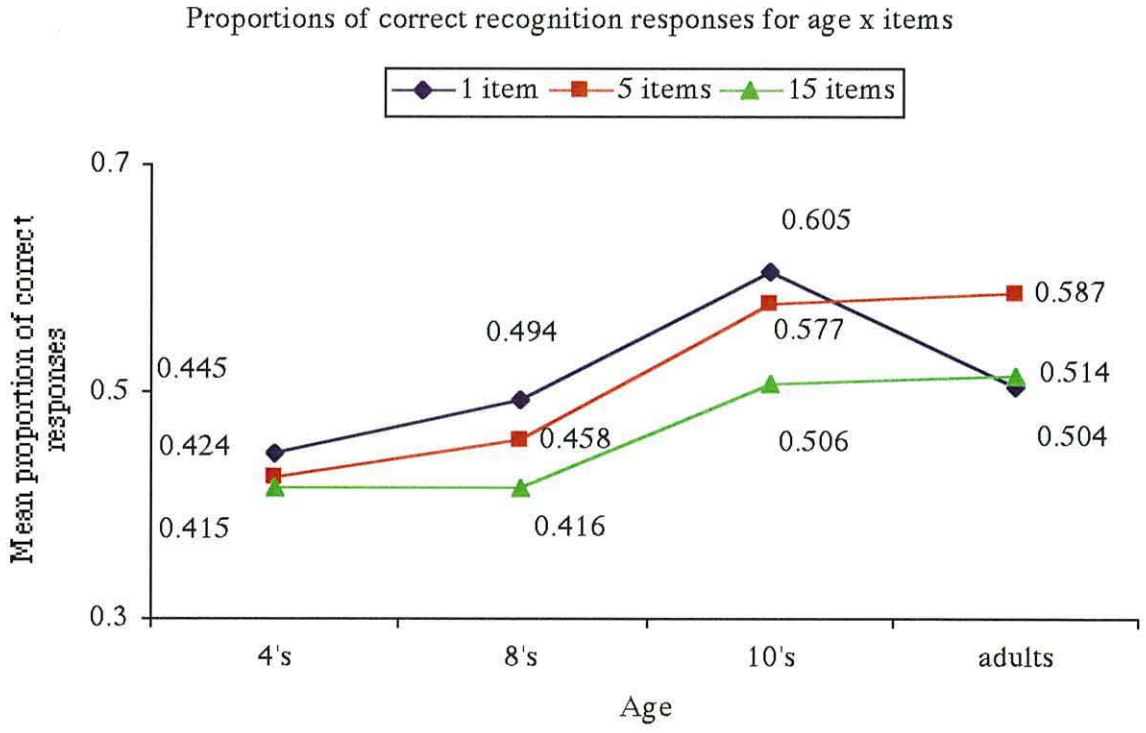


Figure 5.8 Accuracy of object recognition at each age for 1, 5 and 15 items

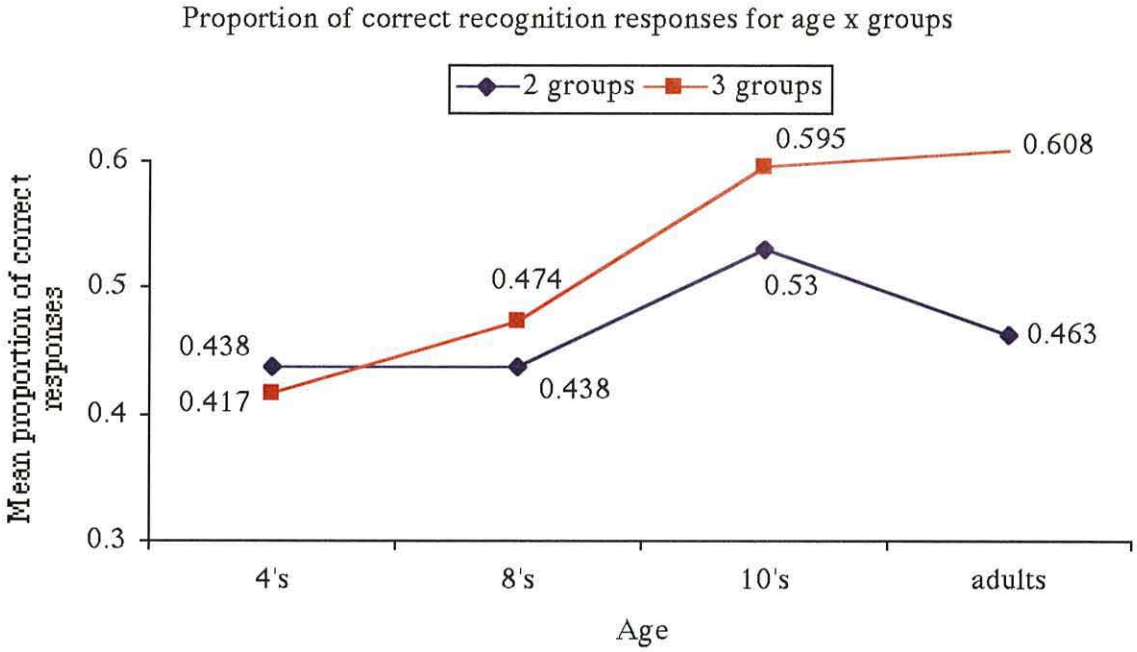


Figure 5.9 Accuracy of object recognition at each age for 2 or 3 groups.

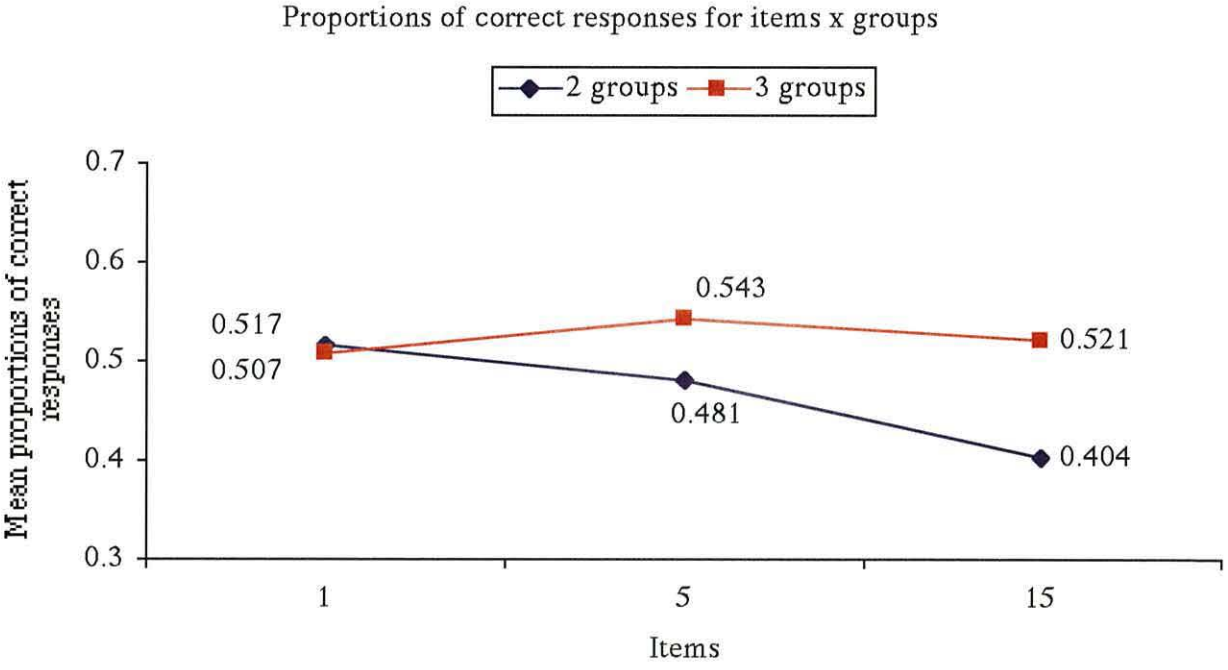


Figure 5.10 Accuracy of object recognition for 2 and 3 groups of 1, 5 and 15 items

Post hoc analysis (LSD) of groups x syntax showed that for 2 groups, items previously seen with an unmarked novel noun were recognized significantly more often than those with a singular (MD = .081, $p<.000$) or plural novel noun (MD = .045, $p<.021$), and items seen with a plural noun were recognized significantly more often than the singular (MD = .037, $p<.036$) nouns. For 3

groups, items presented with a plural novel noun were recognized significantly more often than those with a singular (MD = .056, $p<.001$) and unmarked (MD = .062, $p<.001$) novel noun. Figure 5.11 shows this interaction.

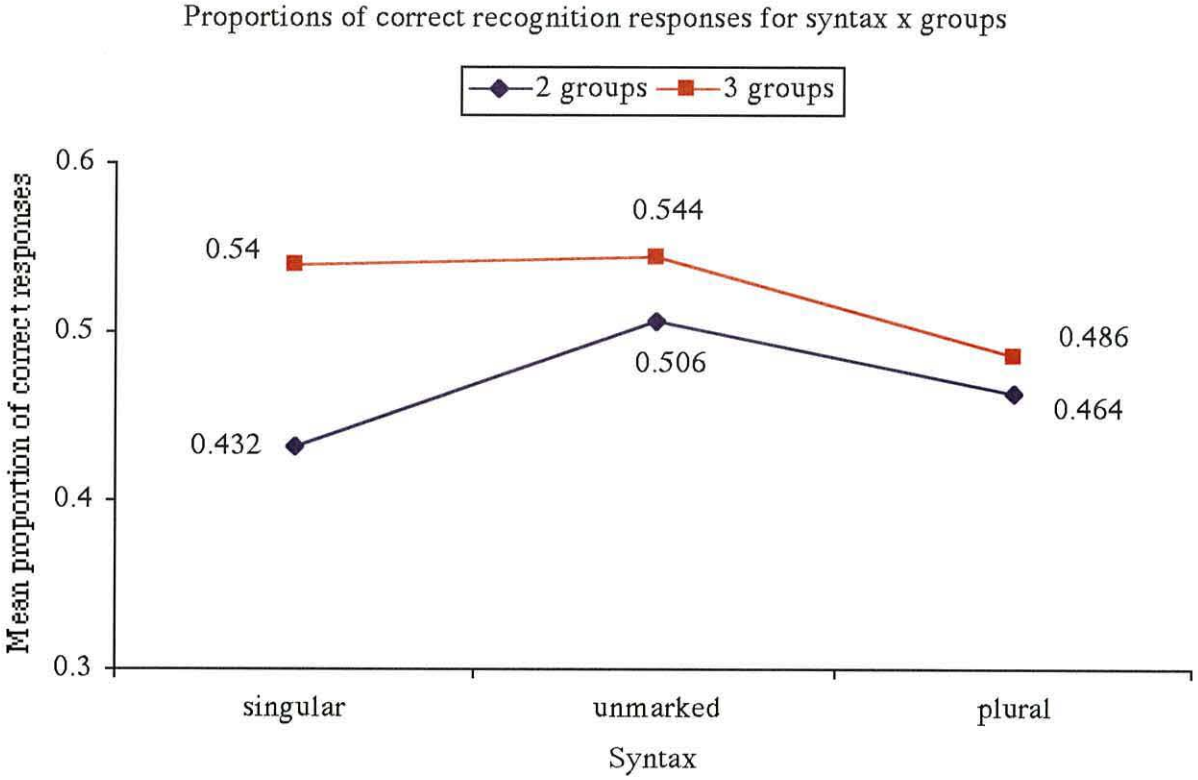


Figure 5.11 Accuracy of object recognition for 2 and 3 groups in each syntax condition

Post hoc analysis of the age x items x groups and the age x groups x syntax interactions were not feasible due to number of pairwise comparisons required. The means and standard deviations for these interactions are shown in Tables 5.13 and 5.14 respectively.

As can be seen from the means in Table 5.13, several trends emerge that are consistent with the previous results. Accuracy of object recognition increased with (1) decreasing number of items and groups and (2) increasing age. However, as highlighted in the Table, the 10-year-olds more accurately recognised 2 or 3 single items and 2 groups of 5 and 15 items than the 4-year-olds, 8-year-olds and adults. Also, items presented in 3 groups of 5 items were better recognised by 10-year-olds and adults than 3 single items.

Table 5.13 Means and standard deviations of correct recognition responses at each age for the groups and items conditions

Age	2 groups						3 groups					
	1 item		5 items		15 items		1 item		5 items		15 items	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
4's	.427	.339	.448	.349	.440	.324	.462	.324	.400	.300	.389	.349
8's	.489	.344	.453	.330	.373	.301	.500	.307	.464	.319	.458	.318
10's	.635	.286	.526	.301	.428	.308	.574	.287	.627	.305	.583	.281
Adults	.517	.330	.495	.300	.376	.296	.491	.297	.680	.267	.653	.282

Similar trends can be seen from the means in Table 5.14. Overall, accuracy of object recognition increased with age and for items presented in 3 groups over 2 groups. Items presented with an unmarked novel noun were better recognised than those presented with singular and plural novel nouns. However, as can be seen in the highlighted cells, the 4-year-olds recognised items presented in 2 groups with a singular and plural noun more accurately than those presented with an unmarked noun. And, 2 groups of items presented with the singular and plural novel nouns were more accurately recognised by 10-year-olds than by adults.

Table 5.14 Means and standard deviations of correct recognition responses at each age for the groups and syntax conditions

Age	2 groups						3 groups					
	singular		unmarked		plural		singular		unmarked		plural	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
4's	.463	.348	.419	.374	.433	.345	.419	.311	.413	.327	.419	.335
8's	.417	.331	.489	.364	.409	.299	.481	.324	.503	.332	.439	.289
10's	.478	.293	.550	.320	.562	.282	.627	.278	.636	.314	.522	.281
Adults	.369	.320	.566	.314	.453	.292	.633	.266	.624	.294	.566	.286

Summary

The main findings of the recognition responses show that overall accuracy is (1) better for the two older age groups than the two younger age groups, (2) better for items presented as single and groups of 5 items than as groups of 15 items, and (3) better for items presented in 3 groups than 2 groups. The increase in accuracy with age is observed for items presented as single items, 5 items, 15 items, and as 3 groups. Items presented in 2 groups were better recognised if they were presented with an unmarked novel noun over the singular and plural nouns, but for 3 groups, items presented with a plural noun were better recognised than those with an unmarked and singular noun. The effect of group was not observed for single items. That is, items presented in 3 groups were better recognised than 2 groups only when 5 or 15 items were in a group. The language x syntax interaction suggests that English speakers more accurately recognised items presented with unmarked and plural novel nouns than the Welsh speakers were.

Reaction Times

As with reaction time data for the collective categorisation task, some 4-year-olds did not type in their responses. To eliminate contamination from the experimenter's reaction times, only the data for the 8-year-olds, 10-year-olds and adults were analyzed for the previously seen items. Four of the 2430 data points were out-liers (above two standard deviations away from the mean). These were suspected to be results of a technical error with the running of the Psyscope program and were disregarded from the analysis.

A language x age x syntax x items x groups ANOVA revealed no significant main effects or interactions for recognition reaction times.

Summary

Language, age or presentation status (previously seen as 2 or 3 groups of 1, 5 or 15 items with a singular, unmarked or plural novel noun) had no effect on participants' reaction times for recognising individual items.

Discussion

The language x syntax interaction for object recognition suggests that language structure does play some role in non-linguistic cognition. Welsh speakers seemed to display poorer recognition of objects presented with unmarked and plural novel nouns than English speakers. The language differences imply that the ability to recognise individual objects presented with unmarked and plural novel nouns was impeded for Welsh speakers when compared to English speakers because of the way speakers of the two languages represented those objects on first exposure. Since the singular novel nouns clearly indicate reference to individuals, this would allow representations of the individuals to be consolidated in memory. However, the reference of the unmarked and plural nouns may be more ambiguous for Welsh speakers in particular, resulting in poorer consolidation of the individual items in memory. Moreover, Welsh speakers took significantly longer to categorise than the English speakers in the collective categorisation task. This suggests that Welsh speakers, regardless of overt categorisation responses, were having more difficulty in deciding how to categorise (or represent) the objects and were less able to formulate robust representations of the items. These findings then support the hypothesis that memory for individual items may be facilitated or hindered by the way these items are represented. When syntax provides a clear reference to individuals, then stronger representations can be made of those individuals thus aiding recognition. When the reference of nouns is more ambiguous though, representations of the individuals are weaker, thereby impeding recognition.

Age was also a key factor in the ability to accurately recognise novel objects. Overall, the two older age groups were better at recognising individual objects than the two younger age groups. This was observed regardless of number of items presented in a group. This suggests that with only one previous exposure to an item the ability to consolidate representations of the individual in memory improves with age.

Further support comes from the effect of object properties on recognition. Items presented as singles or groups of 5 were more accurately recognised than items presented in groups of 15 items. Again, recall from the collective categorisation task, participants gave more collective responses with increasing number of items and groups, and took significantly longer to decide with increasing number of items and groups available. This implies that when participants are unsure about how to interpret novel nouns they are less likely to formulate strong representations of individual items for recognition. In addition, given that more collective responses were given when more items were available, participants may be representing the items as collections, not individuals, thereby interfering with the ability to recognise these items.

In contrast, items presented in 3 groups were better recognised than items in 2 groups but only when 5 or 15 items were in a group, not when 2 or 3 single items were presented. Moreover, items presented in 2 groups were better recognised if they were presented with an unmarked novel noun over the singular and plural nouns. But, for 3 groups, items presented with a plural noun were better recognised than those with an unmarked and singular noun. One possible explanation for this is that a plural context allows attention to be focused on the individual items if more groups are available. With fewer groups available, a more ambiguous context (i.e. unmarked nouns) allows attention to be focused on the individuals. If this is the case then paying more attention to the individual allows the individual items to be consolidated in memory, thereby aiding recognition for those items. Nevertheless, these findings support the hypothesis that memory for individual items may be facilitated or hindered by the way these items are represented as a consequence of the context in which they were first presented.

The lack of significant differences in reaction times for the recognition task allows important conclusions to be drawn from the previous reaction time data in the collective categorisation task. First, the results for the reaction times in

the collective categorisation task must be due to the categorisation process and not due to differences in motor skills. Second, the observed language effect in the reaction times of the categorisation task are not due to English speakers being generally faster than Welsh speakers. Third, the age effect also observed in the reaction times data for the collective categorisation task again are not due to speed of motor responses but rather, are to do with the categorisation process. Language processing is a key factor involved in the categorisation task that is not involved in the recognition task. It is likely then that the differences in the reaction times for the categorisation task stem mainly from the language demands of the task.

Follow-up study

Although the overt categorisation of novel objects by Welsh speakers did not differ significantly to the English speakers that is not to say that differences in the grammatical structure of each language have no effect. Indeed, the reaction time data observed for the categorisation task show that Welsh-speaking participants took significantly longer to categorise novel objects than their English-speaking counterparts did. Since there were no differences in reaction times between the two language groups on the recognition task, it can be argued that Welsh-speakers are slower due to the language demands of the task rather than extraneous reasons. Therefore, as previously mentioned, there are two possible explanations for these results.

Firstly, when having to decide what novel words mean and therefore categorise novel objects, Welsh speakers are faced with more options and possible referents of the nouns than the English speakers. This is because the structure of Welsh allows unmarked noun forms to refer to single items, substances and collections. It is feasible that the Welsh speakers took significantly longer in the categorisation task because their language does not constrain the possible referents in the same way as English.

Alternatively, it could be argued that Welsh speakers took significantly longer to categorise because they are bilingual. Essentially, doing a language oriented task would require a bilingual speaker to inhibit one language whilst undertaking the task in the other. This follow-up study attempts to address these two possible explanations by conducting the same experiment with Welsh/English bilinguals through the medium of English. It is hypothesised that if Welsh speakers take longer because of the structure of their language, then it would be expected that bilingual participants' reaction times will not differ to those of the English speakers when the task is undertaken in English. However, if Welsh speakers take longer as a consequence of being bilingual and the interference of

another language, then it would be expected that even when the task is carried out in English, the reaction time data should be similar to the Welsh group.

Method

Participants

Thirteen bilingual adults took part in the follow-up study. There were 12 females and 1 male. The mean age was 19;6 and the range was 18;2 – 25;9. All the participants were recruited from the University subject panel and completed a language background questionnaire. Only participants that rated themselves as bilingual and stated that the home language was between 40-60% Welsh took part. All participants undertook all the tasks as detailed in Study 3.

Apparatus

The same equipment and computer programme was used as detailed in Study 3.

Procedure

The procedure was carried out as detailed in Study 3 for the English-speaking participants.

Results

Only the reaction time data were used for analysis in this follow-up study. The reaction times of the bilingual participants were compared to the Welsh and English adult data from the collective categorisation task in Study 3. A language x items x groups x syntax ANOVA was carried out on the reaction time data from the categorisation task. Significant main effects were found for language, F 2,43 =3.60,p<.036, Items F 1,43 = 23.89,P<.000, and Groups F 1,42 = 12.67,p<.001. The main effects of items and groups are consistent with the results found for the Welsh and English participant in the original study and will not be discussed further here. See main results for reaction times of the collective categorisation task for discussion of these effects.

Post hoc (LSD) analysis of language indicates that English speakers were significantly faster than Welsh speakers (MD = 989.10,p<.010). However, bilingual participants were not significantly faster than the Welsh speakers (MD = 490.74,p<.221) or significantly slower than English speakers (MD = 498.36,p<.208). This language effect is depicted in Figure 5.12

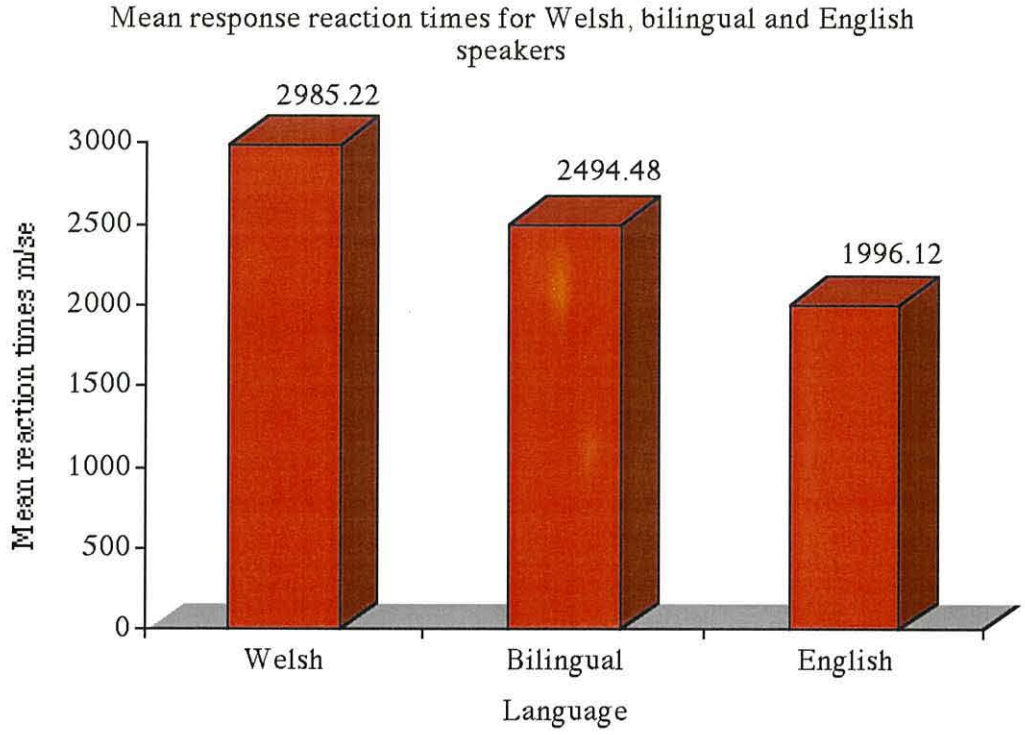


Figure 5.12 Mean reaction times for Welsh, English and Bilingual participants

Summary

The results of the reaction time data for bilingual participants undertaking the collective categorisation task in English neither support or contradict the hypothesis that decreased reaction times were a consequence of differences in language structure. These findings suggest that both differences in language structure and an effect of bilingualism may be involved in the reaction times of the Welsh speakers in the categorisation task.

General Discussion

This study aimed to address three main questions. First, does language structure guide categorisation for language learning? Second, do object properties and syntax play an important role for categorising collections as whole individuals? And third, do the ways collections of objects are categorised (and represented) influence memory for the individual items that make up the collection?

The collective categorisation task was designed to investigate the first question. It was hypothesised that Welsh speakers would be more likely to interpret novel nouns as referring to collections than English speakers because of differences in the way these two languages label collection categories. It was expected then that Welsh speakers would categorise novel objects as collections more often than English speakers. The categorisation responses do not offer strong support for this hypothesis and indeed indicate little difference in the way speakers of the two languages categorised the novel objects. It seems that both Welsh- and English-speaking participants interpreted novel nouns in similar ways.

However, a trend emerges suggesting that Welsh-speaking 10-year-olds categorised novel objects as collections more than English-speaking 10-year-olds, especially when more items are available for categorisation. The reverse is true for adults. These results imply that the effects of differences in language structure emerge with age and may reflect a greater understanding of language systems as children become more experienced with their language. Note that the trend of language difference mentioned above are not due to differences in basic arithmetic abilities, as shown in the arithmetic task. This does lend some support to the notion that language structure guides categorisation for language learning.

Moreover, a robust and consistent finding that Welsh speakers took significantly longer to categorise than the English speakers suggests that language structure does influence on-line language processing. When interpreting novel

nouns, more choices of possible interpretations are available to Welsh speakers than to English speakers. Although categorisation responses were similar across the two languages, the process of deciding on the meanings of new words took more time for Welsh speakers.

Note that Welsh speakers, for the most part, are proficient English speakers. This may account for the fact that Welsh and English speakers categorised novel objects in a similar way. Welsh speakers' knowledge of English may interfere with their categorisation behaviour.

To rule out the possibility that Welsh speakers took longer to categorise as a result of being bilingual, a follow-up study examined the reaction times of Welsh/English bilinguals undertaking the collective categorisation task in English. The results neither support nor contradict the hypothesis that decreased reaction times were a consequence of differences in language structure. These findings suggest that both differences in language structure and an effect of bilingualism may be involved in the reaction times of the Welsh speakers in this task.

The absence of robust differences between the two languages on collective categorisation responses does not mean that language has no role in categorisation for language learning. The reaction time data indicate that Welsh speakers are influenced by the options that their language gives them when interpreting novel words and categorising novel objects.

Age also seems to have an important role in categorisation behaviours. A general trend is seen across the two language groups that participants were more likely to categorise novel objects as collections with increasing age. The youngest participants were less willing to interpret novel nouns as referring to collections than the older participants were. This again may reflect the increasing knowledge of language gained with age.

This is also shown by the reaction time data. It would be expected that if the structure of the language being learned guides the way that speakers

categorise objects and learn new words, then the more experience people have with the structure of their language the quicker they will be at making decisions about novel nouns. The more participants know about their language, the quicker they will be at learning new words. With increasing age, participants were faster at categorising objects and deciding how to interpret novel nouns. Again this may be due to greater experience with language gained with age.

The collective categorisation task also addressed the second question; do object properties and syntax play an important role for categorising collections as whole individuals? The categorisation responses and reaction time data indicate that both object properties and syntactic cues are important factors in interpreting novel nouns as collections.

First, a general pattern indicates that novel objects are categorised as collections more often with increasing numbers of items and with increasing numbers of groups for both languages. Increasing numbers of items and groups may act as cues for participants to interpret novel nouns as collections. These cues, particularly alongside additional syntactic cues, act to facilitate categorisation of objects as collections, by children and adults. The speed with which participants categorise novel objects also seems to be influenced by object properties. Participants took longer to categorise novel objects when more objects were available. Given that participants are more willing to consider a collective interpretation of the novel nouns - but took longer to categorise - when more items and groups are available, it would seem that increasing the numbers offers greater choices about how to think about the objects and the meanings of novel nouns. These results are consistent with the hypothesis that increasing the numbers of items and groups available for categorisation facilitates collective categorisation.

Second, syntax was also shown to be an important cue in learning the meanings of new words. Unmarked nouns are ambiguous as to the referent, especially in Welsh, while singular nouns make a clear reference to individual

items. Although plural nouns can refer to several groups, it was expected that plural nouns would be interpreted as referring to many individuals. Participants were more willing to adopt a collective interpretation for an unmarked noun. Fewer singular nouns were interpreted as referring to collections. For the singular and plural conditions, additional cues such as increased number of items appeared to be used to aid the categorisation process. The reaction time data also support the idea that unmarked syntax offers more options about possible referents when interpreting new words. Participants took longer to categorise items in the unmarked condition than the singular and plural conditions, especially when more items were available. This suggests that participants are more willing to accept a collective interpretation of a noun when there is no clear indication from the noun as to its referent. This added option allowed participants to think about the objects as collections and slowed down the speed of categorisation. These data are consistent with the hypothesis that unmarked syntax would elicit greater collective responses than singular and plural syntax.

The recognition task was designed to address the third question, do the ways collections of objects are categorised (and represented) influence memory for the individual items that make up the collection? It was hypothesised that the way collections of objects were categorised or represented would influence memory for the individual items. The data are consistent with this hypothesis and indicate that participants' recognition of items was influenced by the context in which they were first exposed to them. The results suggest that language structure does play some role in non-linguistic cognition. Welsh speakers displayed poorer recognition for objects presented with unmarked and plural novel nouns than English speakers. This language effect suggest that the ability to recognise individual objects presented with unmarked and plural novel nouns was impeded for Welsh speakers when compared to English speakers because of the way speakers of the two languages represented those objects on first exposure. Since the singular novel nouns clearly indicated reference to individuals, this would

allow representations of the individuals to be consolidated in memory. However, the reference of the unmarked and plural nouns may be more ambiguous, for Welsh speakers in particular, resulting in poorer consolidation of the individual items in memory. Moreover, Welsh speakers took significantly longer to categorise than the English speakers in the collective categorisation task. This suggests that Welsh speakers, regardless of overt categorisation responses, were having more difficulty in deciding how to represent the objects and were less able to formulate robust representations of the items. These findings then support the hypothesis that memory for individual items may be facilitated or hindered by the way these items are represented. When syntax provides a clear reference to individuals, then stronger representations can be made of those individuals, thus aiding recognition. When the reference of a noun is more ambiguous, with the possibility of reference to a collection, then representations of the individuals may be weaker, thereby impeding recognition.

Age was also a key factor in the ability to accurately recognise novel objects. Overall, the two older age groups were better at recognising individual objects than the two younger age groups across all conditions. This consistent and robust finding suggests that the ability to consolidate representations of individual items in memory improves with age. It may also be that the ability of younger participants to formulate robust representations of the individuals for recognition was impeded because they were unsure how to categorise the objects. Again these findings are consistent with the hypothesis that memory for individual items may be facilitated or hindered by the way these items are represented on initial exposure. Having more choices about how to interpret novel nouns may have important consequences for the way objects are represented and in turn recognised.

In contrast, differing effects on recognition are observed with different object properties. Items presented as singles or groups of 5 were more accurately recognised than items presented in groups of 15 items. Again, recall the collective

categorisation task. Participants gave more collective responses and took significantly longer to categorise with increasing number of items and groups. The effect of items supports the idea that more options for interpreting novel nouns interferes with the formation of a strong representation of the individual items for recognition.

Conversely, the effect of groups (only for groups of 5 or 15 items) indicates that *more* groups allow stronger representations of the individual items. This result does not support the idea that stronger representations of individuals are formed when more options for interpreting novel nouns are available. It is difficult to interpret this result but it may be that 2 groups have distinct properties that differ from any other number of groups. When 2 groups of items are presented, attention may be distributed across the two groups when deciding on how to categorise. This would hinder the ability to formulate strong representations. But when 3 or more groups are presented, attention may focus on one group because of the cognitive demands of processing many individuals. Consequently, individual items in 3 groups may have stronger representation that aid recognition. Nevertheless, object properties and syntax have an important role in the way people think about novel objects and have important consequences for object recognition. The way novel objects are recognised is influenced by the way those objects were initially presented. This may be due to the way the objects are attended to, represented and ultimately stored in memory.

The lack of significant differences in the reaction times for the recognition task allows important conclusions to be drawn. Primarily, the results for the reaction times in the collective categorisation task must be due to the categorisation process and not due to differences in motor skills. These include the language and age effects found in the reaction time data of the categorisation task. These results are not due to differences in speed of responding but instead, are likely to stem from the categorisation process. The collective categorisation task involves language processing while the recognition task does not. It appears

that the differences in the reaction times for the categorisation task are a consequence of the language demands of that task.

Many factors seem to play a role in the categorisation of novel objects. Often, multiple cues are required when acquiring the meanings of new words. This, in turn, impacts on memory and recognition for those items. First, the structure of the native language has important influences on the way children acquire the meanings of new words, as does age. And second, the context of novel nouns also has important influences on the way the nouns are interpreted. The syntax of the noun as well as object properties such as number of items or groups seem to play an important role in facilitating collective interpretation of novel nouns.

Although it seems that language differences do not directly affect overt categorisation, the structure of the language being learned may contribute, as one of the many factors, to the overall categorisation process for language learning. Furthermore, these cues involved in categorisation have important effects on the way items are represented and ultimately recognised. Language differences then seem to have some role in guiding language learning but also impact on non-linguistic cognition.

To further examine the role of language structure on linguistic and non-linguistic cognition the following study uses both verbal descriptions and picture selection procedures. The aim is to investigate the role of different noun type categories on attention and memory. The verbal descriptions addresses attention to different noun types in both Welsh and English, measured by mention of target items. The picture selection procedures address attention to number changes in short- and long-term memory without verbal behaviours.

Study 4: Cognition of pictorial scenes – the role of language in guiding attention

Lucy (1992) conducted a study comparing Yucatec Mayan and American English speakers' attention to, and sensitivity to changes in, the *number* of items depicted in line drawings. He observed that Yucatec and English differed mainly with regard to pluralising nouns that refer to implements. In English, it is obligatory to pluralise these nouns. However, this is not the case in Yucatec. For Yucatec speakers, plural forms are possible but rarely used for implements.

In his study, Lucy (1992) found that speakers of the two languages attend to *number* of items in ways that were consistent with these differences across the two languages. Specifically, he found that attention to, and sensitivity to changes in *number* for animate objects and substances were similar across the two languages. Speakers of the both Yucatec and English attended to *number* for animate objects but insensitive to changes in *number* for substance. For implements however, speakers of the two languages performed differently. English speakers attend more to, and were more sensitive to changes in the *number* of implements than their Yucatec Mayan counterparts.

According to Lucy (1992) these findings are consistent with the differences across the two languages and as such, are a direct result of the grammatical structure of the two languages. Less attention is paid to *number* if the nouns that label them do not require grammatical modification (in this case, pluralisation). This study aimed to investigate the role of grammatical structure of Welsh and English in guiding speaker's attention to *number*.

Welsh differs from English with regard to collection categories. Both English and Welsh have nouns that require pluralisation and nouns that do not. In general, countable entities are labelled with singular/plural nouns while substances are labelled with singular only nouns in both languages. Unlike English, Welsh has a

set of nouns that denote collections that are not pluralised. These entities are labelled with cluster/unit nouns in Welsh but with singular/plural nouns in English.

Using a similar methodology to Lucy's (1992), this study addressed Welsh- and English-speaking participants' attention to, and sensitivity to changes in *number*. It was expected that Welsh- and English speaking children and adults would attend to and would be sensitive to changes in *number* for entities labelled with singular/plural nouns and for substances in similar ways. However, for entities labelled with cluster/unit nouns in Welsh (singular/plural nouns in English), it was hypothesised that English speakers would attend more and be more sensitive to changes in *number* than their Welsh-speaking counterparts for two reasons. First, cluster/unit nouns share some syntactic and semantic properties with nouns denoting substance. Cluster nouns allow the referents to be thought of as substance-like. Second, talking about entities as collections reduces the need to attend to *number*.

This study explored Welsh and English speakers' attention to and sensitivity to changes in *number* by (1) examining picture descriptions and (2) using picture selection procedures. Verbal descriptions of pictures, with and without the pictures in view, were examined to indicate participants' attention to target items depicted in the scenes. Picture similarity judgements were aimed to explore the importance of *number* for judging a picture to be most similar. Recall of *number* changes across short- and longer-term memory also indicate the importance of *number* and the ease with which *number* changes are noticed. The role of *number* in picture identification (i.e. does memory for *number* degrade over time?) was also explored by comparing short- and long-term recall for *number*.

In addition, this study aimed to explore a foreground and background effect that was not introduced in Lucy's (1992) study. One possible explanation for Lucy's (1992) findings may be that substances and implements tended to be background material in the picture scenes while animate objects were often the foreground items. This study attempted to explore the possibility that participants'

attention to and recall of items will be influenced by whether the item is presented in the foreground or background of the scene.

Method

Design

The study involves 5 tasks with the aim of exploring attention to specific items depicted in the picture stimuli. All participants undertook all five tasks in a set order. This was to ensure that all conditions of the study were equated across all the participants and to replicate Lucy's methodology. The five tasks differed in degree of difficulty. To allow participants to become accustomed to the procedure, and to minimise confounding effects of prior task experience on responses, the presentation order of tasks was from easiest to most difficult as follows.

Task 1: Picture description

This task involved participants describing a scene depicted in cartoon-like pictures. Participants were given an unlimited amount of time to examine the picture and to describe it with the picture in view.

Task 2: Picture description with no picture in view

For this task, participants were shown a picture for 30 seconds. Following a 30-second delay period, they were asked to describe the picture while the picture was **not** in view. Participants were given an unlimited amount of time to recall and describe the picture.

Task 3: Judging the variant picture most like the original

Participants were shown a picture and 3 corresponding variant pictures. Participants were asked to determine the differences between each variant and the original. Once all differences were established, participants were asked to select the variant picture most like the original. The original and all variant pictures were shown simultaneously.

Task 4: Finding the original (short-interval)

In this task, participants were shown a picture for a period of 30 seconds. Following a 30 seconds delay period, participants were shown the original and 3 corresponding variant pictures individually in a random order. Participants were asked to judge whether the picture was the original.

Task 5: Finding the original (long-interval)

As in the previous task, participants were shown a picture for a period of 30 seconds. A 30-minute interval was then introduced. Again, participants were shown the original and 3 corresponding variant pictures individually in a random order. Participants were then asked to judge whether the picture was the original.

The picture description tasks were expected to give a measure of the ways Welsh and English speakers attended to the *number* of items in a scene and the noun phrases used by the participants. The picture selection tasks aimed to indicate participants' attention to *number* for target items depicted in the scene because the variant pictures differed from the original with regard to *number* only. Short-term and long-term recall of the *number* of items was examined in the two tasks where the participants were asked to select the original. Table 6.1 summarises the five tasks in this study.

Since the objective of the study was to compare participants' responses, given the particular demands of the task, it was necessary that all participants were given the same conditions in each task. This is especially pertinent when comparing picture descriptions and judgement responses. Therefore, the same picture sets were used in each task for all participants but were presented in a random order.

Table 6.1 Summary of the tasks

Task	Task description	Exposure time	Delay period
1	Picture description with picture in view	unlimited	
2	Picture description without the picture in view	30 seconds	30 seconds
3	Judging a variant picture most like the original	unlimited	
4	Finding the original (short-term memory)	30 seconds	30 seconds
5	Finding the original (long-term memory)	30 seconds	30 minutes

Apparatus

A video recorder was used to record participants undertaking the tasks and a stopwatch was used to measure presentations times of the stimuli. Fifteen sets of four picture stimuli were created; three sets for each task (picture stimuli can be seen in Appendix V). All pictures were of equal size; they were colour printed on A4 size paper and laminated. A scoring sheet (Appendix VI) was devised to score picture selection and to make notes during the picture description tasks.

Stimuli

The fifteen picture scenes were created using Microsoft paint package. Each scene consisted of 3 different types of target items. The target items were referents of nouns from three noun type categories. These noun categories were (1) singular/plural nouns, (2) cluster/unit noun in Welsh (singular/plural nouns in English), and (3) nouns denoting substances. For example, the beach scene (Picture 1, Appendix V) contained the target items SHELLS (singular/plural category in both languages); BIRDS (cluster/unit noun *adar* in Welsh; singular/plural noun 'birds' in English); and SAND (substance category in both languages).

For each of the five tasks there were three picture scenes. These were designed so that in one picture the sg/pl category was in the foreground (i.e. was what

the picture was 'about'), in one the cluster/unit category was in the foreground and in one the substance category was in the foreground. For example, in the beach the sg/pl category, the SHELLS, was in the foreground, and the remaining categories BIRDS and SAND, were in the background.

The *number* of target items for all three categories in a given picture was the same. For example, in the beach scene, there were 3 shells, 3 birds and 3 heaps of sand. However, *number* varied across the three pictures within each task; Each task contained three pictures, one with 3 of each target item type, one with 4 of each target item type, and one with 5 of each target item type.

For each original picture in task 3, 4 and 5, three variant pictures were created. Each variant differed from the original in the *number* of items for one of the category types. One variant picture had one fewer instance of the sg/pl category type (e.g. 2 shells compared to 3 in the original); one had fewer of the cluster/unit category type (e.g. 2 birds compared to 3 in the original); and one had fewer of the substance category type (e.g. 2 heaps of sand compared to 3 in the original). A Table listing the picture scenes, their foreground category type, the target items in each scene, the number of target items in each scene and their corresponding variant pictures can be seen in Table 6.2. The picture stimuli used for each task can be seen in Appendix V.

Table 6.2 Summary of the picture stimuli

Task	Picture name	Foreground category	Target items	No. of items	sg/pl variant	c/u variant	substance variant
1	Beach	Sg/pl	Shells (sg/pl) Birds (c/u) Sand (substance)	3	-1 shell	-1 bird	-1 heap of sand
	Bees	C/U	Clouds (sg/pl) Bees (c/u) Wood (substance)	4	-1 cloud	-1 bee	-1 piece of wood
	Kitchen	Substance	Cups (sg/pl) Mice (c/u) Flour (substance)	5	-1 cup	-1 mouse	-1 pile of flour
2	Park	Sg/pl	Balls (sg/pl) Trees (c/u) Grass (substance)	4	-1 ball	-1 tree	-1 tuft of grass
	Farm	C/U	Hens (sg/pl) Pigs (c/u) Paint (substance)	3	-1 hen	-1 pig	-1 pot of paint
	Party	Substance	Cakes (sg/pl) Strawberries (c/u) Juice (substance)	5	-1 cake	-1 strawberry	-1 glass of juice
3	Pond	Sg/pl	Frogs (sg/pl) Ducks (c/u) Wool (substance)	5	-1 frog	-1 duck	-1 piece of wool
	Snow	C/U	Dogs (sg/pl) Children (c/u) Snow (substance)	4	-1 dog	-1 child	-1 mound of snow
	Drawing	Substance	Books (sg/pl) Pears (c/u) Paper (substance)	3	-1 book	-1 pear	-1 piece of paper
4	Night	Sg/pl	Cars (sg/pl) Stars (c/u) Water (substance)	3	-1 car	-1 star	-1 pool of water
	Train	C/U	Houses (sg/pl) Leaves (c/u) Smoke (substance)	5	-1 house	-1 leaf	-1 cloud of smoke
	Ostrich	Substance	Eggs (sg/pl) Feathers (c/u) Oil (susbtance)	4	-1 egg	-1 feather	-1 pool of oil
5	Garden	Sg/pl	Butterflies (sg/pl) Carrots (c/u) Soil (substance)	4	-1 butterfly	-1 carrot	-1 bucket of soil
	Bakery	C/U	Cakes (sg/pl) Roses (c/u) Bread (substance)	5	-1 cake	-1 rose	-1 loaf of bread
	Ants	Substance	Apples (sg/pl) Ants (c/u) Coal (substance)	3	-1 apple	-1 ant	-1 piece of coal

Participants

One hundred and forty four Welsh- and English- speaking children and adults participated in this study. They were recruited from the local community, the University of Wales, Bangor, subject panel, four English medium schools, and nine Welsh medium schools.

Participants fell into one of the two language groups, the Welsh group or the English group. As noted in the previous studies, recruiting Welsh monolinguals is problematic. Consequently, strict inclusion criteria were set to maximise the probability of tapping into a strong Welsh mindset. Assignment to language group was based on self-reports for adults and parental reports for children on language background via questionnaires (Appendices III & IV). Adult participants were allocated to the Welsh language group if they reported that their first language was Welsh and that the primary language spoken at home throughout childhood was Welsh (80 - 100% Welsh at home). Child participants were allocated to the Welsh language group if parental reports indicated that the child's first language was Welsh, the primary language spoken at home was Welsh (80-100% Welsh at home), and the child attended a Welsh medium school. Similarly, adult participants were allocated to the English language group if their first language was English and the primary language spoken at home throughout childhood was English (80 -100% English at home). Child participants were allocated to the English language group if parental reports indicated that the child's first language was English, the primary language spoken at home was English (80-100% English at home), and the child attended an English medium school.

The participants also fell into 4 age groups: Adults (mean age: 28;1, range:18;7 – 72;5), 10-year-olds (mean age:10;11, range:10;5 – 11;4), 8-year-olds (mean age:7;11, range: 7;4 – 8;2), and 4-year-olds (mean age: 4;11, range: 4;5 – 5;5).

The adult group consisted of 32 participants: 15 Welsh speakers (mean age: 32;2, range: 19;7 – 72;5, 12 females, 3 males); and 17 English speakers (mean age: 24;0, range: 18;7 – 45;9, 11 females and 6 males).

The group of 10-year-olds consisted of 38 participants: 21 Welsh speakers (mean age: 10;11, range: 10;5 – 11;4, 11 females and 10 males); and 17 English speakers (mean age 10;11, range: 10;5 – 11;2, 4 females and 13 males).

The group of 8-year-olds consisted of 34 participants: 19 Welsh speakers (mean age: 7;11, range: 7;4 – 8;1, 11 females and 8 males); and 15 English speakers (mean age: 8;0, range 7;4 – 8;2, 9 females and 6 males).

The group of 4-year-olds consisted of 45 participants: 21 Welsh speakers (mean age: 5;1, range: 4;5 – 5;5, 9 females and 12 males); and 24 English speakers (mean age 4;10, range 4;5 – 5;1, 12 females and 12 males).

All participants undertook all five tasks. Note that due to attrition during task 3, the data for two Welsh-speaking 4-year-olds and one Welsh-speaking 8-year-old were eliminated from the analysis of that task.

Procedure

All participants undertaking the tasks were video recorded. Initially, participants were given general instructions as follows:

English

"I have some pictures that I'd like to show you and I will be asking different questions about them"

Welsh

"Mae genna' i ychydig o luniau hoffwn ddangos i chi ac mi fydda i'n gofyn gwestiynnau gwahanol amdanynt."

Task 5 was a long-term memory task and required a 30-minute interval period between exposure to the picture and the test phase. For this reason, all participants were given the pictures for this task at the beginning of the session. Participants were given each picture in this task for 30 seconds and were given the following instructions:

English

" Now I'm going to show you some pictures for a few seconds so you can see the kinds of pictures we will be looking at. All you need to do for the moment is look at the picture. You don't have to say anything yet"

Welsh

"Rwan, 'dw i am ddangos i chi ychydig o luniau i chi gael gweld y math o luniau byddwn ni'n edrych ar. Yr hyn oll ydych 'i angen ei wneud ar hyn o bryd ydi edrych ar y llun. Does 'na'm angen i chi ddweud dim byd am y tro"

Task 1: Description with picture in view

For this task, participants were presented with an original picture for an unlimited time. Participants were asked to describe the scene while the picture was in their view. They were given the following instructions:

English

"Take a few seconds to look at this picture, when you're ready can you describe the picture to me?"

Welsh

"Cymmerwch ychydig o eiliadau i edrych ar y llun yma, pan rydych yn barod allwch ddisgrifio'r llun i mi?"

Participants' responses were recorded on the scoring sheet to encourage child participants to verbalise about the scene.

Task 2: Description without the picture in view

In this task, participants were presented with an original picture to view for a period of 30 seconds timed with a stopwatch. The instructions were as follows:

English

"I'm going to show you a picture for a few seconds"

Welsh

"'Dw i am ddangos llun i chi am ychydig o eiliadiau"

Following a 30-second delay period (again timed with a stopwatch), participants were asked to describe the scene in the absence of the picture. They were given the following instructions:

English

"Can you to describe the picture you have just seen?"

Welsh

"Allwch chi ddisgrio'r llun rydych newydd ei weld"

Again responses were recorded on the scoring sheet.

Task 3: Judging the variant picture most like the original

Participants were shown an original picture to examine while the experimenter shuffled the 3 variant pictures. The original picture was then placed in the bottom left hand side of the table nearest the participant. The 3 variants were then placed bottom right hand side, top left hand side and top right hand side in a random order for all participants. This ensured that all participants had the original in the same location for comparison with the variant pictures that were randomly placed in the remaining positions.

Participants were given the following instructions:

English

"I'm going to show you some more pictures that are very similar to the one you have there. Can you tell me how each picture is different from that first one?"

Welsh

"'Dw i am ddangos lluniau i chi sy'n debyg iawn i'r un sydd gennych ar hyn o bryd. Allwch ddweud wrtha' i sut mae bob un o'r lluniau yma yn wahanol i'r un cyntaf 'na?"

This ensured that all participants were aware of the specific differences in *number* of items in the pictures. Once all the differences between each variant and the original were established, participants were asked to select the picture that was most similar to the original with the following instructions:

English

" Which picture is the one most similar to that first one?"

Welsh

"Pa lun sy'n debycach i'r un cyntaf 'na?"

Participants' choice of picture was recorded.

Task 4: Finding the original (short-interval)

In this task, participants were shown an original picture for a period of 30-second (timed with a stopwatch) and given the following instruction:

English

"I'm going to show you a picture for a few seconds. All you need to do for the moment is look at the picture. You don't have to say anything yet"

Welsh

"'Dw i am ddangos llun i chi. Yr oll ydych 'i angen ei wneud ar hyn o bryd ydi edrych ar y llun. Does 'na'm angen i chi ddweud dim byd am y tro"

The original picture was then placed with the 3 variant pictures and shuffled. Following a 30 seconds delay period (again timed with a stopwatch) each picture was shown individually in a random order. Participants were asked to judge whether the picture was the original one they had seen. The instructions were as follows:

English

"I'm going to show you that picture again but it will be mixed in with pictures that are very similar. I will show you each picture one-by-one and I would like you to tell me if you think it was the one you saw first"

Welsh

"'Dw i am ddangos y llun yna i chi eto ond y tro yma, mi fydd y llun wedi ei gymysgu efo rhai eraill sy'n debyg iawn. Mi 'na i ddangos y lluniau bob yn un a hoffwn i chi ddweud wrtha' i pa un ydych 'i'n meddwl ydi'r un welsoch 'i yn gyntaf"

Participants' selection/s were recorded on the coding sheet.

Task 5: Finding the original (long-interval)

Participants were presented with the original pictures for this task at the start of the session. Following a 30-minute interval, participants were asked to select the original picture from the array with the following additional instructions:

English

"At the very beginning I showed you some pictures and asked you just to look at them. Now I'm going to show you each picture again but it will be mixed in with pictures that are very similar. I will show you each picture one-by-one and I would like you to tell me if you think it was the one you saw first"

Welsh

"I gychwyn efo, nes i ddangos lluniau i chi a gofyn i chi edrych ar y llun yn unig. Rwan 'dw i am ddangos y llun yna i chi eto ond y tro yma, mi fydd y llun wedi ei gymysgu efo rhai eraill sy'n debyg iawn. Mi 'na i ddangos y lluniau bob yn un a haffwn i chi ddweudd wrtha i pa un ydych 'i'n meddwl ydi'r un welsoch 'i yn gyntaf"

Participants' selection/s were recorded on the coding sheet.

The experimenter engaged the participants in a everyday conversation during the short intervals in order to provide distraction from the task. For the long-interval task, participants undertook the remaining tasks as distraction during the 30 minutes interval period of task 5. On completion of all five tasks, the participants were debriefed about the study.

Results

Two sets of data were obtained from this study, description responses and picture selection responses.

Description responses

The videotapes of participants undertaking tasks 1 and 2 were transcribed. Only the noun phrases referring to the target items were analysed. The nature of naturalistic responses allows participants to use any number of terms to describe the target items. Indeed, considerable variability was found to be the case in the description responses for both tasks 1 and 2. Often, superordinate or subordinate terms were used to refer to the target items (e.g. target noun 'strawberries' would often result in 'fruit', whilst 'birds' would often elicit 'seagulls' as a response). For this reason it was necessary to establish a set of criteria that would allow some alternate nouns to be included in the analysis. This was to ensure that participants attending to the target items were not overlooked due to using a different term, and that those using very different terms were not included when they were not attending to the target item[1]. Alternate nouns were included if

(1) the term was the same noun number-type category as the target noun (e.g. 'mugs' instead of 'cups' but not 'drinks' instead of 'juice')

(2) the term was the same category level as the target noun (e.g. 'teisennau' (cakes) instead of 'cacennau' (cakes) but not 'fruit' instead of 'strawberries')

[1] Occasionally Welsh children would use an English word such as 'sand' or 'strawberries'. Such words were not included in the analysis. Exclusion involved removing those target items from the calculations resulting in proportional data. An arcsine transformation of the data was performed for data analysis to ensure normality of the data. A list of the alternate word forms used can be seen in Appendix VII.

The data were coded in two ways in accordance with the inclusion criteria stated above. First, the data for both tasks 1 and 2 were coded for mention of target item for each of the target noun types. Second, the data were coded for mentions of the target items in the foreground and the background of the pictures.

The initial coding strategy was to determine whether Welsh and English participants performed differently on the c/u category since it is this category that differs across the two languages. It was hypothesised that the Welsh- and English-speaking participants would attend to, thereby mention, the target items of the sg/pl noun type equally often. It was also expected that there would be no difference between the two language groups for the substance items; neither language group would be inclined to mention items of this type. For the c/u target items, it was hypothesised that English-speaking participants will attend to and mention the items of this type in the same way as sg/pl types. However, the Welsh-speaking participants were expected to attend to and mention the c/u items less than English speakers thereby treating these items more like substance items.

The second coding strategy aimed to examine whether Welsh- and English-speaking participants performed differently on foreground items over background items. It was hypothesised that Welsh and English speakers overall would attend to and mention foreground target items over background target items. This was expected regardless of the category type.

Task 1: Description with picture in view.

The first coding strategy allowed for the analysis of the effects of language structure on participants' attention to target items labelled by different noun types. The means and standard deviations for mentions of target items for each type in this task are shown in Table 6.3.

A language x age x type (sg/pl, c/u, substance) ANOVA found significant interactions for language x type, $F(1,140) = 4.96$, $p<.028$, and age x type $F(3,140) = 3.01$, $p<.032$. These are shown in Figures 6.1 and 6.2 respectively.

Table 6.3 Means and Standard deviations for mentions of target items for each noun type.

Language	**Age**	**N =**	**Sg/pl**		**C/u**		**Substance**	
			M	**SD**	**M**	**SD**	**M**	**SD**
Welsh	4's	21	.6357	.3107	.7062	.3537	.6348	.3717
	8's	19	.6047	.4207	.9126	.2441	.6405	.3444
	10's	21	.5000	.3734	.9290	.1535	.7305	.3054
	adults	15	.6207	.3667	.8221	.2307	.6186	.3370
	Total	76	.5871	.3655	.8425	.2697	.6600	.3367
English	4's	24	.8133	.2268	.8054	.2775	.5833	.2839
	8's	15	.6333	.4096	.9560	.1161	.7013	.2832
	10's	17	.6271	.3714	.8829	.2018	.7059	.3317
	adults	17	.7347	.3498	.8835	.1626	.6282	.3316
	Total	73	.7147	.3368	.8726	.2125	.6466	.3051

Mean mentions of target item of each type by Welsh and English speakers

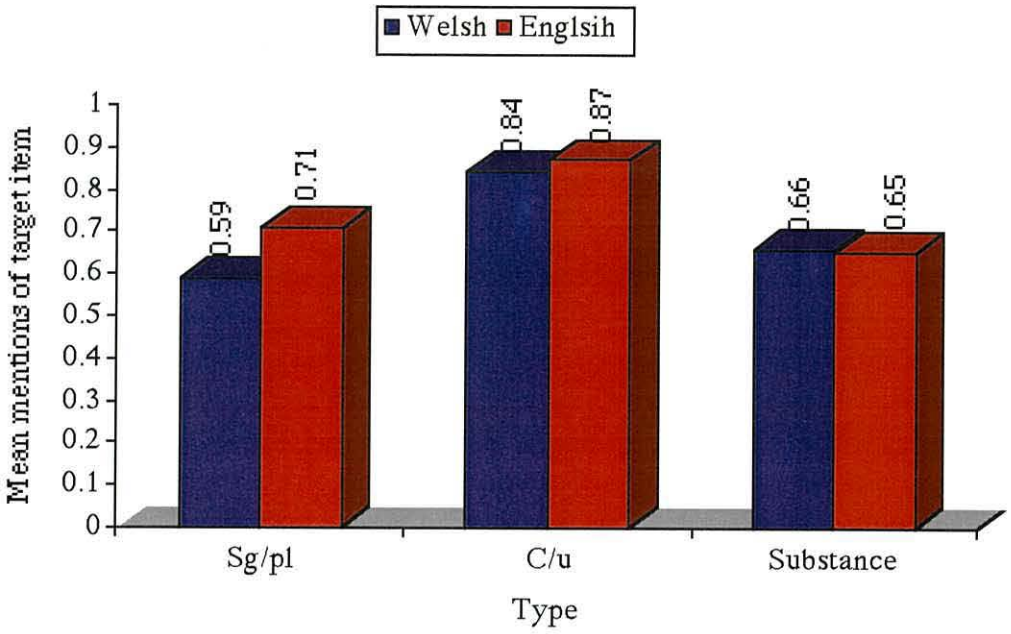


Figure 6.1 Mean mentions of each type by Welsh and English Speakers

Post hoc analysis (LSD) for language x type indicates that English-speaking participants mentioned the target items significantly more than the Welsh-speaking participants for the sg/pl type, $F(1,147) = 8.97$, $p<.003$, and the c/u type, $F(1,147) = 5.43$, $p<.021$.

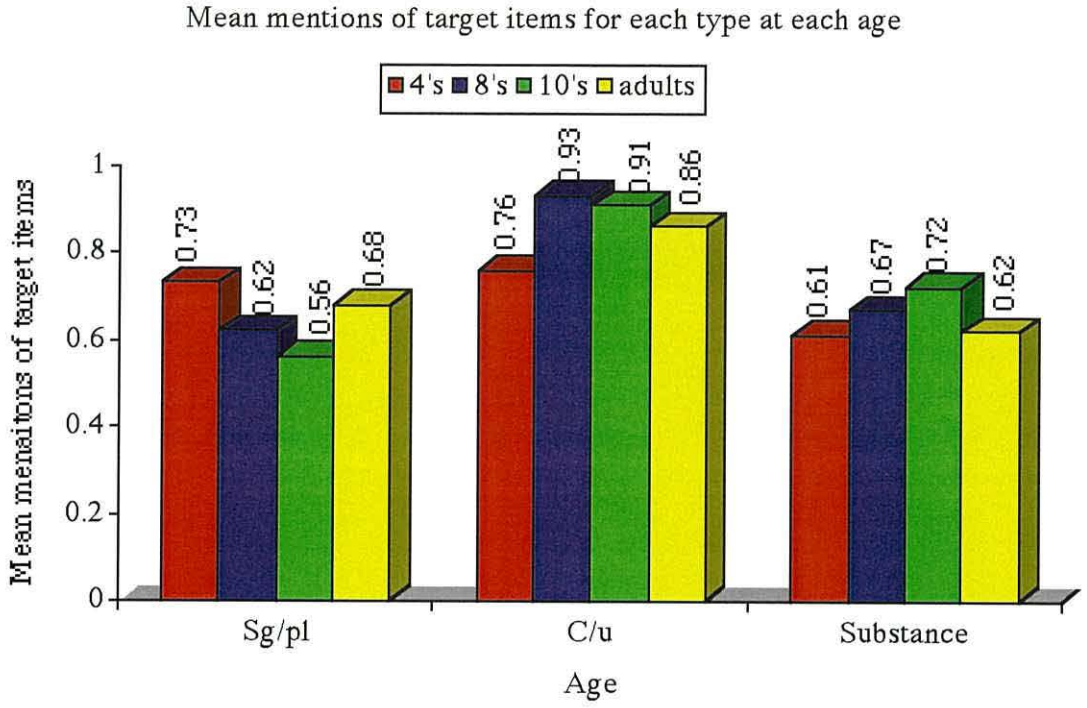


Figure 6.2 Mean mentions for each noun type at each age

Post hoc analysis (LSD) of the age x type interaction shows the 4-year-olds mentioned the c/u target items significantly more often than the substance type (MD = .296,p<.002); the 8-year-olds mentioned the c/u target items significantly more often than the sg/pl type (MD=.532,p<.000) and substance type (MD=.533,p<.000); the 10-year-olds mentioned the target items for the c/u type significantly more often than the sg/pl type (MD= . .604,p<.000) and substance type (MD= .349,p<.000), and mentioned the substance type items significantly more than the sg/pl type (MD= .255,p<.017). Adults also mentioned the c/u type significantly more often than both the sg/pl type (MD = .268, p,.019) and substance type (MD = .398, p<.000).

The second coding strategy allowed for the analysis of ground (foreground/background) effects on participants' attention to target items. The means and standard deviations for mentions of the target items in the foreground and background are shown in Table 6.4.

A language x age x ground (foreground/background) ANOVA was carried out. A significant main effect of ground, $F(1,140) = 8.18$, $p<.005$, and a significant interaction for age x ground, $F(3,140) = 7.43$, $p<.000$ was found. Participants mentioned foreground target items significantly more than the background target items as shown in Figure 6.3.

Table 6.4 Means and Standard deviations for mentions of target items in the foreground and background.

Language	Age	N =	Foreground		Background	
			M	SD	M	SD
Welsh	4's	21	.5665	.4171	.6650	.3043
	8's	19	.7374	2957	.6895	.2688
	10's	21	.7929	.3083	.7343	.2413
	adults	15	.8113	.3143	.5913	.2981
	Total	76	.7221	.3464	.6759	.2764
English	4's	24	.6254	.2330	.7683	.2339
	8's	15	.6900	.2266	.7813	.2603
	10's	17	.8241	.2914	.6818	.2773
	adults	17	.8429	.2675	.7012	.2514
	Total	73	.7356	.2667	.7352	.2521

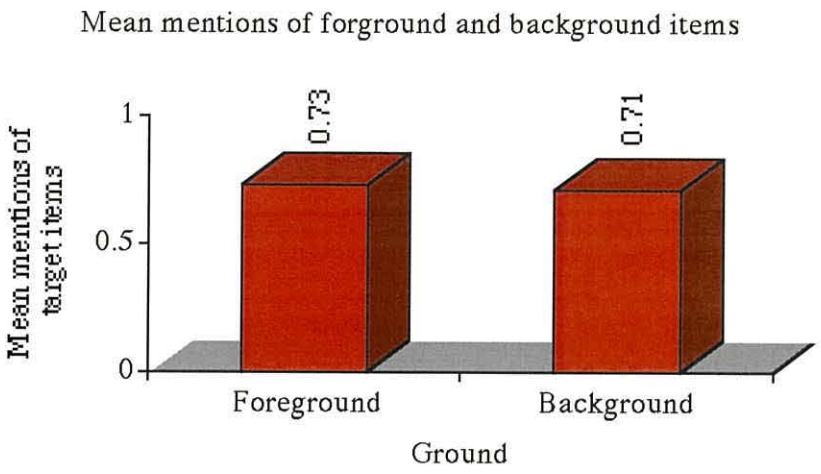


Figure 6.3 Mean mentions of foreground and background items.

Post hoc analysis (LSD) for the age x ground interaction revealed that the foreground target items were mentioned significantly more often than background items by the 10-year-olds (MD= .234, p<.041) and adults (MD= .431, p<.000). This interaction is depicted in Figure 6.4.

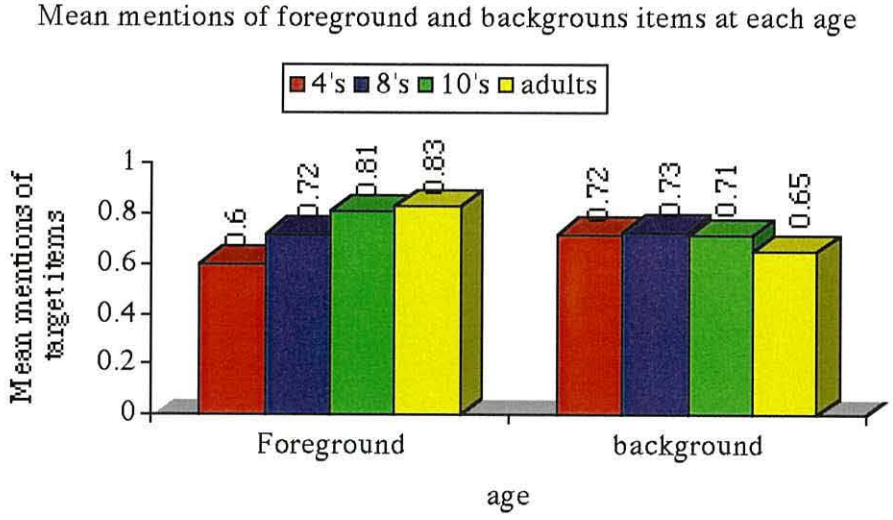


Figure 6.4 Mean mentions of target items in the foreground and background at each age

Summary

When describing a picture with the picture in view, English speakers mentioned items of the sg/pl and c/u type more often than Welsh speakers did. There is also a trend across the age groups for c/u target items to be mentioned more often than both the sg/pl and substance types. Overall, target items in the foreground were mentioned significantly more often than background items. This is largely due to the robust findings for the two older age groups.

Task 2: Description without the picture in view.

As with task 1, the first coding strategy allowed for the analysis of the effects of language structure on participants' attention to target items labelled by different noun types. The means and standard deviations for mentions of the target items for each of the three noun types for task 1 are shown in Table 6.5.

Table 6.5 Means and Standard deviations for mentions of target items for each noun type.

Language	Age	N =	Sg/pl		C/u		Substance	
			M	SD	M	SD	M	SD
Welsh	4's	21	.5071	.3108	.5557	.3009	.2048	.1882
	8's	19	.6421	.2382	.6674	.2893	.5084	.3712
	10's	21	.7467	.2568	.8457	.2209	.4290	.2876
	adults	15	.7871	2107	.7233	.2259	.5479	.1670
	Total	76	.6607	.2783	.6851	.2781	.4085	.2975
English	4's	24	.4579	.2391	.6113	.2907	.2213	.2341
	8's	15	.5567	.2431	.6340	.2388	.4327	.2672
	10's	17	.6294	.2332	.7259	.2638	.5388	.2810
	adults	17	.7476	.1443	.8141	.2495	.5394	.3521
	Total	73	.5856	.2423	.6899	.2721	.4127	.3104

A language x age x type (sg/pl, c/u, substance) ANOVA found significant main effects of type $F(1,141) = 35.97$, $p<.000$, and age $F(3,141) = 14.79$, $p<.000$. These are depicted in Figures 6.5 and 6.6 respectively.

Post hoc analysis (LSD) of type showed that the c/u type were mentioned significantly more often than the sg/pl type (MD = .144, p<.003) and substance (MD = .424, p<.000), and the sg/pl type were mentioned significantly more often than the substance (MD = .280, p<.000).

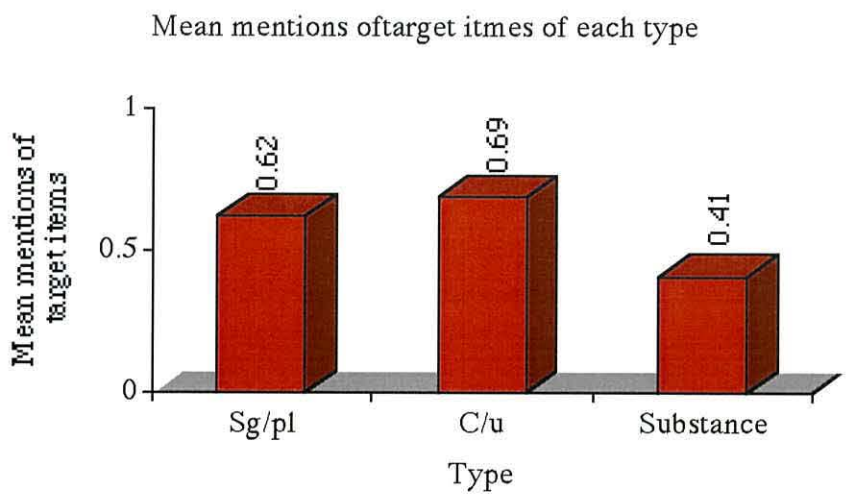


Figure 6.5 Mean mentions of target items for each noun type

For age, post hoc analysis (LSD) showed that adults mentioned the target items significantly more often than the 4-year-olds (MD = .452, p<.000), the 8-year-olds (MD = .256, p<.001) and the 10-year-olds (MD = .156, p<035). Both the 8-year-olds (MD = .196, p<.006) and the 10-year-olds (MD = .296, p<000) mentioned the target items significantly more often than the 4-year-olds.

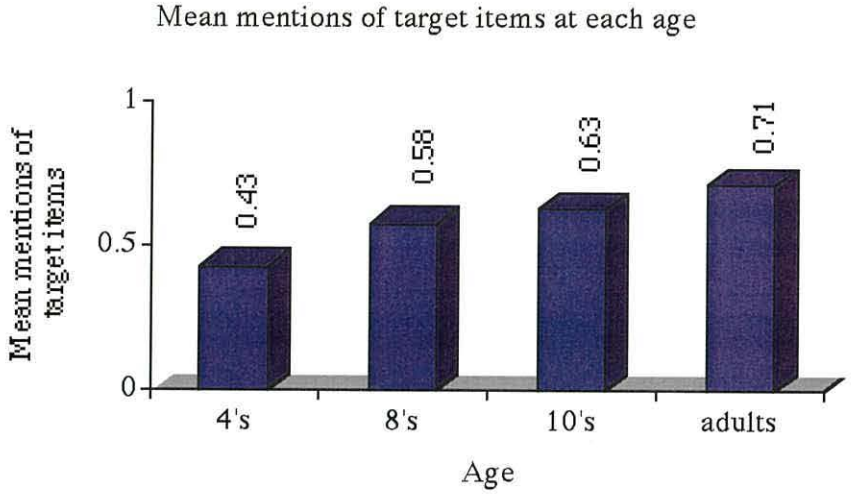


Figure 6.6 Mean mentions of target items at each age

A near significant interaction for language x type, $F(1,141) = 3.63$, $p<.059$ was also found. Post hoc analysis found that Welsh speakers mentioned the target items of the sg/pl type significantly more often than the English speakers did (MD = .168, $p<.022$). This is opposite to what was found in task 1 and is shown in Figure 6.7.

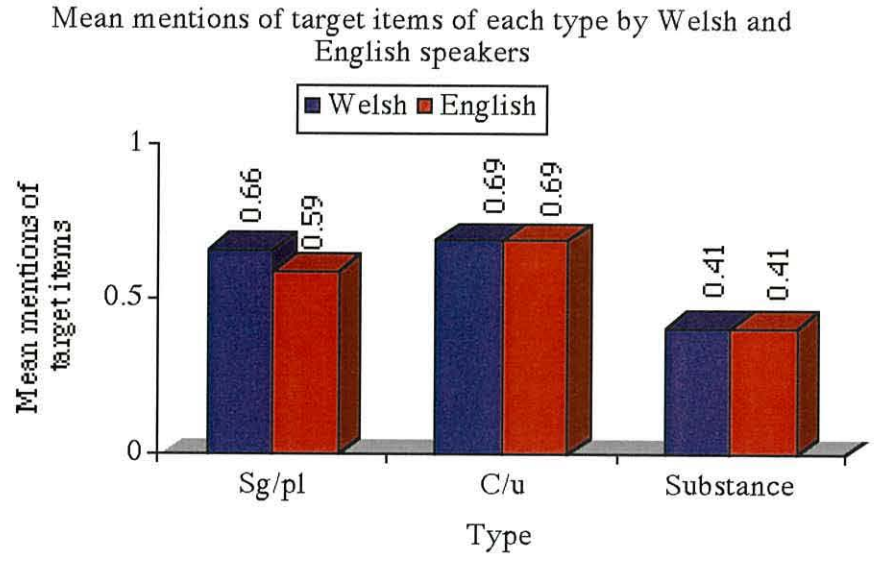


Figure 6.7 Mean mentions of target items of each noun type across the two language groups

As in task 1, the data were also coded and analysed for mentions of foreground and background items. The means and standard deviations for mentions of the target items in the foreground and background at each age for each language group are shown in Table 6.6.

A language x age x ground ANOVA revealed significant main effects of ground $F(1,141) = 257.52$, $p<.000$ and age $F(3,141) = 26.93$, $p<.000$, and a significant interaction for age x ground $F(3,141) = 5.85$, $p<.001$. The effect of ground indicates that target items in the foreground were mentioned significantly more often than items in the background of the pictures. This is illustrated in Figure 6.8.

Table 6.6 Means and Standard deviations for mentions of target items in the foreground and background.

Language	Age	N =	Foreground		Background	
			M	**SD**	**M**	**SD**
Welsh	4's	21	.6200	.2655	.3262	.2116
	8's	19	.7784	.3018	.5184	.2462
	10's	21	.9057	.1528	.5200	.1957
	adults	15	.9780	.0852	.6340	.1852
	Total	76	.8092	.2585	.4886	.2350
English	4's	24	.6121	.2142	.3346	.2063
	8's	15	.7567	.2348	.4380	.2044
	10's	17	.8729	.2167	.5059	.2375
	adults	17	.9418	.1297	.5859	.2230
	Total	73	.7793	.2393	.4542	.2342

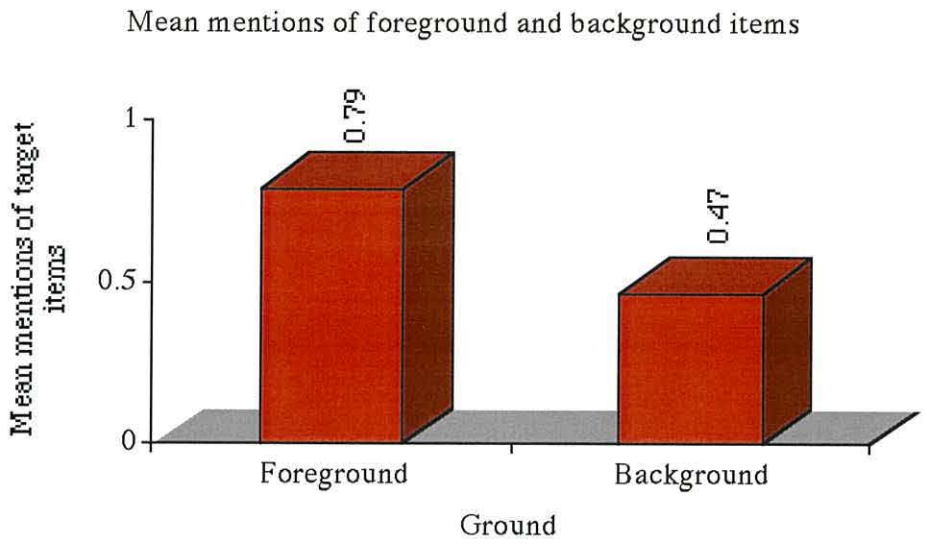


Figure 6.8 Mean mentions of target items in the foreground and background of a scene

The effect of ground shows that again, foreground items were mentioned significantly more often than background items as is depicted in Figure 6.8. The effect of age was found previously and can be seen in Figure 6.6.

Post hoc analysis (LSD) of the age x ground interaction shows that 4-year-olds mentioned foreground target items significantly less often than 8-year-olds (MD = -.322, p<.001), the 10-year-olds (MD = -.571, p<.000) and adults (MD = -.571, p<.000). The 8-year-olds also mentioned foreground target items significantly less often than the 10-year-olds (MD = -.248, p<.012), and adults (MD = -.402, p<.000). Background target items were also mentioned significantly less often by 4-year-olds than by 8-year-olds (MD = -.186, p<.005), 10-year-olds (MD = -.221, p<.001) and adults (MD = -.351, p<.000). The 8-year-olds also mentioned background items significantly less often than the adults (MD = -.164, p<.020). Figure 6.9 depicts this interaction and illustrates the finding that foreground items were mentioned more often than background items for each age group.

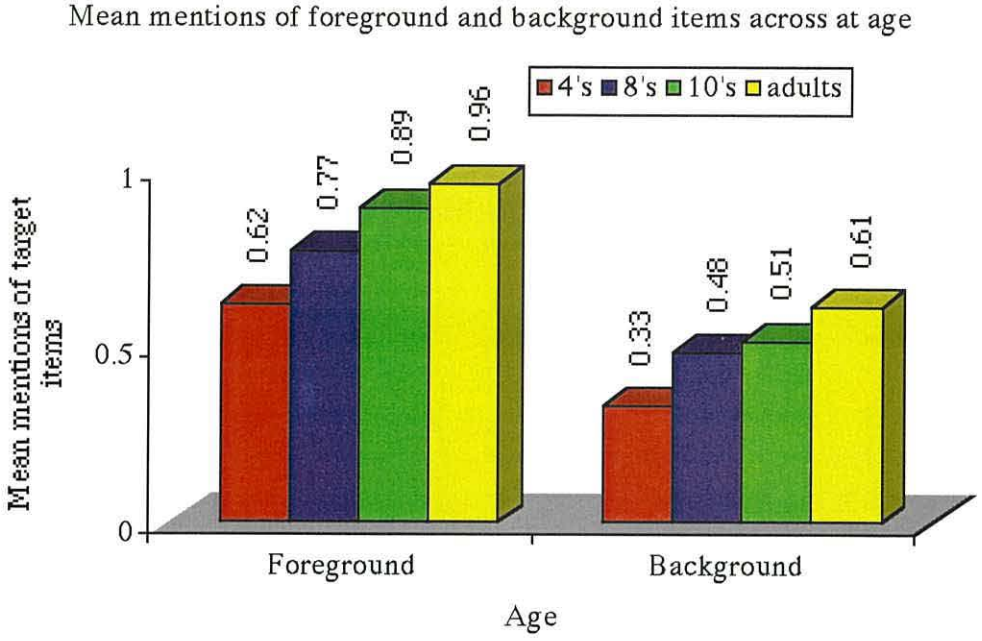


Figure 6.9 Mean mentions of target items in the foreground and background at each age

Summary

Participants describing pictures when the picture was not in view mentioned both c/u and sg/pl target item types more often than the substance type. The trends across the age groups shows an increase in mentions of target items with increasing age. Welsh speakers mention sg/pl target items more often than the English speakers did. As was found in task 1 when the picture was in view, target items in the foreground were mentioned significantly more often than items in the background.

Tasks 1 and 2: Descriptions with and without the picture in view

Further analysis was carried out to compare mentions of target items when a picture was or was not in view. Means and standard deviations can be seen in previous Tables for each task.

A language x age x task x type ANOVA revealed significant main effects of task $F(1,140) = 70.72$, $p<.000$, type $F(1,140) = 14.95$, $p<.000$, and age $F(3,140) = 4.56$, $p<.004$. These are depicted in Figures 6.10, 6.11, and 6.12 respectively.

The effect of task shows that participants mentioned the target items more in task 1 when the picture was in view than in task 2 when the picture was not in view as shown in Figure 6.10.

Post hoc analysis (LSD) of the type effect again shows that across the two tasks, the target items of the c/u type were mentioned significantly more often than the sg/pl (MD = .260,$p<.000$) and the substance type MD = .406,$p<.000$). The target items of the sg/pl type were also mentioned significantly more often than the substance type (MD = .147,$p<.000$). This can be seen in Figure 6.11.

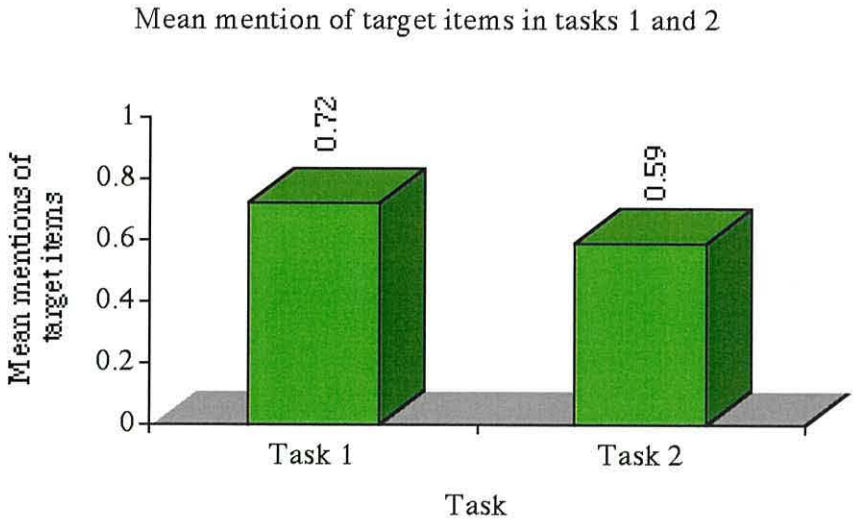


Figure 6.10 Mean mention of target items across the two tasks

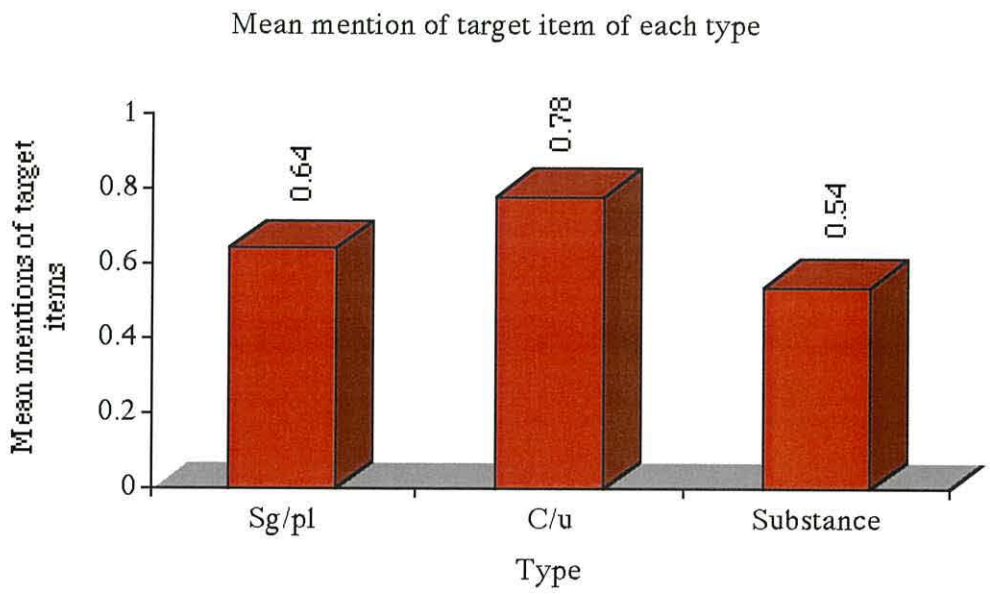


Figure 6.11 Mean mention of target items for sg/pl, c/u and substance type

Post hoc analysis for the effect of age across the two tasks shows that the 4-year-olds mentioned the target items significantly less than the 8-year-olds (MD = .153,p<.028), 10-year-olds (MD = .187,p<.006) and adults (MD = .239,p<.001). This is shown in Figure 6.12.

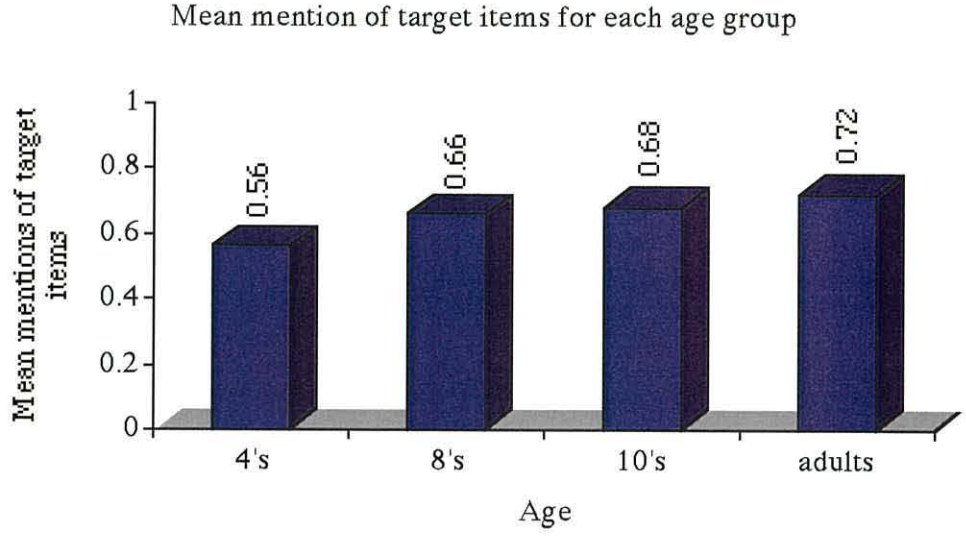


Figure 6.12 Mean mentions of target items at each age

Significant interactions were found for task x age $F(3,140) = 7.07$, $p<.000$, task x type $F(1,140) = 16.64$, $p<.000$, and task x type x language $F(1,140) = 11.15$, $p<.001$.

Post hoc analysis (LSD) for task x age indicates that target items were mentioned significantly more often in task 1 than in task 2 by the 4-year-olds (MD = .458,$p<.000$), the 8-year-olds (MD = .358,$p<.000$) and the 10-year-olds (MD = .233,$p<.001$). This interaction can be seen in Figure 6.13.

Mean mentions of target items for each age across the two tasks

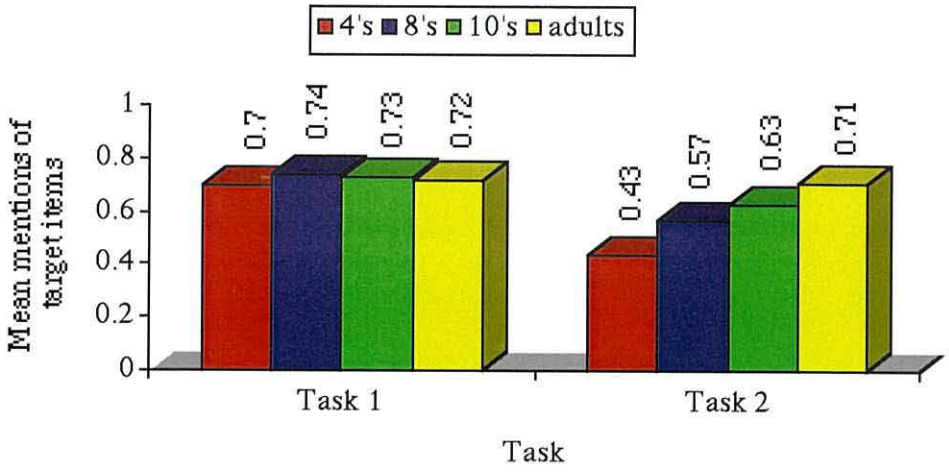


Figure 6.13 Mean mentions of target items in tasks 1 and 2 at each age

Post hoc analysis (LSD) of task x type indicates that for task 1, c/u target items were mentioned significantly more often than both the sg/pl type (MD = .364,p.000), and substance type (MD = .356,p.000). In task 2 the c/u target items were also mentioned significantly more often than both the sg/pl (MD = .146,p.002) and substance types (MD = .424,p.000), and the sg/pl type was mentioned significantly more often than the substance type (MD = .278,p.000). This interaction is shown in Figure 6.14 and illustrates the trend that target items were mentioned more often in task 1 than in task 2 for each type, but especially for the c/u and substance type.

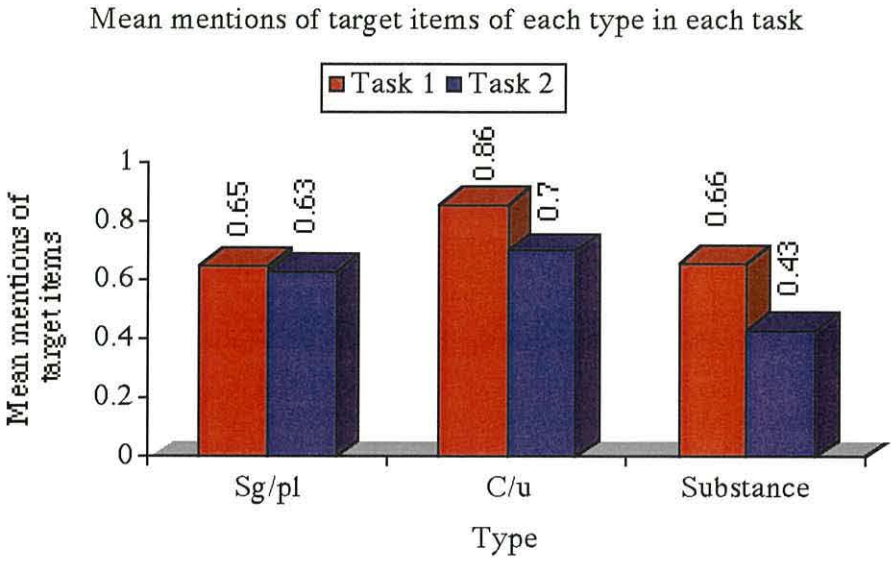


Figure 6.14 Mean mentions of target items for each type in tasks 1 and 2

For task x type x language interaction, post hoc analysis (LSD) shows that English speakers mentioned the target items significantly more often than the Welsh speakers for the sg/pl type in task 1 only (MD = .128, $p<.028$). Figure 6.15 depicts this interaction.

A near significant interaction was found for task x language $F(1,140) = 3.61$, $p<.059$. This is shown in Figure 6.16. No significant differences were found between the two language groups in the post hoc analysis, but a general trend indicates that English speakers mentioned more target items than Welsh speakers in task 1 as can be seen from Figure 6.16.

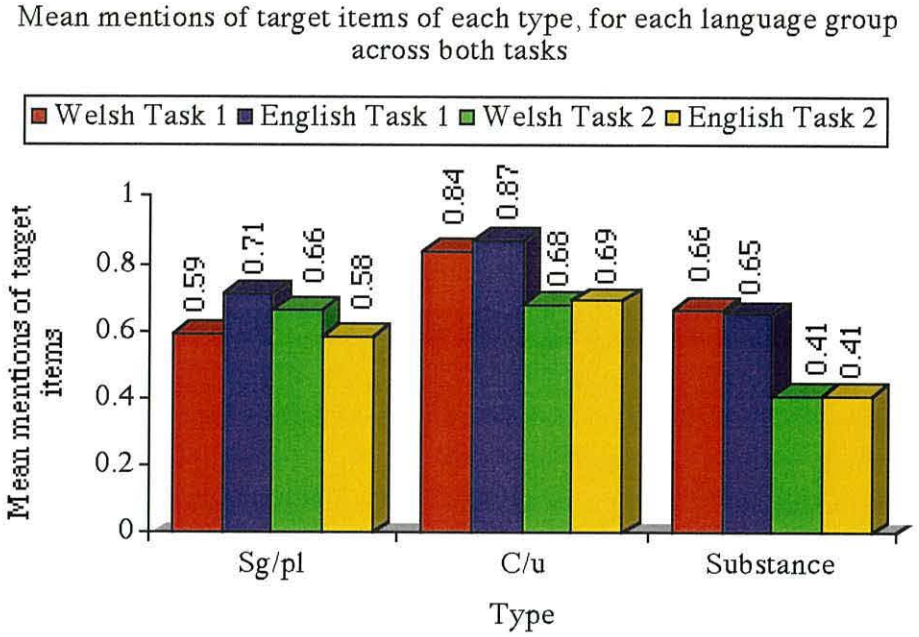


Figure 6.15 Mean mentions of target items of each type by each language group for tasks 1 and 2.

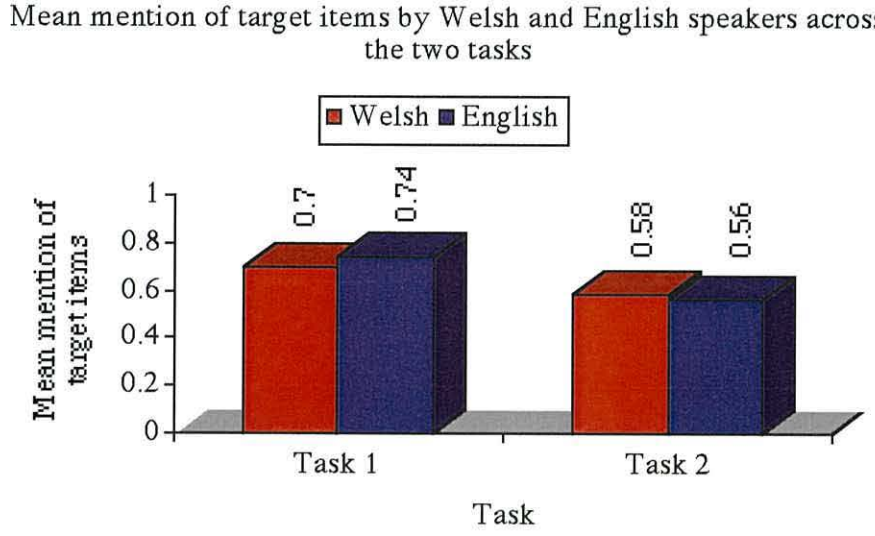


Figure 6.16 Mean mention of target items in each task by each language group

An analysis was also carried out to compare the foreground and background effects across the two tasks. A language x age x task x ground ANOVA revealed significant main effects of task $F(1,138) = 14.89$, $p<.000$, ground $F(1,138) = 148.05$, $p<.000$, and age $F(3,140) = 11.23$, $p<.000$.

Again, the main effect of task shows that participants mentioned target items more in task 1 than in task 2 as was previously shown in Figure 6.11. Similarly, the main effect of age across the two tasks shows that participants mention target items more often with increasing age, as was previously shown in Figure 6.13. The main effect of ground indicates that across the two tasks, participants mentioned the target items in the foreground significantly more often than background target items. This is depicted in Figure 6.17.

Mean mentions of target items in the foreground and background

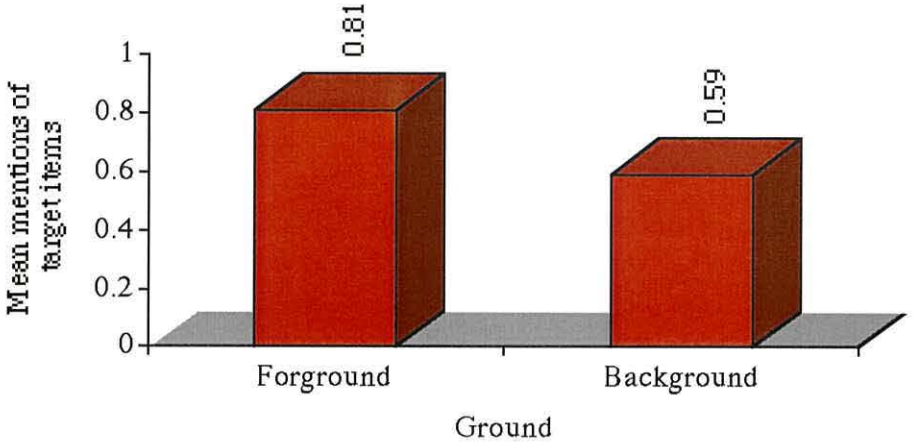


Figure 6.17 Mean mention of target items in the foreground and background

Significant interactions were found for task x age $F(3,138) = 5.352$, $p<.002$, ground x age $F(3,138) = 12.539$, $p<.000$, and task x ground $F(3,140) = 4.56$, $p<.004$. Again, the task x age interaction was shown previously in Figure 6.14.

Post hoc analysis (LSD) of the ground x age interaction indicates that across both tasks, the 4-year-olds mention foreground items significantly less often than the 8-year-olds (MD = -.251,$p<.003$), 10-year-olds (MD = -.488, $p<.000$) and adults (MD = .584, $p<.000$). The 8-year-olds also mention foreground items significantly less often than the10-year-olds (MD = --.237, $p<.007$) and adults (MD = -.333, $p<.027$). No differences were observed for the background items. Figure 6.18 shows this interaction and also illustrates the trend that foreground items are mentioned more often than background items at each age.

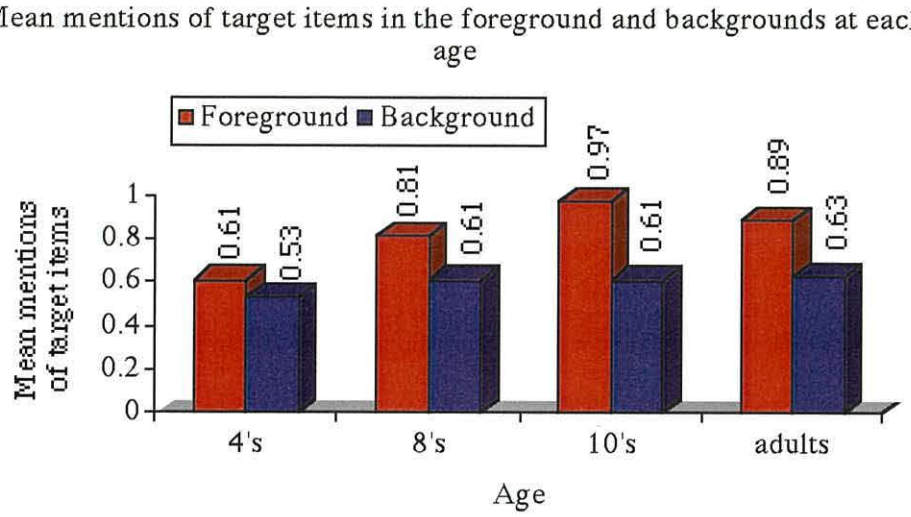


Figure 6.18 Mean mentions of target items in the foreground and background for each age group

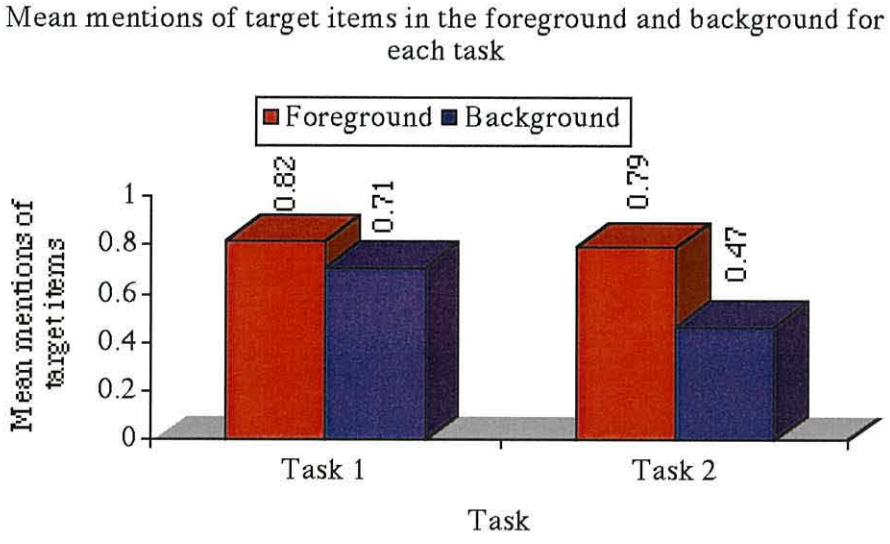


Figure 6.19 Mean mentions of target items in the foreground and background in tasks 1 and 2

Post hoc analysis (LSD) for the ground x task interaction indicates that target items were mentioned significantly more often in task 1 than in task 2 (MD = .393, $p<.000$) for background items only. This interaction is depicted in Figure 6.19.

Summary

Overall, participants mentioned target items more often in task 1 (with the picture in view) than in task 2 (with no picture in view). This effect is robust for the younger three age groups but not for the adults. Target items of each type were mentioned more often in task 1 than task 2, and, across both tasks, the sg/pl and c/u type were mentioned more often than the substance type. The c/u was also mentioned more often than the sg/pl type. An age trend shows that mentions of target items increase with increasing age for both tasks. Target items in the foreground were mentioned more often than background items at each age, older children and adults mentioned foreground items more than younger children, and only background items were mentioned more often in task 1 than task 2.

The results show that (1) adults attended to more items than children when a picture was not in view, (2) foreground items were attended to more than background items, and (3) the youngest children do not distinguish between foreground and background items.

Picture selection responses

Picture selections were coded in specific ways for the three picture selection tasks.

Task 3: Judging the variant picture most like the original

For this task, picture selections were coded in two ways. First, the choice of variant picture was coded for each type. Second, picture selections were coded for choice of variant with foreground and background *number* changes. A score of 1 was assigned to the chosen variants and a score of 0 was assigned to variants not chosen.

The initial coding strategy was to determine whether Welsh and English participants performed differently on the c/u category since it is this category that differs across the two languages. It was hypothesised that neither the Welsh- nor

English- speaking subjects would be inclined to choose the sg/pl as being the variant most similar to the original. For the English-speaking participants a similar prediction was made regarding the c/u category because it is a sg/pl category in English. It was expected then that English speaking participants would favour the substance variant over the other two types. For Welsh speaking participants, it was expected that they would choose the c/u variant as often as the substance as being most similar to the original picture. This is because Welsh speakers were not expected to pay as much attention to the *number* of items in the c/u category, as they did to items in the sg/pl category.

The second coding strategy aimed to examine whether Welsh- and English-speaking participants performed differently according to whether items were in the foreground or the background. It was hypothesised that Welsh and English speakers overall would be more likely to select variant pictures with background *number* changes. This was expected regardless of the category type. However, it was expected to be most striking in relation to pictures with the sg/pl type in the foreground. For the picture with the substance in the foreground it was expected that both language groups might be willing to choose the substance variant even when this category was foreground material. For the pictures with c/u in the foreground it was expected that English speakers would be reluctant to select the c/u variant, whereas Welsh speakers might be more prone to doing so.

The means and standard deviations for picture choice are shown in Table 6.7 for each age and each language group. Figure 6.20 depicts the choice of each variant picture type for both language groups. Picture selections of each variant type for each age group is shown in Figure 6.21.

Table 6.7 Means and standard deviations for picture choice

Language	Age	Sg/pl variant picture		C/u variant picture		Substance variant picture	
		M	SD	M	SD	M	SD
Welsh	4's	.3333	.2222	.2982	.2485	.3684	.2918
	8's	.2315	.2672	.1574	.1850	.6111	.3078
	10's	.1270	.1659	.1746	.1706	.6984	.2083
	Adults	.0667	.1380	.1333	.1690	.8000	.1690
	Total	.1864	.2090	.1864	.2018	.5877	.3121
English	4's	.3333	.3108	.3056	.1945	.3611	.2766
	8's	.2222	.2722	.3333	.2817	.4444	.3253
	10's	.2549	.2508	.1765	.1715	.5686	.3284
	Adults	.0980	.1960	.2353	.1566	.6667	.2041
	Total	.2374	.2749	.2648	.2076	.4977	.3049

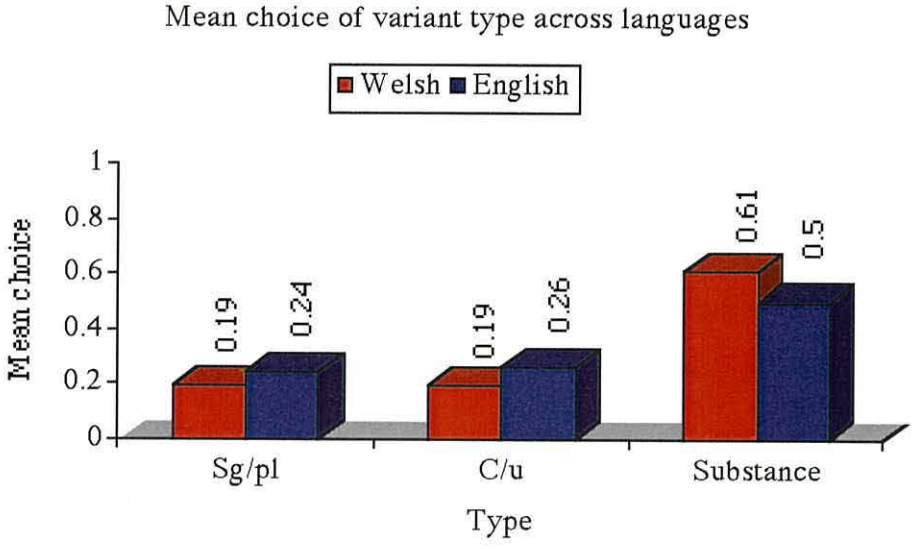


Figure 6.20 Mean choice of variant pictures for each language group

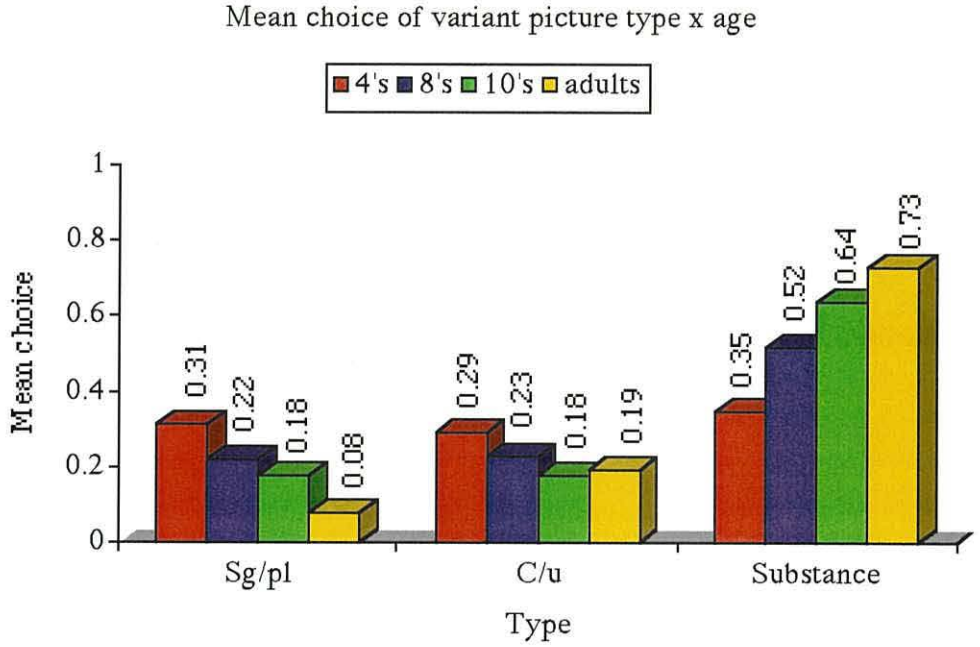


Figure 6.21 Mean choice of variant picture for each age group

Recall that participants chose only one variant picture. Picture selections then were not mutually exclusive. The data for one category type is dependent on the data of another category type. Consequently, analysis was carried out for each type individually. However, it is immediately clear from Figure 6.21 that responses by 4-year-olds did not differ from chance (33%) across types.

Initially, analysis was carried out on the c/u choices because it is this category that differs across the two languages. A language x age ANOVA of the c/u choices revealed significant main effects of language, $F(1,141) = 5.91$, $p<.016$ and age, $F(3,141) = 87.21$, $p<.047$. The main effect of language shows that the English speakers chose the c/u variant as being most similar significantly more often than Welsh speakers. This can be seen for the c/u type in Figure 6.20 (Note that this is contrary to the prediction).

Post hoc analysis (LSD) for age shows that the 4-year-olds chose the c/u variant type significantly more often than the 10-year-olds (MD = .114, $p<.011$) and the adults (MD = .101, $p<.031$) as depicted for the c/u type in Figure 6.21.

To examine the effects of variant type on picture selection further, analysis was carried out on sg/pl and substance variant choices as well. A language x age ANOVA of the sg/pl choices revealed a significant main effects of age only, $F(3,141) = 6.541$, $p<.000$ and can be seen for the sg/pl type in Figure 6.21. Post hoc analysis (LSD) shows that the 4-year-olds chose the sg/pl variant type significantly more often than the 10-year-olds (MD = .134, $p<.009$) and the adults (MD = .235, $p<.000$). The 8-year-olds also chose the sg/pl variant significantly more often than the adults (MD = .137, $p<.017$).

For substance variant choices, a language x age ANOVA indicated a significant main effect of language, $F(1,141) = 4.11$, $p<.044$ and age, $F(3,141) = 14.18$, $p<.000$. The main effect of language shows that the Welsh speakers chose the substance variant significantly more often than the English speakers. This language effect can also be see in Figure 6.20.

Post hoc analysis (LSD) of age shows that the 4-year-olds chose the substance variant type significantly less often than the 8-year-olds (MD = -.172, $p<.007$), the 10-year-olds (MD = -.292, $p<.000$) and the adults (MD = -.381, $p<.000$). The 8-year-olds also chose the substance variant significantly less often than the adults (MD = -.210, $p<.017$). This effect can also be seen in Figure 6.21 for the substance type.

The second coding strategy allowed analysis of the effects of foreground/background *number* changes on picture selection. The means and standard deviations for picture choice of foreground or background *number* changes are shown in Table 6.8. Picture selection of foreground and background variants for each type is depicted in Figure 6.22.

Note that participants were asked to select only one variant picture. Again, foreground and background variant choices were not mutually exclusive. Indeed, background choices are inversely proportional to foreground choices. Furthermore, participants had twice as many chances of choosing a variant picture

with background *number* changes than with foreground *number* changes. For these reasons, only the choices of variants with foreground *number* changes were analysed.

Table 6.8 Mean and standard deviations for choice of variant with foreground and background *number* changes.

Language	**Age**	**Foreground *number* change**		**Background *number* change**	
		M	SD	M	SD
Welsh	4's	.2807	.2294	.7192	.2294
	8's	.2507	.2294	.7500	.2816
	10's	.1111	.1610	.8888	.1610
	Adults	.1333	.1690	.8666	.1690
	Total	.1864	.2227	.7742	.2704
English	4's	.3611	.3095	.6388	.3096
	8's	.2444	.1979	.7556	.1978
	10's	.1961	.1691	.8040	.1690
	Adults	.1569	.2915	.8432	.2914
	Total	.2511	.2652	.7488	.2652

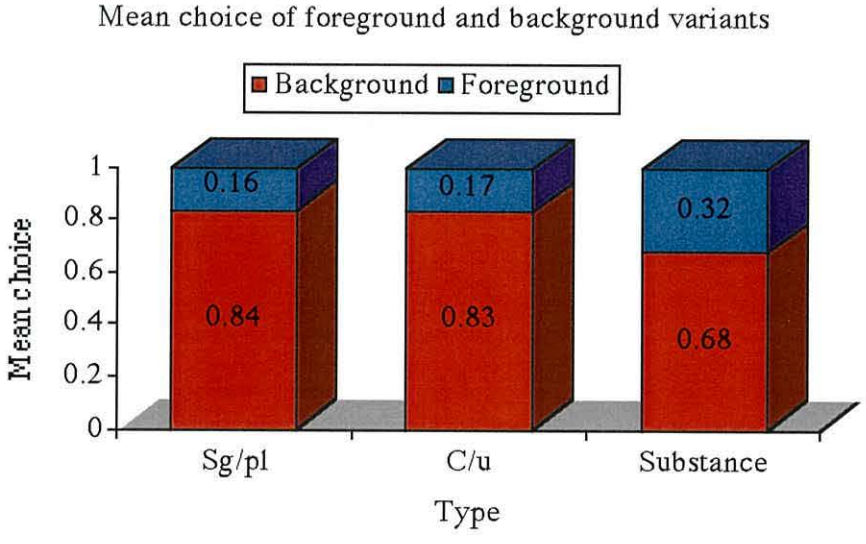


Figure 6.22 Mean choice of variant pictures with foreground and background *number* changes

Again, because picture choices are not independent, only a language x age ANOVA could be carried out on choice of variant with foreground *number* changes. A significant main effect of age only $F(1,141) = 35.04, p<.000$ was found. However, as can be seen from Figure 6.22, participants chose the substance variant more often even when substances were in the foreground.

Post hoc analysis (LSD) shows that 4-year-olds chose variant pictures with foreground *number* changes significantly more often than the 10-year-olds (MD = .162, $p<.002$) and adults (MD = .165, $p<.003$). This interaction can be see in Figure 6.23.

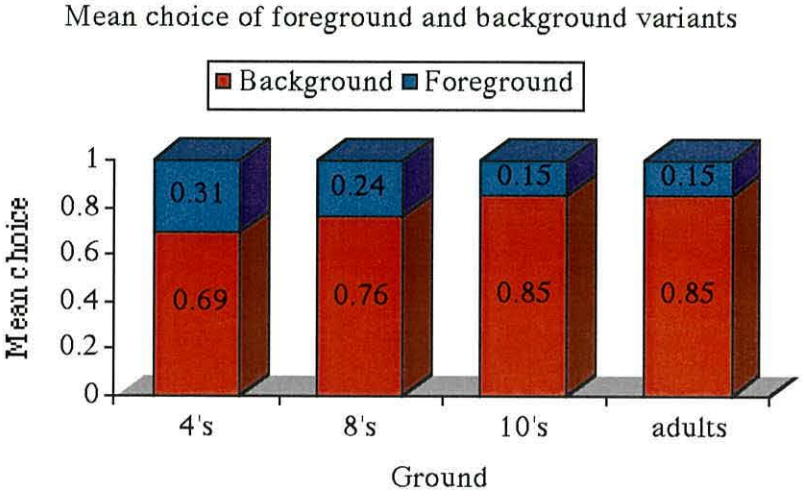


Figure 6.23 Mean choice of variants with foreground and background number changes at each age

Summary

English speakers chose the c/u variant pictures to be most like the original more often than the Welsh speakers. Welsh speakers chose the substance variant pictures more often than the English speakers. The younger age groups chose the c/u and sg/pl as most like the original more often than the older age groups while the older age groups chose the substance variant as most like the original more than the younger age groups. This was also true for variants with foreground *number* changes;

a decrease in choices of foreground variants was found with increasing age. When items were in the foreground, participants chose the substance variants more often than the sg/pl and c/u variants.

Task 4: Finding the original (short-term memory)

Again, two coding strategies were adopted for tasks 4 and 5. First, the data were coded for correctness of response because the tasks required that participants select from an array the original picture they had seen. A correct selection was assigned a score of 1, and an incorrect selection was assigned a score of 0. Some participants selected more than one picture to be the same as the original. Selecting the original was coded as correct only when no other alternative pictures were selected in addition. Second, an analysis of errors was required to determine which alternative pictures were being incorrectly selected. Here, the data are nominal.

The first coding strategy was to examine whether Welsh and English speakers were equally sensitive to changes in *number* and thereby correctly identified the original picture. The second coding strategy was to determine whether there were any differences across the two language groups in the picture type that may have been incorrectly chosen. It was hypothesised that neither Welsh- nor English-speaking participants would choose the sg/pl variant as the original because they would be sensitive to changes in *number* for this type. Conversely, it was also hypothesised that both Welsh and English would be more inclined to choose the substance variant because both language groups were expected to be insensitive to *number* changes for this type. For the c/u type, it was hypothesised that Welsh speakers would be more inclined than the English speakers to choose c/u variant pictures because the Welsh were expected to be less sensitive to changes in *number* for this type.

Analysis of correct choice of original picture

For task 4, Table 6.9 shows the mean number of correct choices and their standard deviation.

A language x age ANOVA showed a significant main effect of age only, $F(3,140) = 24.66$ $p<.000$. Post hoc analysis (LSD) of age indicated that adults chose the correct picture significantly more often than the 4-year-olds (MD = .580, p<.000), the 8-year-olds (MD = .490, p<.000) and the 10-year-olds (MD = .266, p<.001). The 10-year-olds also chose the correct picture significantly more often than the 4-year-olds (MD = .314, p<.000) and 8-year-olds (MD = .224, p<.003). These findings are depicted in Figure 6.16.

Table 6.9 Means and standard deviations for correct picture choices

Language	Age	M	SD
Welsh	4-year-olds	.2333	.2882
	8-year-olds	.3684	.3832
	10-year-	.5873	.3637
	Adults	.8444	.2477
	Total	.4889	.3926
English	4-year-olds	.3333	.2408
	8-year-olds	.3778	.3301
	10-year-	.6078	.3581
	Adults	.8824	.2340
	Total	.5342	.3589

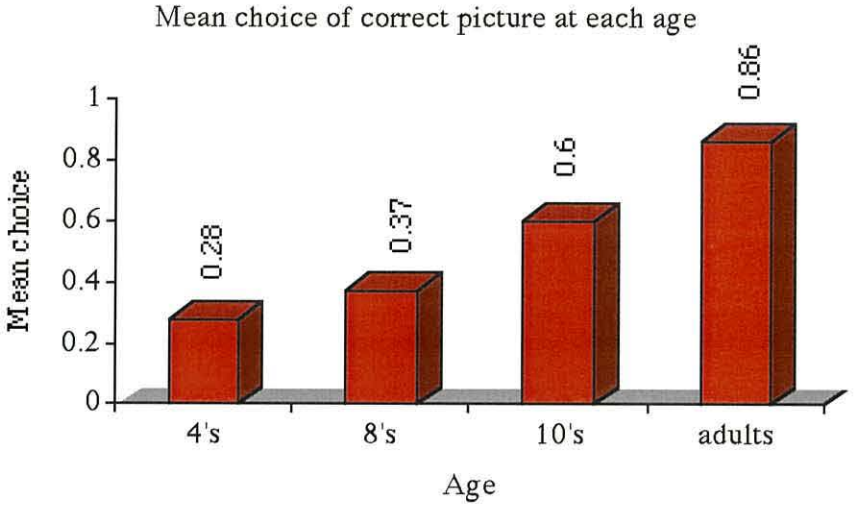


Figure 6.24 Mean choice of correct picture at each age

Analysis of picture selection errors

To examine whether language differences influenced picture selection when variant pictures were erroneously chosen, analyses was carried out in two ways. First, the frequencies of incorrect choices of each variant picture type were analysed. These data are shown in Table 6.10. Second, frequencies of incorrect choices of variant pictures with foreground versus background *number* changes were analysed. These data are shown in Table 6.11. Note that some participants chose more than one variant picture to be the original and some chose none. Analyses are based on all the incorrect responses regardless of whether two or more pictures were chosen by the same participant. This was to ensure that all errors were included in the analysis.

Table 6.10 Number of incorrect variant picture choices for each type

Language	Age	N	Sg/pl variant picture	C/u variant picture	Substance variant picture	Total
Welsh	4's	21	23	17	16	56
	8's	19	13	18	14	45
	10's	21	6	11	11	28
	Adults	15	3	1	4	8
	Total	73	45	47	45	137
English	4's	24	18	15	15	48
	8's	15	7	11	10	28
	10's	17	8	9	3	20
	Adults	17	2	2	1	5
	Total	73	35	37	29	101
	Overall Total	146	80	84	74	**238**

A language x type chi square analysis was carried out on the frequency of incorrect picture selections at each age. No effect of language or type was detected for any age group. As can be seen from Table 6.10, participants did not appear to favour any variant picture type over another. But, trends indicate that Welsh speakers made more errors than English speakers and fewer errors were made with age. This suggests that Welsh speakers, overall, were attending to *number* details less than English speakers. And, attention to *number* improves with age. However, Table shows that the Welsh 10-year-old participants erroneously selected the c/u and substance type more often than the sg/pl type, while the English 10-year-olds erroneously selected the sg/pl and c/u type more often than the substance type.

To further explore picture selection errors, the data were examined for foreground and background effects. Table 6.11 shows the number of incorrect choices of variant pictures with foreground and background *number* changes. A language x ground chi square analysis was carried out on the frequency of incorrect picture selections at each age. A significant interaction was found for language x

ground for the 10-year-olds only (χ^2 (1, n=48) = 5.486, p<.019). As can be seen from the Table 6.11, Welsh 10-year-olds chose variant pictures with foreground changes almost twice as often as variants with background changes. The English 10-year-olds however, chose variants with background changes more often than variants with foreground changes as expected. Recall that participants had twice as many opportunities to select a variant with background changes, yet overall, the frequency of errors do not reflect this. More variants with foreground changes were erroneously chosen than was expected.

Table 6.11 Number of incorrect choices of variant pictures with foreground and background *number* changes

Language	**Age**	**N**	**Foreground**	**Background**	**Total**
Welsh	4's	21	18	38	56
	8's	19	20	25	45
	10's	21	18	10	28
	Adults	15	3	5	8
	Total	73	59	78	137
English	4's	24	19	29	48
	8's	15	11	17	28
	10's	17	6	14	20
	Adults	17	2	3	5
	Total	73	38	63	101
	Overall Total	146	97	141	**238**

Summary

Selecting the correct original picture was largely dependent on age. More correct selections were made with increasing age; adults made very few incorrect choices.

When variant pictures were chosen in error, there was little difference across the two languages or across the different variant types. A trend indicates that Welsh 10-year-olds appear to favour the c/u and substance variants over the sg/pl variants while the English 10-year-olds favoured the sg/pl and c/u type over the

substance type. The Welsh and English 10-year-olds also selected variants with foreground and background changes differently. Welsh 10-year-olds selected almost twice as many variants with foreground changes than background changes whilst the English 10-year-olds selected almost twice as many variants with background changes than foreground changes. Overall, Welsh speakers seemed to erroneously choose variant pictures as the original more often than English speakers. This suggests that Welsh speakers attend to *number* details less, and may be less sensitive to *number* changes, than English speakers.

Task 5:Finding the original (long-term memory)

The coding strategy for task 5 was identical to that undertaken in task 4.

Analysis of correct choice of original picture

The means and standard deviations for correct choices of picture can be seen in Table 6.12.

Table 6.12 Means and standard deviations for correct choices of picture for each type

Language	Age	M	SD
Welsh	4's	.1667	.2757
	8's	.3509	.3420
	10's	.5556	.3177
	Adults	.8000	.3034
	Total	.4489	.3806
English	4's	.2778	.2538
	8's	.2444	.3204
	10's	.4902	.3144
	Adults	.7059	.4230
	Total	.4201	.3686

A language x age x type ANOVA showed a significant main effect of age, $F\,3,140 = 20.08$, $p<.000$. Post hoc analysis (LSD) indicates that adults chose correct

pictures significantly more often than the 4-year-olds (MD = .531, p<.000), 8-year-olds (MD = .455, p<.000) and 10-year-olds (MD = .230, p<.003). The 10-year-olds also chose correct pictures significantly more often than the 4-year-olds (MD = .301, p<.000) and 8-year-olds (MD = .225, p<.003). These findings can be seen in Figure 6.24.

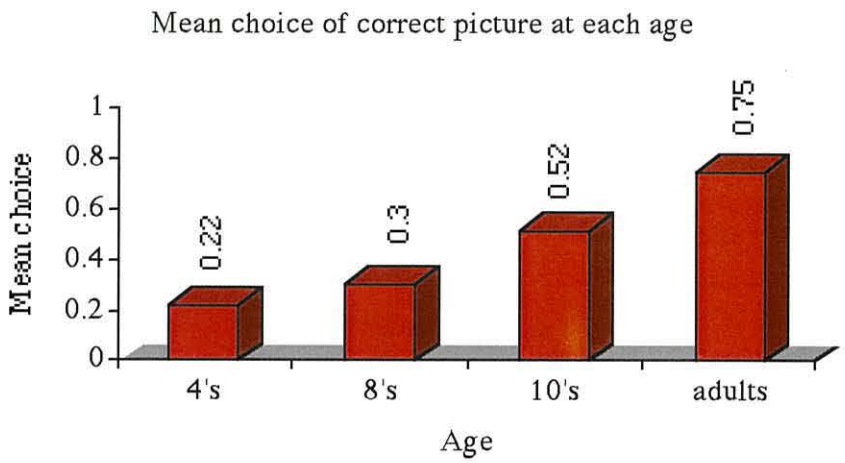


Figure 6.25 Mean correct choices at each age.

Analysis of picture selection errors

As with the short-interval task, an analysis was carried out to examine whether language differences affected picture selection when variant pictures were erroneously chosen. Again, only the choices of each variant picture type and choices of pictures containing foreground *number* changes were analysed. Table 6.13 shows the number of incorrect variant picture choices for each type and Table 6.14 shows the number of incorrect variant picture choices with foreground and background *number* changes.

Table 6.13 Number of incorrect variant picture choices for each type

Language	Age	N	Sg/pl variant picture	C/u variant picture	Substance variant picture	Total
Welsh	4's	21	24	20	20	64
	8's	19	14	11	10	35
	10's	21	6	14	7	27
	Adults	15	0	7	4	11
	Total	73	44	52	41	137
English	4's	24	14	20	18	48
	8's	15	16	10	8	28
	10's	17	7	12	8	20
	Adults	17	2	8	5	5
	Total	73	39	50	38	127
	Overall Total	146	83	102	79	**264**

As for task 4, a language x type chi square analysis was carried out on the frequency of incorrect picture selections for each age. Again no effect of language or type was found for any age group. As can be seen from Table 6.13, trends indicates that Welsh speakers made more errors than English speakers and that fewer errors were made with increasing age but again participants did not favour any variant picture type over another.

Further analysis was carried out to explore picture selection errors by foreground and background *number* changes. Table 6.14 shows the number of incorrect choices of variant pictures with foreground and background *number* changes. A language x ground chi square analysis was carried out on the frequency of incorrect picture selections at each age and again no effect of language or ground was detected at any age. Table 6.14, shows that, in this task, more variant pictures with background changes were incorrectly chosen than those with foreground changes across the language and age groups. This reflects the fact that there were

twice as many chances of selecting variants with background changes than foreground changes.

Table 6.14 Number of incorrect choices of variant pictures with foreground and background *number* changes

Language	Age	N	Foreground	Background	Total
Welsh	4's	21	20	44	64
	8's	19	7	28	35
	10's	21	10	17	27
	Adults	15	2	9	11
	Total	73	39	98	137
English	4's	24	18	34	52
	8's	15	10	24	34
	10's	17	9	17	26
	Adults	17	5	10	15
	Total	73	42	85	127
	Overall Total	146	81	183	**264**

Summary

As was found with the short-interval task, selecting the correct original picture was largely dependent on age. More correct selections were made and fewer erroneous choices of variants were made with increasing age.

Of the erroneously chosen variant pictures, there was no effect of language, type, or ground on incorrectly choosing a picture at any age. As was noted in task 4, Welsh speakers seemed to make more incorrect choices than English speakers, but not to the same extent, and fewer incorrect choices were made with increasing age.

Task 4 and 5: short- and long-interval comparison

The short- and long-interval tasks were contrived to be similar with exception to the interval period between exposure to the picture and the test phase. The short-interval task had an interval period of 30 seconds aimed to address short-term memory or recall while the long-interval task had an interval of 30 minutes aimed to address long-term memory or recall. The similarity in the tasks allows for the data from the two tasks to be compared. This is to examine whether language differences are persistent across short-term and long-term. Consequently, long- and short-term recall was compared to determine whether attention to *number* decayed over time and to determine the relevance of *number* for the target items in longer-term memory. The means and standard deviations for correct choices and choices of variant pictures can be seen above for each task.

Analysis of correct choice of original picture

A language x age x task ANOVA was carried out on correct choices. A significant main effect of task, F 1,140 = 7.64, p<.006 and age, F 3,140 = 32.23, p<.000 was found.

The main effect of task shows that significantly more correct choices were made by participants in the short-interval task than the long interval task as can be seen in Figure 6.26.

Post hoc analysis (LSD) of age indicates that, across the two tasks, adults chose the correct picture significantly more often than the 4-year-olds (MD = .555, p<.000), the 8-year-olds (MD = .473, p<.000) and 10-year-olds (MD = .248, p<.000). The 10-year-olds also chose the correct picture significantly more often than the 4-year-olds (MD = .307, p<.000) and 8-year-olds (MD = .225, p<.004). An overall age effect would be expected given the age effects for each task and will therefore not be depicted graphically.

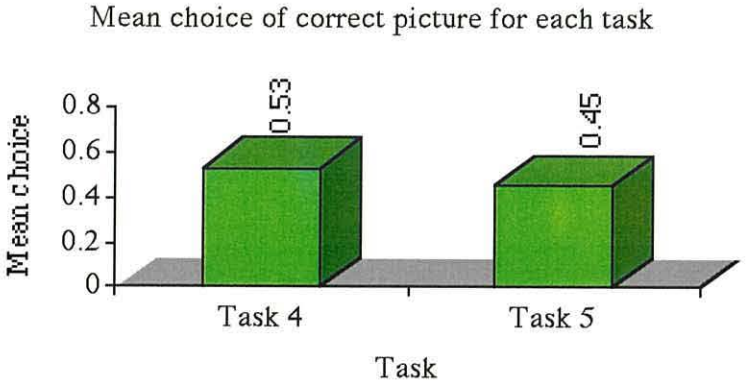


Figure 6.26 Mean correct picture choice across the two tasks.

Analysis of picture selection errors

More incorrect errors were made in tasks 5 (long interval) than 4 (short interval). Language and variant picture type had little effect on incorrect choices of variants in both tasks. Welsh and English speakers showed similar patterns of selecting variants at each age. In task 4, the data indicate that the Welsh 10-year-olds favoured the c/u and substance type over the sg/pl type while English 10-year-olds favoured the sg/pl and c/u over the substance type. Likewise, Welsh 10-year-olds selected variants with foreground changes more often than variants with background changes while the reverse was true for English 10-year-olds. This pattern was not found in the data for task 5.

Summary

Overall, participants correctly selected the original picture more in task 4 (short interval) than in task 5 (long interval). Again, age is a key factor in picture selection whereby more correct choices and fewer errors were made with increasing age. Language and variant type had little effect where variants were incorrectly chosen. Welsh 10-year-olds differed from English 10-year-olds with regard to variant

picture type chosen and selecting variants with foreground and background changes only in task 4. No differences were detected in task 5 at any age. Overall, Welsh speakers made more incorrect choices than English speakers, particularly in task 4.

Discussion

The tasks in this study aimed to examine Welsh and English speakers' attention to referents of different noun types and sensitivity to *number* changes by picture descriptions and picture selection procedures. For picture descriptions, mentioning target items was taken to indicate that participants were paying attention to the items of that type. In contrast, picture selections were taken to indicate that participants were paying less attention to *number* for the type of items that differed from the original.

Picture description

When describing a picture, with or without the picture in view, there was a tendency for participants to mention sg/pl type and c/u type target items more often than substance type. And, when the picture was in view, the c/u types were mentioned more often than the sg/pl type. This suggests that participants paid more attention to items of the sg/pl and c/u type than substances and that c/u items were particularly salient for both language groups. These findings are consistent with the findings of Lucy's study in that less attention was given to substances than countable items. However, the data do not support the prediction that Welsh speakers would treat the c/u target items like the substance type while the English speakers would treat the c/u like the sg/pl type.

A consistent and robust finding was that participants mentioned foreground items more often than background items. This indicates that Welsh- and English-speaking participants were paying more attention to foreground items than background items regardless of the target item type. Although it is intuitive that attention is drawn to foreground items, it is nevertheless an important consideration for studies that aim to measure attention to different items in a scene. In Lucy's study, target items were not controlled for foreground/background effects. Such effects may have serious implications for his findings.

Age was also a key factor. Mentioning the target items increased with age. However, differences in mentions of the distinct category types are demonstrated strongly in the older age groups only. Younger participants mention the target item types equally often. Similarly, differences in mentions of foreground and background items are more evident in the older age groups. The youngest children mention both foreground and background items equally often. Countable objects were mentioned more often than substances. Children seem to learn that foreground items are more relevant for discussion than background items.

Target items were mentioned significantly more often when a picture was in view than when a picture was not in view by the three youngest age groups. This would be expected in that attention may shift from one object to another more freely when the picture is in view. Participants were able to discuss relevant items and then search the scene for other items to be described. When a picture was not in view, participants were restricted by memory for what they had seen. There were no differences in the type of items that participants attended to for each task. Although participants' mentioned target items more often when the picture was in view, the patterns of mentioning items were similar across both tasks.

There was very little difference between the two language groups in mentioning target items. Although English speakers mentioned target items more often than Welsh speakers when the picture was in view, it is likely that this is due to social factors (e.g. Welsh speakers being less verbal) rather than being due to language difference per se. The data do not support the hypothesis that Welsh and English speakers differ with regard to attending to the c/u type. Both Welsh and English speakers attend to the each target item in a similar way, and speakers of both languages attended to sg/pl and c/u types more than substance type.

Picture selection

When selecting a picture most like the original English speakers chose the c/u variant pictures (but not the sg/pl or substance variant pictures) more often than the Welsh speakers. This indicates that English speakers were attending to c/u items less than Welsh speakers and were insensitive to *number* changes for this type. It was expected that English speakers would mainly select the substance type as most like the original because changes in *number* of substances are less relevant to the scene. It was expected that Welsh speakers would be treated c/u items like substances and therefore choose the c/u type more often than English speakers. The data contradict the hypothesis that Welsh speakers would be less sensitive to *number* changes for this type.

Again age was a key factor. Choice of substance variant increased with age while choice of sg/pl or cu variants decreased with age. Similarly, choice of variant with foreground *number* changes decreased with age. This supports the idea that as children become more experienced, they learn to attend to relevant items in a scene for discussion and consequently become insensitive to changes in number for less relevant items. Substance items and background items are often irrelevant to a scene and do not draw attention in the same way as countable items and foreground items. The data illustrate a developmental pattern that children learn to distinguish between these different types of items and the amount of attention that should be paid to them.

Correctly selecting the original, following a short or long interval, was largely dependent on age. Regardless of interval length, more correct selections, and fewer errors, were made with increasing age. However, more errors were made at each age with a long interval than with a short interval. This indicates that memory for items increases with age, and that memory for items is better following a short interval than a long interval. There were no differences across the two language groups for correctly selecting the original.

When errors were made (i.e. variant pictures were incorrectly chosen as the original) there were no differences across the two language groups. Both Welsh and English speakers overall displayed similar patterns in their choice of variant picture. Generally, participants did not favour a particular variant picture type. This was true for both the long and short interval tasks. This suggests that participants were equally sensitive to changes in *number* for each item type.

In the short interval task though, a trend showed that Welsh- and English-speaking 10-year-olds differed with regard to variant type choices and choice of variants with foreground and background changes. The Welsh 10-year-olds chose the c/u and substances more often than the sg/pl type, as was expected. Thus, Welsh 10-year-olds were less sensitive to changes in *number* for c/u and substance types. Surprisingly, the English 10-year-olds chose the sg/pl and c/u over the substance type, indicating that they were insensitive to changes in *number* for sg/pl and c/u type, contrary to what was expected. Likewise, the Welsh 10-year-olds chose variants with foreground changes twice as often as variants with background changes while the English 10-year-olds did the reverse.

Animacy

Some of the results, in particular those relating to the c/u type, were contrary to those expected. The findings suggest that the target items of the c/u type were particularly salient for both the Welsh and English speakers. It was observed that many c/u nouns in Welsh refer to animate objects. It was hypothesised that the results may reflect the salience of animacy rather than the target item type. It was noted that all substance type target items were inanimate objects while 3 out of 15 sg/pl type items and 6 out 15 c/u type items were animate. Further analysis was carried out to explore whether animacy influenced participants' responses.

Only the responses in tasks 1 and 2 were analysed for animacy. This was because (1) participants seemed to attend to the c/u items more in these tasks, and (2)

there were insufficient numbers of animate items in the remaining tasks. Mentions of target items in task 1 and 2 were coded as animate or inanimate regardless of target item type or task. A language x age x animacy ANOVA was performed and a significant main effect of animacy $F(1,141) = 216.30$, $p < .000$ was found.

The effect of animacy shows that animate items (M = .8379, SD = .2116) were mentioned significantly more often than inanimate (M = .5658, SD = .2079) items by speakers of both languages at each age. Thus, participants were attending to the animate items over the inanimate items. In task 1, all the c/u items were animate while none of the sg/pl items were animate. The finding that c/u items were mentioned more than sg/pl may reflect the fact that participants attended to animate items more than inanimate items rather than properties of the c/u category.

Summary

The structure of Welsh and English did not have a significant effect on the way Welsh and English speakers attended to items in a scene or their sensitivity to *number* changes. A developmental pattern emerged suggesting that children learn what items in a scene they need to attend to (e.g. countable items or items in the foreground) and become less sensitive to changes in *number* for items they do not attend to (e.g substances or items in the background). Participants consistently and robustly attended to foreground items over background items. This is particularly important in light of the fact that foreground/background effects may have significant implications for Lucy's findings. Another important finding of this study is the animacy effect. Animate objects seem to draw more attention than inanimate objects. This explains why participants mentioned the c/u items more than the other two types. Again this poses some questions on the findings of Lucy's study where he examined animate objects, implements and substances in a scene. He found that Yucatec and English speakers attended more to animate items than substance items. With regard to implements, English speakers attended to these items more than Yucatec speakers. Lucy argues that these findings reflect the fact that in both languages, the animate

category requires pluralisation but the substances do not, and, English pluralises implements while Yucatec does not. However, the finding that the animate items are attended to more than substances may reflect an animacy effect in that participants may be attending more to animate objects because they are animate rather than language influences such as requiring pluralisation.

General Discussion

There were three main objectives to this thesis:

(1) To use a cross-linguistic approach to explore the influence of language structure on cognitive processes and language acquisition.

(2) To develop empirically based knowledge of the acquisition of collection categories to contribute to a wider understanding of the acquisition of collective nouns and language acquisition in general.

(3) To examine the acquisition of collection categories across Welsh and English because the grammatical structure of these two languages differ in crucial ways with regard to collection categories and the way they encode *number* in general.

This chapter addresses these objectives by discussing the findings of the studies detailed in this thesis in relation to existing theories of linguistic relativity and language acquisition. First, the findings of the studies examining language differences between Welsh and English will be discussed. Second, the findings of the studies exploring cognitive influences of language structure will follow. Methodological limitations and the implications for future research will conclude the chapter.

Language differences

The two studies carried out to measure the differences between Welsh and English clearly demonstrate that Welsh differs from English with regard to collection categories. Although similarities were identified across the two languages with many shared or common patterns, crucially, the way each language refers to number is different.

The frequency and distributions of nouns in written texts and mothers' input were examined. Nouns from the two languages were classified into different noun types. The results show that nouns of the singular/plural (sg/pl) type make up

most of the noun types but very few collective nouns were found in either language. These findings suggest that both Welsh and English treat the referents of most nouns as countable entities with little emphasis on collections. Furthermore, the results indicate that more nouns occur as singular forms than as plural forms across the two languages, suggesting that there is greater emphasis on individuals than on multiple entities or collections.

However, the main disparity between the two languages involves the cluster/unit (c/u) noun types found only in Welsh and making up 2.5% of all nouns in Welsh. This c/u system operates in opposition to the sg/pl system. And the semantic nature of cluster nouns (more substance-like meaning) are distinct from collective nouns (meaning whole individual groups) and plural nouns (meaning many). Within the c/u system, the emphasis is on the collection, not the individuals. This is supported by the finding in written texts that c/u nouns occur as cluster forms (44.7%) almost as often as they occur as unit forms (55.3%). In comparison, the other noun types occur in forms that refer to multiple entities only 1.30% of the time in Welsh. Singular forms are more frequent in Welsh than in English. Recall that in Welsh numerals can occur with singular forms and unmarked nouns can refer to clusters. This means that Welsh may not have a clear distinction between singular and multiple reference. This may highlight 'collectiveness' for Welsh speakers, allowing them to think more readily about entities as collections.

Overall, the study of mothers' input supports the findings on written texts. As a semi-naturalistic study of mothers' speech to their infants in two play conditions, the results overall were similar to those found in study 1. The distributional patterns of different nouns types and the frequency of use were, for the most part, similar across written texts and mothers' input. The findings suggest that mothers' speech to children reflects the structure of their language.

However, two important findings emerged from the data on mothers' speech that did not support the findings on written texts. First, of the nouns used to refer to multiple referents in Welsh, fewer were cluster forms than was expected. Far more were plural forms. Second, when c/u nouns were used, most were used in their unit forms. Only 16.0% of the nouns used to refer to multiple referents in Welsh were cluster forms, surprisingly low when compared to 58.9% in written texts. Similarly, almost half of the c/u nouns (44.7%) occurred in their cluster forms in written texts, yet only 15.2% were cluster forms in the mothers' speech. Further analysis of the use of singular and unit forms revealed that Welsh mothers used a proportion of these forms with numerals. Thus, in the input to Welsh speaking children, a singular or unit form does not necessarily indicate reference to an individual. Instead, singular and unit forms co-occur with numerals to refer to several individuals. Welsh nouns are often non-specific with regard to number. This, together with the scarceness of plural and cluster forms, and the range of ways in which plurals are formed, means that the referent of a noun (i.e. reference to number) may not always be clear to a child learning Welsh.

Research has demonstrated a high correlation between frequency in language input and child language acquisition (Huttenlocher et al, 2002; Rowland & Pine, 2000; Mintz et al, 2002; Theakston et al, 2001; Naigles & Hoff- Ginsberg, 1998; Gathercole, Sebastian and Soto, 1999). Although Welsh-speaking children hear fewer cluster forms than was expected in the input, they are nevertheless exposed to nouns that refer to collections more than English-speaking children and may therefore acquire collective nouns more easily than English speakers.

Cognitive differences

What impact do these language differences have on the acquisition of collection categories? Previous research has suggested that learning collective nouns is difficult for children because they have not yet developed the cognitive complexity to understand their meanings. However, this research has focussed on English-speaking children and adults. Different languages have ways of encoding number that differ from English. As has been demonstrated in this thesis, Welsh clearly differs from English with regard to the way the two languages refer to number, in particular, collection categories. Although neither language has many collective nouns, Welsh does have the additional c/u nouns that also refer to collection categories. Welsh does not make clear distinctions in the number reference of nouns in the same way as English. Consequently, the difficulty in acquiring collective nouns may reflect the structure of English rather than the degree of difficulty in conceptualising collections. Children's difficulty in acquiring them could merely reflect the lack of sufficient exemplars in the input. Given that Welsh has the additional c/u system, Welsh speakers may be exposed to more examples of collection categories in their language and as a result may be more willing to conceptualise entities as collections.

Two studies were carried out to investigate the influence of the language differences on Welsh and English speakers' willingness to accept nouns as referring to collections. The first (Study 3) investigated the extent to which Welsh and English speakers would interpret novel nouns as referring to collections by examining categorisation and recognition behaviours. The second (Study 4) examined Welsh and English speakers' attention to and memory for number.

Study 3 addressed three main questions (1) Does language structure guide categorisation for language learning? (2) Do object properties and syntax have a role in categorising collections as whole individuals? And (3) Do the ways collections of

objects are categorised (and represented) influence memory for the individual items that make up the collection?

In answer to the first question, the findings of the overt categorisation behaviours do not support the notion that Welsh speakers would be more likely to interpret novel nouns as referring to collections than English speakers and indeed indicate little difference in the way speakers of the two languages categorised the novel objects. However, a robust and consistent finding that Welsh speakers took significantly longer than the English speakers to categorise items when items were presented in groups suggests that language structure does influence on-line language processing. This finding is not due primarily to Welsh speakers being bilingual, as was demonstrated in the follow-up study. Note that language, age, object properties and syntax had no influence on the speed of recognising individual items. Thus, the findings of the speed of categorisation are a consequence of the language demands of that task and not due to poorer motor skills or other extraneous factors.

These results offer some indication that language structure does influence categorisation for language learning. Indeed, Hunt and Banaji (1988) argue that subtle differences may be observed in processing times across speakers of different languages in the same conditions because of the variety of ways of construing the conditions linguistically. Hunt and Agnoli (1991) also argue that reaction time measures can be utilised to demonstrate cross-linguistic processing differences where languages can facilitate or hinder non-linguistic reaction times in speakers of different languages.

Hohenstein (2001) agrees and suggests that cognitive processing allows online thoughts to be measured rather than just the products of processing. Differences in non-linguistic cognition produced by language differences can be shown in implicit learning or speed of processing than in explicit responses to classification tasks. Thus, examining reaction times to stimuli, related to different

linguistic conditions, allows us to measure cognitive processing both within the individual and across groups. The results show that although Welsh and English speakers do not differ significantly in response choices, language structure does influence cognitive processing.

Numerous studies have been carried out to investigate the role of language structure on categorisation for a range of conceptual domains including spatial concepts (Choi and Bowerman, 1991; 1996; Bowerman and Choi; 1997; Choi, Mcdonough, Mendler, and Bowerman, 2000), motion concepts (Naigles & Hoff-Ginsberg, 1998) and object concepts (Lucy, 1992a; 1992b; Mazuka and Friedman, 2000). Little research has examined this issue with regard to collection concepts. The finding that Welsh speakers differ from English speakers with regard to processing collection categories demonstrates the influence of language structure on cognition of objects. These findings lend some support to the notion of linguistic relativity in that language structure does influence thinking.

Age also seems to have been an important factor in categorising collections. A general trend is seen across the two language groups: Participants were more likely to categorise novel objects as collections, and were faster at categorising with increasing age. It seems that interpreting novel nouns as referring to collections is influenced by age as was suggested by Bloom et al (1994; 1996).

In answer to the second question, the results demonstrate that both object properties and syntactic cues are important factors when interpreting novel nouns as collections. Participants were more likely to think of the objects as collections and took longer to categorise as the number of items and groups increased.

In addition, participants were more willing to adopt a collective interpretation and took longer to categorise when unmarked novel nouns were used. Unmarked syntax seems to offer more options about possible referents when interpreting new words. Bloom (1994; 1996) argues that singular syntax plays an

important role in acquiring collective nouns. However, the findings of Study 3 suggest that the unmarked forms are more crucial for acquiring collective nouns. This may be because unmarked nouns are more ambiguous with regard to their referents than singular forms, especially in Welsh. Blooms' observation may hold for English speakers because unmarked nouns are generally singular nouns in English. But in Welsh unmarked nouns do not necessarily have singular reference. Distinguishing between unmarked and singular syntax is important in language research, if theories are to account for the acquisition of languages other than English. These findings suggest that both object properties and syntax have an important role to play in language acquisition because they provide additional information that guide children in acquiring the meanings of new names.

There are proposals that children are biased or constrained to think in certain ways when acquiring new words. These include the whole object bias (Markman, 1994) and shape bias (Landau et al 1988). Given that children will interpret novel words as collections when the conditions are right (e.g. with increasing number of items and groups), does this mean that they are overriding constraints and biases? If so, what functions do constraints and biases have if children can think about objects in different ways depending on the conditions? Biases may arise as a consequence of the structure of the language being learned rather than being inherent in every child. The finding that interpreting novel nouns as collections is facilitated by object properties and syntax supports the notion suggested by Gathercole et al (1999) that language acquisition is a matter of children learning to co-ordinate multiple cues regarding the referents of nouns. Increasing exposure to language structure, through the input and with age seems to help guide children's categorisation and language acquisition, together with syntactic context and object properties.

In answer to question three, the findings of Study 3 suggest that the ways collections of objects are categorised (and represented) does influences memory for

the individuals that make up the collection. The findings indicate that the conditions for initial exposure to novel objects effects non-linguistic cognition. Participants' recognition of items was markedly influenced by the syntactic context and object properties of the stimuli when they were first encountered. Importantly, Welsh speakers demonstrated poorer recognition for objects that had been presented with unmarked and plural novel nouns than English speakers. It seems that the ability to recognise individual objects presented with unmarked and plural novel nouns was more difficult for Welsh speakers than English speakers because of the way speakers of the two languages represented those objects on first exposure. Given that Welsh speakers took significantly longer to categorise than the English speakers, it is possible that Welsh speakers were having more difficulty in deciding how to represent the objects and were less able to formulate robust representations of the items. Language differences seem to influence non-linguistic cognition because memory for individual items may be facilitated or hindered by the way these items are first represented[1].

Increased number of items on first exposure also seemed to interfere with the ability to accurately recall the items. Memory was better when fewer items were available to form representations. In contrast, increasing the number of groups available on first exposure facilitated the ability to recall the individual items. Memory was better when more groups were available on first exposure. One possible explanation for this is the way attention is distributed across groups when forming representations of the items. When two groups of items are available, attention may be distributed across the two groups with less attention given to individual items in each group. However, the cognitive demand of having several groups of items may mean that attention is drawn to one of the groups, allowing more attention to be given

[1] Increasing age also improves the ability to recognise the individual items, indicating that memory improves with age.

to the individuals. This would then allow stronger representations of items to be consolidated in memory thereby aiding recognition.

In summary, Study 3 underlines that many factors play a role in the categorisation of novel objects. Often, multiple cues are required when acquiring the meanings of new words. This, in turn, impacts on memory and recognition for those items. First, the structure of the native language has important influences on the way children acquire the meanings of new words. Second, the linguistic context of novel nouns has important influences on the way new nouns are interpreted. Finally, object properties such as number of items or groups also seem to play a role in the interpretation of novel nouns as collective or not.

To further examine the role of language structure on linguistic and non-linguistic cognition, Study 4 used both verbal descriptions and picture selection procedures. The aim was to investigate the role of different noun type categories on attention and memory. The verbal descriptions addressed attention to different noun types in Welsh and English, measured by mention of target items. The picture selection procedures addressed attention to *number* changes in short- and long-term memory without verbal behaviours.

The findings suggest that language structure had little effect on the way Welsh and English speakers attended to items in a scene or their sensitivity to *number* changes. There was very little difference between the two language groups in mentioning target items. (English speakers did mention target items more often than Welsh speakers when the picture was in view, but it was hypothesised that this is likely to be due to social factors rather than being an effect of language difference since it affected all three noun types.) Speakers of both languages attended to sg/pl and c/u items more than substance items, and attended to foreground items more than background items. Similar patterns also emerged across the two language groups for picture selections with no differences for correct or incorrect picture choices. If

anything, the data provide counter-evidence to the hypothesis that Welsh speakers would treat c/u items like substances and would be less sensitive to *number* changes for this type than English speakers: English speakers chose the c/u variant (but not the sg/pl or substance variant pictures) as most like the original more often than Welsh speakers did. (Thus, contrary to what was expected, English speakers were less sensitive to *number* changes for this type than Welsh speakers.).

Furthermore, c/u items were mentioned most by both Welsh and English speakers. This suggests that c/u items were particularly salient for speakers of both languages. This seems to be primarily due to an animacy effect. Several c/u items in the study were animate and it is likely that animate, objects draw attention more than inanimate objects. Although this finding confounds the data to some extent, it highlights the need to examine the effect of animacy on attention, and to control for such effects in future research.

Generally, participants attended to c/u and sg/pl items more than substances both when mentioning target items in picture descriptions and when selecting variant pictures as most like the original. This supports the findings of Lucy's study that less attention is given to substances than to countable items. However, when participants were asked to identify which picture they had seen before, no variant type was favoured over another.

Attention to and sensitivity to *number* changes of the different types was affected by age. A developmental pattern emerged suggesting that children learn with time to attend to countable items in a scene and to overlook substance type referents. While 4-year-olds did not mention items or select variants of a given type more than at chance level, by adulthood clear differences were detected.

Participants consistently and robustly attended to foreground items over background items and were less sensitive to changes in *number* for background items. (This brings into question the findings of Lucy's study where foreground/background

effects were not taken into consideration.) There is a clear developmental pattern in that attention to foreground items over background items progresses with age.

Naigles and Eisenberg (1998) point out that to demonstrate cognitive effects of language differences, it is necessary to

(1) show a developmental pattern by testing children before and after the acquisition of the linguistic devices in question,

(2) demonstrate that cognitive differences are due to linguistic as opposed to cultural differences, and

(3) illustrate that cognitive differences are related to language but use non-linguistic tasks to show that these differences are not solely related to language.

The findings of the studies reported in this thesis first demonstrate a developmental pattern in the acquisition of collective nouns. With increasing age, children are more willing to accept that novel nouns refer to collections and seem to attend more to *number* in picture scenes. Recall that Welsh 10-year-olds interpreted more nouns as collections than English 10-year-olds while the adults did the reverse. It is likely that collection categories are particularly relevant for Welsh speakers at this age. These differences across the two languages in the 10-year-olds support the idea that cognitive differences are largely due to language differences. Also note that a trend emerged indicating that Welsh- and English-speaking 10-year-olds differed with regard to variant type choices and choice of variants with foreground and background changes. Welsh 10-year-olds were less sensitive to changes in *number* for c/u and substance types, as was expected. The English 10-year-olds were less sensitive to the *number* changes for the sg/pl and c/u, contrary to what was expected. In addition, the Welsh 10-year-olds chose variants with foreground changes twice as often as variants with background changes while the English 10-year-olds did the reverse.

Second, the cognitive differences observed across the two language groups are not due to cultural or educational backgrounds because all the participants in the studies live in the same region of Wales and engage in the same educational and cultural systems. Third, although attention and sensitivity to *number* changes did not differ across the two languages, differences between the two language groups on the recognition task (i.e. another non-linguistic task) shows that the cognitive differences between Welsh and English speakers are not solely related to language. In sum, the findings in this thesis lends support to the notion that differences in the grammatical structure of Welsh and English does influence, to some extent, both linguistic and non-linguistic cognitive processes for the acquisition of collection categories.

Future research

Research on language acquisition is fraught with many methodological difficulties, not least the considerable variation of language use within and across languages. A key weakness of the research detailed in this thesis is that Welsh speaking participants are largely proficient English speakers. Welsh speakers live in a bilingual environment where English has a dominant place in the ambient language. On the other hand, problems associated with cultural and educational differences between the populations being studied, as suggested by Mazuka and Friedman (2000) and Naigles and Eisenberg (1998), are overcome in the studies in this thesis because Welsh- and English-speaking participants live in the same areas in Wales and engage in the same educational and cultural systems. Nevertheless, the influence of English on the acquisition of language may mask the developmental patterns expected for Welsh speakers. There is also the added complexity of language evolution. Ravid and Hayek (2003) suggest a trend of regularisation in Palestinian Arabic whereby speakers tend to use a plural form instead of a collective form. This may also be true in Welsh. Although anecdotal, it is often observed that Welsh speakers use the

English plural -s with many Welsh nouns to indicate plural reference. Some Welsh children also use Welsh plural inflections inappropriately on cluster nouns (e.g. *moch-od* 'pigs-Pl' instead of *moch* 'pigs'). And unit inflections are often used on singular nouns (e.g. *plat-en*, 'plate-Sg', instead of *plât* 'plate'), which may be indicative of a need to provide a stronger singular reference given the absence of an indefinite article in Welsh. Given the complexity of the number marking system in Welsh, and the knowledge that Welsh speakers have of English, it is not surprising that Welsh speakers may be regularising the number system. Moreover, the classification of participants into the language groups, for the studies contained in this thesis, were based on self-reports. Thus, classification is subjective and allows for considerable variation in the sample populations. To fully understand the influence of language structure on language acquisition it is important that further research be done on the effects of bilingual environments on language ontogeny and evolution. To achieve this, it may be necessary to establish standardised language production and comprehension tasks, in particular for Welsh speakers, that would allow more rigorous classification of participants into appropriate language groups to be studied.

Another interesting aspect of the Welsh number system that requires further investigation is that of the partitive system. The fact that Welsh speakers can use numerals directly with singular or unit forms, and that plural and cluster forms are used with the partitive *o* 'of', means that nouns do not necessarily carry all the information needed for number reference. Likewise, singular-only nouns (e.g. *bara* 'bread') occur with the partitive *o* 'of', which means that mass and count properties of nouns are not as clear-cut in Welsh as they are in English. Further research is needed to understand the way different noun types in Welsh are conceptualised, individuated and quantified.

Although Welsh has far more nouns that refer to collections than English, this does not necessarily mean that the collection categories to which the nouns refer

are the same. In English, only collective nouns refer to collection categories but in Welsh, there are two types of nouns (collective nouns and cluster nouns) that refer to collection categories. It may be that in Welsh, the two types refer to different kinds of collection categories with different semantic properties. As previously mentioned, cluster nouns refer to entities that have a more substance-like meaning while collective nouns refer to entities consisting of discrete individuals. Consequently, concepts of collection may vary not only across languages but also within a given language. The idea that collection categories may not be universal across languages is also supported by Ravid and Hayek's (2003) study on Palestinian Arabic. The collective nouns in Arabic do not correspond to collective nouns in English but instead have a sense of homogenousness in a similar way to cluster nouns in Welsh. It is important that further research be done to fully understand the different kinds of collection categories and the ways in which these categories are formed.

References

Ankrum, C. and Palmer, J. (1991). Memory for objects and parts. *Perception and Psychophysics, 50 (2),* 141-156.

Aronoff, M. (1994). *Morphology by itself*. London: The MIT Press.

Baker, C. (1985). *Aspects of Bilingualism in Wales.* Clevedon, Avon: Multilingual Matters.

Berlin, B. and Kay, P. (1969). *Basic color terms: Their universality and evolution.* Berkeley: University of California Press.

Berlin, B., Breedlove, D. E. and Raven, P. H. (1974). *Principals of Tzeltal plant classification.* New York: Academic.

Bloom, P. (1994). Possible names - the role of syntax-semantics mappings in the acquisition of nominals. *Lingua, 92 (1-4),* 297-329.

Bloom, P. (1996). Possible individuals in language and cognition. *Current Directions in Psychological Sciences, 5,* 90-94.

Bloom, P. (1998*). How children learn the meaning of words.* School of Psychology seminar, University of Wales, Bangor, Thursday 26th Feb.

Bloom, P. (2001a). Precis of how children learn the meaning of words. Behavioural and Brain Sciences, 24(6), 1095-1123.

Bloom, P. (2001b). Roots of word learning. In M. Bowerman and S. C. Levinson (Eds.), *Language acquisition and conceptual development.* Cambridge: Cambridge University Press.

Bloom, P. (2001c). Controversies in the study of word learning – response. Brehaviour and Brain Science. 24(6), 1124-1134.

Bloom, P. and Kelemen, D. (1995). Syntactic cues in the acquisition of collective nouns. *Cognition, 56,* 90-94.

Bloom, P., Kelemen, D., Fountain, A., and Courtney, E. (1995). The acquisition of collective nouns. In D. MacLaughlin and S. McEwan (Eds). *Proceedings of the 19th Boston University Conference on Language Development*: Cascadilla Press.

Bloom, P. and Markson, L. (1998). Capacities underlying word learning. *Trends in Cognitive Science, 2, 2*, 67-73.

Bloom, P. and Veres, C. (1999). The perceived intentionality of groups. *Cognition, 71*, B1-B9.

Bloom, P. and Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language, 24 (3)*, 511-33.

Bloomfield, L. (1933). *Language*. London: George Allen and Unwin.

Bowerman, M. (1996a). Learning how to structure space for language: A crosslinguistic perspective. In P. Bloom, M. Peterson, L. Nadel, and M. Garrett (Eds.), *Language and space* (pp.385-436). Cambridge, MA: MIT Press.

Bowerman, M. (1996b). The origins of children's spatial semantic categories: cognitive versus linguistic determinants. In J.J. Gumperz and S. C. Levinson (Eds). *Rethinking Linguistic Relativity*. Cambridge, MA: MIT Press.

Bowerman, M. (2000). Where do children's word meanings come from? Rethinking the role of cognition in early semantic development. In L. P. Nucci, G. Saxe, and E. Turiel (Eds.), *Culture, Thought, and Development* (pp.199-230). Mahwah, NJ: Lawrence Erlbaum Associates.

Bowerman, M. and Choi, S. (2001). Shaping meanings for language: universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman and S. C. Levinson (Eds), *Language acquisition and conceptual development*. Cambridge: Cambridge University Press.

Brown, R. (1958). How shall things be Called? *Psychological Review, 65*: 14-21.

Choi, S. (1997). Language-specific input and early semantic development: Evidence from children learning Korean. In D. I. Slobin (Ed). *The crosslinguistic study of language acquisition: Vol. 5. Expanding the contexts* (pp.41-133). Mahwah, NJ: Lawrence Erlbaum Associates.

Choi, S. and Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition, 41*, 83-121.

Choi, S., McDonough, L., Bowerman, M., and Mandler, J. M. (1999). Early sensitivity to language-specific spatial categories in English and Korean. *Cognitive Development, 14*, 241-268.

Coley, J. D. and Medin, D. L. (1997). Does rank have its privilege? Inductive inferences within folkbiological taxonomies. *Cognition, 64*, 73-112.

Crago. M. B., Allen, S. E. M., and Pesco, D. (1998). Issues of complexity in Inuktitut and English child directed speech. In E. V. Clark (Ed). The Proceedings of the Twenty-ninth Annual Child Language Research Forum. USA: Stanford Linguistic Association.

Craig, C. (1986). *Noun classes and categorisation.* Amsterdam: John Benjamin's Publishing Company.

Croft, W. (2000). Parts of speech as language universals and as language-particular categories. In Petra Maria Vogel and Bernard Comrie, (Eds*) Approaches to the typology of word classes.* Berlin: Mouton de Gruyter.

Croft, W. (2001a). *Radical construction grammar: Syntactic theory in typological perspective.* Oxford: Oxford University Press.

Croft, W. (2001b). Some contributions of typology to cognitive linguistics and vice versa. In Theo Janssen and Gisela Redeker, (Eds*) Foundations and scope of cognitive linguistics.* Berlin: Mouton de Gruyer.

Crystal, D. (1997). *A dictionary of linguistics and phonetics.* Oxford: Blackwell Publishers.

Davidoff, J. (2001). Language and perceptual categorisation. *Trends in Cognitive Sciences, 5, 9,* 382-387.

Davies, J. (1999). *The Welsh language: A pocket guide.* Cardiff: University of Wales Press and The Western Mail.

Dean, A. L., Chabaud, S., and Bridges, E. (1981). Classes, collections, and distinctive features: Alternative strategies for solving inclusion problems. *Cognitive Psychology, 13,* 84-112.

Deuchar, M. (In press). Minority language survival in Northwest Wales: An introduction. *Proceedings of the 4th International Symposium on Bilingualism.*

Diesendruck, G. and Bloom, P. (2003). How specific is the shape bias? *Child Development, 74 (1),* 168-78.

Ellis, N. C., O'Dochartaigh, C., Hicks, W., Morgan, M., and Laporte, N. (2001). Cronfa Electroneg o Gymraeg/ *A 1 million word lexical database and frequency count for Welsh.* University of Wales, Bangor, School of Psychology.

Francis, N. W. and Kucera, H. (1982) *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Muffin Company.

Gathercole, V. C. and Min, H. (1997). Word meaning biases or language specific effects? Evidence from English, Spanish and Korean. *First Language, 17,* 31-56.

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

Gathercole, V. C. M. (1997b). Word meaning biases or language-specific effects? Evidence from English, Spanish and Korean. *First Language, 17,* 31-56.

Gathercole, V. C. Mueller, Thomas, E. M. and Laporte, N. I. (2001). The

acquisition of grammatical gender in Welsh. *Journal of Celtic Language Learning, 6,* 53-87.

Gathercole, V. C. Mueller, Sebastian, E., and Soto, P. (1999). The early acquisition of Spanish morphology: Across-the-board or piecemeal knowledge? *The International Journal of Bilingualism, ,3 (2-3),* 133-182.

Gathercole, V. C., Thomas, E. M. and Evan, D. (1999). What is a noun? Welsh-, Englsih, and Spanish-speaking children see it differently. *First Language, 20,* 55-90.

Gathercole, V. C., Thomas, E. M. and Kim, S. (1999). Cue Coordination: An alternative to word meaning biases. In B. A. Fox, D. Jurafsky and L. A. Michaelis (Eds). *Cognition and Function in Language.* California. CSLI Publications.

Gleason, H. A. (1955). *An introduction to descriptive linguistic.* New York: Holt, Rinehart and Winston.

Golinkoff, R.M., Mervis, C.B. and Hirsh-Pasek, K. (1994). Early object labels - the case for a developmental lexical principles framework, *Journal of Child Language, Vol. 21, Iss. 1,* 125-155.

Gumperz, J. J. and Levinson S. C. (1996). *Rethinking Linguistic Relativity.* Cambridge, MA: MIT Press.

Hall, D. G. and Graham, S. A. (1999). Lexical form class information guides word-to-object mappings in preschoolers. *Child Development, 70, 1,* 78-91.

Hayes, R. A., Slater, A. and Brown, E. (2001). Infants' ability to categorise on the basis of rhyme. *Cognitive Development, 15 (4),* 405-419.

Heider, E. R. (1971). 'Focal' color areas and the development of color names. *Developmental Psychology 4:* 447-55.

Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology, 93,* 10-20.

Hohenstein, J. M. (2001). Motion event similarities in English- and

Spanish-speaking children. Dissertation Abstracts International: Section B: *The Sciences and Enginieering, 62 (3-B)*,1616.

Hunt E. and Banaji, M. (1988). The Whorfian hypothesis revisited: A cognitive science view of linguistic and cultural effects on thought. In J. Berry, S. Irvine, and E. Hunt (Eds.). *Indigenous cognition: Functioning in cultural context.* Dordecht, The Netherlands: Martinus Nijhoff.

Hunt, E. and Agnoli, F. (1991). The Whorfian hypothesis: A cognitive psychology perspective. Psy*chological Review, 98*, 377-389.

Huntley-Fenner, G. (1995). The effect of whole object bias on preschoolers' understanding of collective nouns. *Proceedings of the 27th Annual Child Language Research Forum*, 145-155.

Huttenlocher, J. Vasilyeva, M., Cymerman, E., and Levine, S. (2002). Language input and child syntax, *Cognitive Psychology.* 45(3), 337-374.

Imai, M. (1999). Constraint on word-learning constraints. *Japanese Psychological Research, 41 (1),* 5-20.

Imai, M. and Gentner, D. (1997). A cross-linguistic study of early word meaning: universal ontology and linguistic influence. *Cognition, 62*, 169-200.

Jackobson, R. (1968). *Child language, Aphasia and Linguistic Universals.* The Hague: Mouton.

Jensen, J.T. (1990). Morphology: Word structure in generative grammar. *Current Issues in Linguistic Theory, 70.* Amsterdam: John Benjamins Publishing Company.

Kay, P. (1996). Intra-speaker relativity. In J. J. Gumperz, and S. C. Levinson (Eds). *Rethinking Linguistic Relativity.* Cambridge, MA: MIT Press.

King, G. (1993). *Modern Welsh. A comprehensive grammar.* London: Routledge.

Lakoff, G. (1986). Classifiers as a reflection of mind. In C. Craig (Ed)

Noun Classes and Categorisation. Amsterdam: John Benjamins Publishing Company.

Lakoff, G. (1987). *Women, fire and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Landau, B., Smith, L. B. and Jones, S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3,* 299-321.

Leach, E. (1964). Anthropological aspects of language: Animal categories and verbal abuse. In E. H. Lenneberg (Ed). *New directions in the study of language,* 23-63. Cambridge, Mass.: MIT Press.

Lucy, J. A. (1992a). *Grammatical categories and cognition.* Studies in the Social and Cultural Foundations of Language, No.13, Cambridge: Cambridge University Press.

Lucy, J. A. (1992B). *Language, thought and diversity. A reformulation of the linguistic relativity hypothesis*, Cambridge: Cambridge University Press.

Lucy, J. A. (1996). The scope of linguistic relativity: An analysis and review of empirical research. In J. J. Gumperz and S.C. Levinson (Eds). *Rethinking Linguistic Relativity*, Cambridge: Cambridge University Press.

Lyons, J. (1968). *Introduction to theoretical linguistics.* Cambridge: Cambridge University Press.

Markman, E. M. (1978). Empirical versus logical solutions to part-whole comparison problems concerning classes and collections. *Child Development, 49,* 168-177.

Markman, E. M. (1994). Constraints on word meaning in early language acquisition. *Lingua, 92,* (1-4), 199-227.

Markman, E. M. and Siebert, J. (1976). Classes and collections: Internal organization and resulting holistic properties. *Cognitive Psychology, 8,* 561-577.

Markman, E. M., Horton, M. S. and McLanahan, A. G. (1986). Classes and collections: Principles in the learning of hierarchical relation. *Cognition, 8,* 227-

241.

Martinez, I. V. and Shantz, M. (1996). Linguistic influence on categorisation in preschool children: A crosslinguistic study. *Journal of Child Language, 23*, 529-545.

Mazuka, R and Friedman, R. S. (2000). Linguistic relativity in Japanese and English: Is language the primary determinant in object classification. *Journal of East Asian Linguistics, 9*, 353-377.

Mintz, T. H., Newport, E. L., and Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science, 26 (4)*, 393-424.

Moravicsik, E. A. and Wirth, J. (1983). Markedness - an overview. In E. A. Moravicsik, and J. Wirth, (Eds). *Markedness*. London: Plenum Press.

Murphy, G. L. and Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92 (3)*, 289-316.

Naigles, L. R, and Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language, 25(1)*, 95-120.

Quinn, P. C., and Johnson, M. H. (1997). The emergence of perceptual category representations in young infants: a connectionist analysis. *Journal of Experimental Child Psychology, 66*, 236-263.

Ravid, D. and Hayek, L. (2003). Learning about different ways of expressing number in the development of Palestinian Arabic, *First Language, 23; 1 (67)*, 41-64.

L. M. Reeves, K. Hirsch- Pasek and R. Golinkoff (1998). Words and meaning: From primitives to complex organisation. Chp 4. In J. B., Gleason and N. B., Ratner, (Eds). *Psycholonguistics*. 2nd Ed. Fort Worth: Harcourt Brace College Publishers.

Rohde, D. L. T., and Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition, 27 (1)*, 67-109.

Rosch, E. and Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology, 7*: 573 - 605.

Rosch, E. (1973). Natural categories. *Cognitive Psychology, 4*. 328 - 50.

Rosch, E. (1978). Principles of categorisation. In Rosch, E. and Lloyd, B. B. (Eds), *Cognition and Caegorization*, 27 - 48. Hillside: Lawrence Erlbaum.

Rowland, C. F. and Pine, J. M. (2000). Subject auxiliary inversion errors and wh- question acquisition: "What children do know*!*", *Journal of Child Language, 27 (1)*, 137-181.

Sapir, E. (1949*). Language: An introduction to the study of speech*. New York: Harcourt, Brace and Company.

Sera, M., Berge, D., and Pintado, J. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development.*

Slobin, D. I. (1996). From "thoughts and language" to "thinking for speaking". In J. J. Gumperz, and S. C. Levinson (Eds). *Rethinking Linguistic Relativity*. Cambridge, MA: MIT Press.

Smith, L. B. (1995). Self-organizing processes in learning to learn words: Development is not induction. *Minnesota Symposium on Child Psychology*, vol 28: Basic and applied perspectives on learning, cognition and development. Mahawah, NJ: Lawrence Erlbaum.

Smith, L. B. and Rizzo, T. A. (1982). Childrens' understanding of the referential properties of collective and class nouns. *Child Development, 53*, 245-257.

Spencer, A. (1991). *Morphological theory*. Oxford: Basil Blackwell, Ltd.

Sproson, H. (2001). *The use of collective nouns in speech to children in English*. Unpublished Final Year Dissertation, University of Wales, Bangor.

Taylor, J. R. (1995). *Linguistic categorisation: Prototypes in linguistic theory*. 2nd Ed. Oxford: Clarendon Press.

Theakston, A., Lieven, E., Pine, J. M., and Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language, 28(1)*, 127-152.

Theakston, A., Lieven, E., Pine, J. M., and Rowland, C. F. (2002). Going, going, gone: The acquisition of the verb "go". *Journal of Child Language, 29 (4)*, 783-811.

Thomas, E. M. (2001). *Aspects of gender mutation in Welsh*. Unpublished Doctoral Thesis, University of Wales, Bangor.

Thomas, E. M. and Gathercole, V. C. M. (In press). Minority language survival: Obsolescence or survival for Welsh in the face of English dominance. *Proceedings of the 4th International Symposium on Bilingualism*.

Thomas, P. W. (1996). *Gramadeg y gymraeg*. Caerdydd: Gwasg Prifysgol Cymru.

Thorne, D. A. (1993). *A comprehensive Welsh grammar*. Oxford: Blackwell Publishers.

Ungerer, F. and Schmid, H-J. (1996). *An introduction to cognitive linguistics*. London: Longman.

Waxman, S. R. and Hall, D. G. (1993) The development of a linkage between count nouns and object categories - evidence from 15-month olds to 21-month olds. *Child Development, Vol. 64, Iss. 4*, 1224-1241.

Waxman, S. R. and Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29*, 257-302.

Whorf, B. (1956). *Language, thought and reality.* Cambridge, MA: MIT Press.

Wierzbicka, A. (1985). *Lexicography and conceptual analysis.* Ann Arbor: Karoma.

Wierzbicka, A. (1988). *The semantics of grammar.* Vol.18. Amsterdam: John Benjimins Publishing Company.

Wijnen, F., Kempen, M. and Gillis, S. (2001). Root infinitives in Dutch early child language: An effect of input? *Journal of Child Language, 28(3),* 629-660.

Williams, S. J. (1980). *A Welsh grammar.* Cardiff: University of Wales Press, Cardiff.

Wittgenstein, L. (1953). *Philosophical investigations.* New York: Macmillan.

Wurzel, W. (1984). *Flexionsmorphologie und naturlichkeit.* Berlin: Akademie Verlag; translated as Inflectional morphology and Naturalness (1989). Dordrecht: reidel.

Yont, K. M., Snow, C. E., and Vernon-Feagans, L. (2002). The role of context in mother-child interactions: and analysis of communicative intents expressed during toy play and book reading with 12-month-olds. *Journal of pragmatics*, 35 (3), 435-454.

Welsh number morphology

I) Singular/Plural (sg/pl) System

The sg./pl system in Welsh employs the following operations:

a) Adding a plural ending: although there are twelve different plural endings in Welsh their morphosyntactic operations can be equated to the English plural -s.

		Examples	
-au	/ai̯/	afal > afalau	"apple/s"
-iau	/iai̯/	esgid > esgidiau	"shoe/s"
-ion	/iɔn/	ysgol > ysgolion	"school/s"
-on	/ɔn/	modrwy > modrwyon	"ring/s"
-edd	/ɛð/	bys > bysedd	"finger/s"
-oedd	/ɔɛð/	mynydd > mynyddoed	"mountain/s"
-ed	/ɛd/	merch > merched	"girl/s"
-aint	/ajnt/	gof > gofaint	"blacksmith/s"
-od	/ɔd/	cath > cathod	"cat/s"
-iaid	/jajd/	creadur > creaduriaid	"creature/s"
-i	/i/	bisged > bisgedi	"biscuit/s"
-ydd	/ɨð/	diod > diodydd	"drink/s"
-i or **-ydd**	/i/ or /ɨð/	tref > trefi/trefydd	"town/s"

b) Vowel changes: again there are several different vowel changes that can occur in Welsh their operations can be equated to vowel changes in irregular English plural nouns (e.g. *man/men*) although much more frequent.

		Examples	
a > ai	/a > aj/	llygad > llygaid	"eye/s"
a > ei	/a > ɛi/	car > ceir	"car/s"

a > **y** /a > ə/	bustach > bustych	"ox/en"
	*alarch >elyrch	"swan/s"
ae > **ai** /aɛ > ai/	draen > drain	"thorn/s"
e > **y** /ɛ > ɨ/	bachgen > bechgyn	"boy/s"
e > **ai** /ɛ > ai/	dyniawed > dyniewaid	"injury/s"
o > **y** /ɔ > ɨ/	ffordd > ffyrdd	"road/s"
w > **y** /ʊ > ɨ/	*asgwrn > esgyrn	"bone/s"
oe > **wy** /ɔɛ > ʊɨ/	croen > crwyn	"skin/s"
y > **ai** /ɨ > ai/	ty > tai	"house/s"
	*note: penultimate a > e	

c) Adding plural ending with vowel change (small class of nouns) : no English equivalent.

Examples

a > **ei** /a> ɛi/	mab > meibion	"son/s"
a > **e** /a> ɛ/	gardd > gerddi	"garden/s"
e > **ei** /ɛ > ɛi/	gefell > gefeilliaid	"twin/s"
ae > **ei** /aɛ > ɛi/	saer > seiri	"carpenter/s"
ae > **ey** /aɛ > ɛɨ/	maes > meysydd	"field"
aw > **ew** /aʊ > ɛʊ/	cawr > cewri	"giant/s"
ai > **ei** /ai > ɛi/	iaith > iaithoedd	"language/s"
au > **eu** /aɨ > ɛɨ/	haul > heuliau	"sun/s"
aw > **o** / aʊ > ɔ/	awr > oriau	"hour/s"
w > **y** /ʊ > ə/	cwestion > cwestiynau	"question/s"
uw > **u** /ɨʊ > ɨ/	buwch > buchod	"stag/s"

au > **aw** /aɨ> aʊ/	cenau > cenawon	"whelp"
au > **af** /aɨ > av/	edau > edafedd	"thread/s"
ai > **a** /ai > a/	gwraig > gwragoedd	"wife"
ai > **ae** /ai > ɛɨ/	Sais > Saeson	"Englishman/men"
y > **y** / ɨ > ə/	llyn > llynnoedd	"lake/s"

d) Irregular plural nouns:

Examples

ci > cwn "dog/s"

or derived from its singular form:

Examples

dosbarth > dosbarthiadau "class/es".

(e) Other forms (small class of nouns): The roots of these nouns are bound forms and only occur with either a unit or plural suffix.

Examples

unigol > unigolion "individual/s"

cwningen > cwningod "rabbit/s"

In addition the roots of some bound forms have a different vowels associated with unit or plural ending.

Examples

cerdyn > cardiau "card/s"

An extension of the bound form class is that nouns denoting person and profession end with either -**wr** /ʊr/ or -**ydd** /ɨð/ for individuals and -**wyr** /ʊir/ for many.

II) Cluster/Unit System

In the c/u system nouns that refer to collections are the base forms and unit nouns are derived from them by using the following operations:

a) Adding a unit ending:

	Examples	
	plu > pluen	"feathers > a feather"
	moch> mochyn	"pigs > a pig"
	ser > seren	"stars > a star"

b) Adding a unit ending with a vowel change:

		Examples	
a > ai	/a > ai/	hwyaden > hwyaid	'duck/s'
a > ei or **y**	/a > ɛi or ə/	tywarchen > tyweirch > tyeyrch	'turf/s'
ei > ai	/ɛi > ai/	deilen > dail	'leaf/leaves'
o > y	/ɔ > ɨ/	cortyn > cyrt	'cord/s'
o > aw	/ɔ > aʊ/	conyn > cawn	'stalk/s'
eu > au	/eɨ > aɨ/	blodeuyn > blodau	'flower/s'
y > y	/ə > ɨ/	gwenynen > gwenyn	'bee/s'
e > a	/ɛ > a/	plentyn > plant	'child/ren'
ew > au	/eʊ > aʊ/	llysewyn > llysau	'herb'
y > w	/ə > ʊ/	cacynen > cacwn	'hornet'

Summary table of the semantic and syntactic characteristics of Welsh and English

Conceptual continuum	Syntactic category	Semantic properties	Syntactic properties	Examples	
				Welsh	English
Individual	Count Nouns	1 individual vs Many individuals All category levels (the label for the 'one' is at the same category level as for the label for the 'many') Subset relation – same word e.g.{many {one}} Singular forms function as 'individuals' occurring with singular verbs and plural forms function as 'many' occurring with plural verbs	Base form are singular and marked forms are plural A _____ / *Y* _____ *6*_____ _____s / _____*oedd* many _____s / *llawer o* _____*oedd* few _____s / *ychydig o* _____*oedd* some _____s / *rhai* _____*oedd* Verb agreement required (e.g. 'this **is** a book' and 'these **are** book**s**')	Bwrdd/byrddau Cwpan/au	Table/s Cup/s
Group	Collective Nouns (A) (B)	1 collection vs many collections Superordinate level category for 'whole' and basic level category for 'parts' Part/whole relation part 1 + part 2 + part 3… = whole	Base forms are singular and marked forms are plural A _____ / *Y* _____ *6*_____ _____s / _____*oedd* many _____s / *llawer o* _____*oedd* few _____s / *ychydig o* _____*oedd* some _____s / *rhai* _____*oedd* Singular forms can occur with singular or plural pronouns and singular or plural verbs. Verb agreement required for plural forms. Singular forms occur with singular pronouns and singular verbs.	Tim/au Teulu/oedd Byddin/oedd Cynilleidfa/oedd Coedwig/oedd	Team/s Family/ies Army/ies Audience/s Forest/s
	Measure Nouns	Modify/quantify nouns Singular forms function as 'individuals' occurring with singular verbs and plural forms function as 'group' occurring with plural verbs	Base forms are singular and marked forms are plural A _____ of_____ / *Y* _____ *o* _____ *6*_____ _____s / _____*oedd o* _____ many _____s / *llawer o* _____*oedd o* _____ few _____s / *ychydig o* _____*oedd o* _____ some _____s / *rhai* _____*oedd o* _____ Verb agreement required ('this **is** a bunch of flowers' and 'these **are** bunches of flowers')	Set/iau Grwp/iau Torfa/oedd	Set/s Group/s Crowd/s Bundle/s Bunch/es
	Welsh Collectives	Collection vs unit (many individuals vs 1 individual) Basic and subordinate category names ('unit' names are at the same category level as 'collection' names) Subset relation – same word {{collection} unit} Collections function as 'groups', used with singular verbs but plural pronoun. Units function as 'individual', used with singular verbs	Base forms are cluster (collections) and marked forms are unit No plural forms. *Y* _____ (unit/cluster form) *6*_____ (unit form) *6 of* _____(cluster) *llawer o*_____ (cluster) *ychydig o* _____ (cluster) Verb agreement required. ('llygoden ydi **hwn**' [this **is** a mouse] and 'llygod ydi **rhain**' [these **are** mice)	Llygod/en [mice/mouse] Pysgod/yn [Fish/ a fish]	
		Collections	Base forms are collection with no marked/unit form Occur with plural verbs and plural pronoun. * No singular/ unit form	Gwartheg [cows]	Cattle
Substance	Mass Nouns	Substance like Function as singular and is used with singular verb	Base forms are singular, no marked/plural forms. much _____ *Llawer o* _____ some _____ little _____ *Ychydig o* _____ Welsh adopts a partitive system to quantify mass nouns that can also be used when quantifying count nouns.	Dodrefn Pren	Furniture Wood

English Adult Questionnaire

Questionnaire

Please provide the following information.

1. Name: _____

2. Birth date:_____

3. Place of birth: _____

4. Have you always resided in North Wales?

YES____ NO____

If you were born outside of North Wales, or spent some time living in another area, where did you live and at what age did you come to North Wales?

Place:_____

Age:_____

5. Your first language is _____

6. You speak (please tick all that apply):

English_____

I began speaking English at age: _____

At present, I speak English approximately

____% of the time.

Welsh_____

I began speaking Welsh at age: _____

At present, I speak Welsh approximately

____% of the time.

Other_____(please

specify:_____)

I began speaking this language at age: _____

At present, I speak this language

approximately ____% of the time.

7. Language/s currently spoken at home (please tick one):

___100% English

___ 80% English, 20% Welsh

___ 60% English, 40% Welsh

___ 50% English, 50% Welsh

___ 60% Welsh, 40 % English

___ 80% Welsh, 20% English

___100% Welsh

___ Other language / ratio, please specify

_____

8. Language/s spoken at home while I was growing up (please tick one):

___100% English

___ 80% English, 20% Welsh

___ 60% English, 40% Welsh

___ 50% English, 50% Welsh

___ 60% Welsh, 40 % English

___ 80% Welsh, 20% English

___100% Welsh

___ Other language / ratio, please specify

_____

9. Language/s spoken at Work (please tick one):

___100% English

___ 80% English, 20% Welsh

___ 60% English, 40% Welsh

___ 50% English, 50% Welsh

___ 60% Welsh, 40 % English

___ 80% Welsh, 20% English

___100% Welsh

___ Other Language / ratio, please specify

_____

10. Languages spoken by my family members:

Mother

First language is _____ which she began to speak at age:_____ and uses approximately _____% of the time.

Other language/s (please specify)_____

_____

Father

First language is _____ which he began to speak at age:_____ and uses approximately _____% of the time.

Other language/s (please specify)_____

_____

Partner

First language is _____ which s/he began to speak at age:_____ and uses approximately _____% of the time.

Other language/s (please specify)_____

_____

Others

On average most of my friends, colleagues brother/s and/or sister/s speak _____ which they use approximately _____% of the time.

11. Languages spoken by family members to yourself:

Mother

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

Father

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

Partner

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

Others

On average most of my friends, colleagues, brothers and/or sisters use

English____approximately ___% of the time,

Welsh____ approximately ___% of the time,

Other____(please specify_____)

approximately ____% of the time when speaking to me.

12. Language/s spoken, on average, by you to:

A) family members

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

B) Others (e.g. friends, colleagues):

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

Thank you for completing this questionnaire and for participating in this study.

English Child Questionnaire

Questionnaire

Please provide the following information.

1. Child's name: _____

2. Child's birth date:_____

3. Child's place of birth: _____

4. Has your child always resided in North Wales? YES____ NO____

If your child was born outside of North Wales, or spent some time living in another area, where did s/he live and at what age did s/he come to North Wales?

Place:_____

Age:_____

5. Child's first language is _____

6. Child speaks (please tick all that apply):

English_____

S/he began speaking English at age: _____

At present, speaks English approximately ____% of the time.

Welsh_____

S/he began speaking Welsh at age: _____

At present, speaks Welsh approximately ____% of the time.

Other_____

S/he began speaking this language at age: _____

At present, speaks this language approximately ____% of the time.

7. Language/s spoken by all members of the family at home (please tick one)

___100% English

___ 80% English, 20% Welsh

___ 60% English, 40% Welsh

___ 50% English, 50% Welsh

___ 60% Welsh, 40 % English

___ 80% Welsh, 20% English

___100% Welsh

___ Other ratio, please specify _____

8. Language/s spoken at the child's school (please tick one)

___100% English

___ 80% English, 20% Welsh

___ 60% English, 40% Welsh

___ 50% English, 50% Welsh

___ 60% Welsh, 40 % English

___ 80% Welsh, 20% English

___100% Welsh

___ Other ratio, please specify _____

9. Languages spoken by family members (please complete those applicable):

Mother/Primary Care Giver

First language is _____ which I began to speak at age:_____ and use approximately _____% of the time.

Other language/s (please specify)_____

_____

Father

First language is _____ which he began to speak at age:_____ and uses approximately _____% of the time.

Other language/s (please specify)_____

_____

Siblings

A) *Brother* or *sister* (please circle one)

First language is _____ which s/he began to speak at age:_____ and use approximately _____% of the time.

Other language/s (please specify)

_____ _____

B) Brother or *sister* (please circle one)

First language is _____ which s/he began to speak at age:_____ and use approximately _____% of the time.

Other language/s (please specify)_____

_____

C) *Brother* or *sister* (please circle one)

First language is _____ which s/he began to speak at age:_____ and use approximately _____% of the time. Other language/s (please specify)_____

_____

D) *Brother* or *sister* (please circle one)

First language is _____ which s/he began to speak at age:_____ and use approximately _____% of the time. Other language/s (please specify)_____

_____

Grandparents

A) *Mother's mother*

First language is _____ which she began to speak at age:_____ and use approximately _____% of the time.

Other language/s (please specify)

_____

_____

B) *Mother's father*

First language is _____ which he began to speak at age:_____ and use approximately _____% of the time.

C) *Father's mother*

First language is _____ which she began to speak at age:_____ and use approximately _____% of the time.

D) *Father's father*

First language is _____ which he began to speak at age:_____ and use approximately _____% of the time.

10. Languages spoken by family members to the child:

Mother/Primary Care Giver

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

Father

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

Siblings

A) *Brother* or *sister* (please circle one)

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

B) *Brother* or *sister* (please circle one)

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

C) *Brother* or *sister* (please circle one)

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

D) *Brother* or *sister* (please circle one)

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

Grandparents

A) *Mother's mother*

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

B) *Mother's father*

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

C) *Father's mother*

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

D) *Father's father*

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

11. Language/s spoken, on average, by the child to:

A) family members:

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

B) Others (e.g. friends):

English____approximately ___% of the time.

Welsh____ approximately ___% of the time.

Other____(please specify_____)

approximately ____% of the time.

Thank you for completing this questionnaire and permitting your child to participate in this study.

Welsh Adult Questionnaire

Holiadur

Rhoddwch y wybodaeth a ganlyn os gwelwch yn dda.

1. Enw: _____

2. Dyddiad geni:_____

3. Lle cawsoch eich geni: _____

4. Ydych wedi byw yng Ngogledd Cymru erioed?

Do____ **Naddo**____

Os cawsoch eich geni y tu allan i Ogledd Cymru, neu os gwnaethoch dreulio peth amser yn byw mewn ardal arall, ym mhle y buoch yn byw a faint oedd eich oed pan ddaethoch i Ogledd Cymru?

Lle:_____

Oed:_____

5. Eich iaith cyntaf yw _____

6. Rydych yn siarad (ticiwch bob un sy'n berthnasol)

Saesneg_____

Dechreuais siarad Saesneg pan oeddwn yn _____ oed

Ar hyn o bryd, rwy'n siarad Saesneg tua ____% or amser.

Cymraeg_____

Dechreuais siarad Cymraeg pan oeddwn yn _____ oed

Ar hyn o bryd, rwy'n siarad Cymraeg tua ____% or amser.

Arall_____(nodwch:_____)

Dechreuais siarad yr iaith hon pan oeddwn yn _____ oed

Ar hyn o bryd, rwy'n siarad yr iaith hon tua ____% or amser.

7. Iaith/ieithoedd a siaredir yn y cartef ar hyn o bryd (ticiwch un)

___100% Saesneg

___ 80% Saesneg, 20% Cymraeg

___ 60% Saesneg, 40% Cymraeg

___ 50% Saesneg, 50% Cymraeg

___ 60% Cymraeg, 40 % Saesneg

___ 80% Cymraeg, 20% Saesneg

___100% Cymraeg

___ Iaith arall/ canran, nodwch

_____

8 Iaith/ieithoedd a siaredid gartef tra oeddwn yn tyfu i fyny (ticiwch un)

___100% Saesneg

___ 80% Saesneg, 20% Cymraeg

___ 60% Saesneg, 40% Cymraeg

___ 50% Saesneg, 50% Cymraeg

___ 60% Cymraeg, 40 % Saesneg

___ 80% Cymraeg, 20% Saesneg

___100% Cymraeg

___ Iaith arall/ canran, nodwch

_____

9. Iaith/ieithoedd a siaredir yn y gwaith (ticiwch un)

___100% Saesneg

___ 80% Saesneg, 20% Cymraeg

___ 60% Saesneg, 40% Cymraeg

___ 50% Saesneg, 50% Cymraeg

___ 60% Cymraeg, 40 % Saesneg

___ 80% Cymraeg, 20% Saesneg

___100% Cymraeg

___ Iaith arall/ canran, nodwch

_____

10. Iaith/ieithoedd a siaredir gan aelodau fy nheulu:

Mam

Ei hiaith gyntaf yw _____. Dechreuodd ei siarad pan oedd yn _____ oed ac mea'n ei defnyddio tua _____% or amser.

Iaith/ieithoedd eraill (nodwch)_____

Tad

Ei iaith gyntaf yw _____. Dechreuodd ei siarad pan oedd yn _____ oed ac mea'n ei defnyddio tua _____% or amser.

Iaith/ieithoedd eraill (nodwch)_____

Partner

Ei (h)iaith gyntaf yw _____. Dechreuodd ei siarad pan oedd yn _____ oed ac mea'n ei defnyddio tua _____% or amser.

Iaith/ieithoedd eraill (nodwch)_____

Eraill

Ar gyfartaledd, mea'r rhan fwyaf o'm ffrindiau, cywiethwyr, brawd/brodyr a/neu chwaer/chwiorydd yn siarad _____ ac maen't yn ei defnyddion tua _____% or amser.

11. Languages spoken by family members to yourself:

Mam

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

Father

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

Partner

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

Eraill

Ar gyfartaledd, mea'r rhan fwyaf o'm ffrindiau, cywiethwyr, brawd/brodyr a/neu chwaer/chwiorydd yn siarad

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

12. Iaith/ieithoedd a siaredir gennych chi, ar gyfartaledd, wrth::

A) aelodau'r teulu

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

B) Eraill (e.e. ffrindiau, cywiethwyr):

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

Diolch ichi am lenwi'r holiadur hwn ac am gymryd rhan yn yr astudiaeth hon.

Welsh Child Questionnaire

Holiadur

Rhoddwch y wybodaeth a ganlyn os gwelwch yn dda.

1. Enw'r plentyn: _____

2. Dyddiad geni'r plentyn:_____

3. Lle ganwyd y plentyn: _____

4. Ydi'ch plentyn wedi byw yng Ngogledd Cymru erioed? **Do**____ **Naddo**____

Os ganwyd eich plentyn y tu allan i Ogledd Cymru, neu os treuliodd peth amser yn byw mewn ardal arall, ym mhle roedd yn byw a faint oedd ei (h)oed pan ddaethoch i Ogledd Cymru?

Lle:_____

Oed:_____

5. Iaith gyntaf y plentyn yw _____

6. Mae'r plentyn yn siarad (ticiwch bob un sy'n berthnasol):

Saesneg_____

Dechreuodd siarad Saesneg pan oedd yn _____oed

Ar hyn o bryd, mea'n siarad saesneg tua ____% or amser.

Cymraeg_____

Dechreuodd siarad Cymraeg pan oedd yn _____oed

Ar hyn o bryd, mea'n siarad Cymraeg tua ____% or amser.

Arall_____

Dechreuodd siarad yr iaith hon pan oedd yn _____oed

Ar hyn o bryd, mea'n siarad saesneg tua ____% or amser.

7. Iaith/ieithoedd a siaredir gan holl aelodau'r teulu yn y cartef (ticiwch un)

___100% Saesneg

___ 80% Saesneg, 20% Cymraeg

___ 60% Saesneg, 40% Cymraeg

___ 50% Saesneg, 50% Cymraeg

___ 60% Cymraeg, 40 % Saesneg

___ 80% Cymraeg, 20% Saesneg

___100% Cymraeg

___ Iaith arall/ canran, nodwch

_____

8. Iaith/ieithoedd a siaredir yn ysgol y plentyn (ticiwch un)

___100% Saesneg

___ 80% Saesneg, 20% Cymraeg

___ 60% Saesneg, 40% Cymraeg

___ 50% Saesneg, 50% Cymraeg

___ 60% Cymraeg, 40 % Saesneg

___ 80% Cymraeg, 20% Saesneg

___100% Cymraeg

___ Iaith arall/ canran, nodwch

_____

9. Ieithoedd a siaredir gan aelodau'r teulu (llenwch y rhai berthnasol):

Mam/Prif Roddwr Gofal

Fy iaith gyntaf yw_____ . Dechreuais ei siarad pan oeddwn yn _____ oed ac rwyf ei defnyddio tua _____% or amser.

Iaith/ieithoedd eraill
(nodwch)_____

Tad

Ei iaith gyntaf yw_____ . Dechreuodd ei siarad pan oedd yn _____ oed ac mae'n ei defnyddio tua _____% or amser.
Iaith/ieithoedd eraill
(nodwch)_____

Brodyr a chwiorydd

A) *Brawd neu chwaer* (rhowch gylch o amgylch un)
Ei (h)iaith gyntaf yw_____ . Dechreuodd ei siarad pan oedd yn _____ oed ac mae'n ei defnyddio tua _____% or amser.
Iaith/ieithoedd eraill
(nodwch)_____

B) Brawd neu chwaer (rhowch gylch o amgylch un)
Ei (h)iaith gyntaf yw_____ . Dechreuodd ei siarad pan oedd yn _____ oed ac mae'n ei defnyddio tua _____% or amser.
Iaith/ieithoedd eraill
(nodwch)_____

C) *Brawd neu chwaer* (rhowch gylch o amgylch un)
Ei (h)iaith gyntaf yw_____ . Dechreuodd ei siarad pan oedd yn _____ oed ac mae'n ei defnyddio tua _____% or amser.
Iaith/ieithoedd eraill
(nodwch)_____

D) *Brawd neu chwaer* (rhowch gylch o amgylch un)
Ei (h)iaith gyntaf yw_____ . Dechreuodd ei siarad pan oedd yn _____ oed ac mae'n ei defnyddio tua _____% or amser.
Iaith/ieithoedd eraill
(nodwch)_____

Teidiau a neiniau

A) *Mam y fam*
Ei hiaith gyntaf yw_____ . Dechreuodd ei siarad pan oedd yn _____ oed ac mae'n ei defnyddio tua _____% or amser.
Iaith/ieithoedd eraill
(nodwch)_____

B) *Tad y fam*
Ei iaith gyntaf yw_____ . Dechreuodd ei siarad pan oedd yn _____ oed ac mae'n ei defnyddio tua _____% or amser.
Iaith/ieithoedd eraill
(nodwch)_____

C) *Mam y tad*
Ei hiaith gyntaf yw_____ . Dechreuodd ei siarad pan oedd yn _____ oed ac mae'n ei defnyddio tua _____% or amser.
Iaith/ieithoedd eraill
(nodwch)_____.

D) *Tad y tad*
Brawd neu chwaer (rhowch gylch o amgylch un)
Ei (h)iaith gyntaf yw_____ . Dechreuodd ei siarad pan oedd yn _____ oed ac mae'n ei defnyddio tua _____% or amser.
Iaith/ieithoedd eraill
(nodwch)_____.

10. Languages spoken by family members to the child:

Mam/Prif Roddwr Gofal

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

Tad

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

Brodyr a chwiorydd

A) *Brawd neu chwaer* (rhowch gylch o amgylch un)

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

B) *Brawd neu chwaer* (rhowch gylch o amgylch un)

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

C) *Brawd neu chwaer* (rhowch gylch o amgylch un)

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

D) *Brawd neu chwaer* (rhowch gylch o amgylch un)

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser..

Teidiau a neiniau

A) *Mam y fam*

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

B) *Tad y fam*

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

C) *Mam y tad*

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

D) *Tad y tad*

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

11. Iaith/ieithoedd a siaredir, ar gyfartaledd, gan y plentyn wrth:

A) Aelodau'r teulu:

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

B) Eraill (e.e. ffrindiau):

Saesneg____tua ___% or amser.

Cymraeg____ tua ___% or amser.

Arall (nodwch) _____ tua ____% or amser.

Diolch i chi am lenwi'r holiadur hwn ac am adael i'ch plentyn gymryd rhan yn yr astudiaeth hon.

Novel word forms used to create Welsh and English novel noun lists

Welsh

PIGELL	LLEIB	TESOL	DWLYDD	FFENYDD	TWNSYDD
TWRDD	RHEIDD	CWGEL	GABIL	ERYDD	PWRF
CERLITH	MURSAN	BATAIL	LLAIN	PESEL	PAIR
BELEM	NIPYN	DARRIG	RHAITH	TWRIN	GOLLT
DOLYN	IAINT	GEFALCH	MEIGEL	CADAN	GALAS
GWIBELL	PEFRYN	LLESARN	NELFA	BARTH	MERIN
LLABER	TOLFA	RHENG	FFALED	POGAIR	TOLUR
RHAIG	CERIN	MELAN	ILWR	TIGAIL	CEDUR
MALLT	BOLED	NOLER	PIBAN	CANAD	TOWL
NIGOL	DELAN	FFAGOR	TWNIN	BWLL	EDDAL
FFALOR	GLEIDD	YSGW	CAIG	DAIG	FORLAN
WANER	LLIBER	FFOLYDD	BLIGIOR	GELAN	TRELEW
POLET	RHIB	GLYPED	DIGW	MAGL	CORNIW
TABER	MEDAN	TWEIL	GARED	FFRITHIN	PEIRILL
CELLIG	NWLYDD	PALL	LLESEM	OBLWR	CLOMP
BALAN	FFAGOL	TABAR	RHODL	NAGUR	GLWRO
DABIR	OLYDD	CINW	MABEN	DIDAN	FFELW
GISARN	PIBOT	BANIAR	NOCHEL	RHATH	CRWYLL

English

PIRENT	PIGELEAM	LLEIBENDOLE	TESOPPERY	DWGERBICK	FFESYLMPLE
TEARTH	TWRIRETTE	RHEIDIGGLENT	CWGEBIRATE	GABILDRK	ERYDISH
CORLIT	CERMARSENT	MURSANILE	BATAIMIVE	LLAIWAIF	PESEINDLE
BELLITE	BELEMPIC	NIPYNGGET	DARRICH	RHAINT	TWRINANT
KEET	DOLYNENT	IAINTVIORID	GEFALCHLE	MEIGEL	CADANDLE
MIBBLE	GWIBELLBE	PEFRYNSTOIN	LLESARNTIC	NELFAOLAGE	BARTHPER
CHARBERT	LLABERKLY	TOLFAFF	RHENGKINT	FFALEDD	POGAIRF
REGORT	RHAIRKIN	CERMELLINT	MELANLOCK	ILWRULL	TIGAIRUNT
HICKET	MALOWLIFF	BOLEDTER	NOLERODEL	PIBANORSE	CANADORK
FLEEK	NIGELANT	DELANGOT	FFAGORY	TWNINEEM	BWLLMBLE
DALOUR	FFALARIT	GLEIDSTER	YSGWKIBET	CAIGROCK	DAIDICKET
WONERT	WANERBORY	LLIBERX	FFOLYDOO	BLIGIORDER	GELARG
POTCH	POLZARIB	RHIBORBITE	GLYPEDLIC	DIGWEEP	MAGLEET
ZABER	TABERIT	MEDWINE	TWEINK	GARED	FFRITHIN
GRALIT	CELLICTIC	NWLYDDETTE	PALLOVINARE	LLESEMUSE	OBLWRE
BALUME	BALANGLE	FFAGOLPOIL	TABARMPER	RHODLG	NAGURB
DABORE	DABIRINC	OLYDRINK	CINWROCKERT	MABENLE	DIDANGE
GRISTERN	GISARNZ	PIBOTUSH	BANIARVENT	NOCHELE	RHATHGE

English

Modified novel noun lists (including singular, plural and unmarked forms) in Welsh and English

Lists for version 1 in order of presentation

Welsh

PIGELLYN	LLEIBION	TESOLED	DWLYDDEN	FFENYDDYN	TWNSYDD
TWRDD	RHEIDDIAU	CWGELYN	GABIL	ERYDDOD	PWRFYN
CERLITHEN	MURSAN	BATAILON	LLAINAID	PESELEDD	PAIR
BELEM	NIPYNEN	DARRIGEN	RHAITH	TWRIN	GOLLTYN
DOLYNEDD	IAINTYN	GEFALCHIAID	MEIGELAID	CADANYN	GALASEN
GWIBELLYN	PEFRYNOEDD	LLESARN	NELFA	BARTHAU	MERINEN
LLABERYN	TOLFAOEDD	RHENGYN	FFALED	POGAIRYDD	TOLURYDD
RHAIG	CERINEN	MELAN	ILWRYN	TIGAILEN	CEDUREN
MALLT	BOLED	NOLER	PIBANEN	CANADION	TOWL
NIGOL	DELANEN	FFAGOR	TWNIN	BWLL	EDDALOD
FFALOR	GLEIDDION	YSGWYN	CAIGEN	DAIG	FORLAN
WANER	LLIBER	FFOLYDD	BLIGIORYN	GELANYN	TRELEWOD
POLETYN	RHIB	GLYPEDI	DIGWAID	MAGLYN	CORNIWEN
TABERYDD	MEDANEN	TWEIL	GAREDI	FFRITHIN	PEIRILL
CELLIGEN	NWLYDD	PALL	LLESEM	OBLWYN	CLOMPYDD
BALANOD	FFAGOLED	TABARON	RHODLIAID	NAGUROEDD	GLWROEDD
DABIR	OLYDDEN	CINWEN	MABENYN	DIDANIAU	FFELWEN
GISARNIAU	PIBOT	BANIARAINT	NOCHEL	RHATHION	CRWYLL

English

A PIRENT	BLEAMES	TENDOLES	A JOPPERY	A GRIBBICK	SHUMPLE
TEARTH	HIRETTES	A QUIGGLENT	GABIRATE	VORKS	A NUSH
A CORLIT	MARSENT	BANTILES	ANOMIVES	TWAIFS	RINDLE
BELLITE	A CLOPIC	A DRIGGET	ROTCH	CRINT	A GLANT
KEETS	A YOVENT	GIVIORIDS	CORTILES	A PLIN	A BINDLE
A MIBBLE	PERTOBES	BLESTOIN	NEBLITIC	FROLAGES	A HAPER
A CHARBERT	TORKLIES	A QUAFF	MOLKINTS	LIGS	CORFS
REGORT	A KARKIN	MELLINT	A SHILLOCK	A RULL	A SPRUNT
HICKET	BOWLIFF	NOLTER	A DROXEL	BORSES	THORK
FLEEK	A DELANT	FINGOT	EEDY	SWEEM	GIMBLES
DALOUR	GLARITS	A TROSTER	AN ORKIBET	GROCK	DICKET
WONERT	LIMBORY	HINX	A MANIDOO	A PLINDER	BROGS
A POTCH	ZARIB	GORBITES	WOTLICS	A FLEEP	A PREET
ZABERS	A MIVIT	TWIVE	DENKS	GIRD	FAND
A GRALIT	NILTICS	BLIGGETTE	JOVINARE	A FLOUSE	PIRES
BALUMES	FORGLES	CARPOILS	GOMPERS	YOGS	CHOBS
DABORE	AN ORBINC	A WARINK	A BROCKERT	JABLES	A LAGE
GRISTERNS	CHUZ	VUSHES	SLOVENT	NEAFS	SADGE

Lists for version 2 in order of presentation

Welsh

TWRDDYN
PIGELL
CWRLITHION
BELEMOEDD
MALLTIAU
GWIBELLAU
LLABERAU
RHEIGIAU
DOLYN
POLET
FFALOREN
WANERYN
NIGOLYN
TABER
CELLIG
BALANEN
DABIREDD
GISARNEN

LLEIBYN
NIPYNION
MURSANEN
RHEIDDYN
PERFRYN
IAINTOD
TOLFA
CERIN
BOLEDOEDD
NWLYN
OLYDDION
LLIBERYN
RHIBYN
MEDANOEDD
DELANOEDD
PIBOTIAU
GLEIDDYN
FFAGOL

TESOLEN
CWGEL
TWEILYDD
DARRIG
GEFALCHIAID
LLESARNEN
RHENGAINT
MELANEN
NOLERYN
FFAGORIAU
YSGWOEDD
FFOLYDDION
PALLI
BATAIL
GLYPED
TABAR
BANIAR
CINWED

DWLYDDEN
GABILEN
LLAINYN
RHEITHYN
MEIGEL
BLIGIORYN
FFALYN
ILWRAID
PIBAN
GARED
CAIG
NELFAYN
DIGWYN
TWNINED
RHODL
LLESEMOEDD
DEIGYN
NOCHELYN

FFENYDD
ERYDD
PESELEN
TWRINED
TOLUREN
BARTH
POGAIR
PEIRILLEN
CANAD
BWLLAU
MABEN
GELANEDD
CRWYLLYN
FFRITHYN
OBLWR
NAGUREN
DIDANEN
RHATH

TWNSYDDEN
GLWRYN
PAIRYN
GOLLTIAU
GALASI
MERIN
PWFRYNOEDD
CEDUR
TOWLIAINT
FORLANYN
EDDAL
TRELEW
CORNIW
TIGAIL
CLOMP
CADAN
FFELW
MAGL

English

A TEARTH
PIRENT
CORLITS
BELLITES
HICKETS
MIBBLES
CHABERTS
REGORTS
KEET
POTCH
A DALOUR
A WONERT
A FLEEK
ZABER
GRALIT
A BALUME
DABORES
A GRISTERN
A BLEAM
CLOPICS

A MARSENT
A HIRETTE
A PERTOBE
YOVENT
TORKLY
KARKIN
BOWLIFFS
A NILTIC
ORBINKS
A LIMORY
A ZARIB
MIVITS
DELANTS
CHUZES
A GLARIT
FORGLE
A TENDOLE
QUIGGLENT
TWIVES
DRIGGET

A GIVIORID
BLESTOINS
QUAFFS
MELLINT
NOLTERS
FINGOTS
TROSTERS
HINXES
BLIGGETTES
BANTILE
GORBITE
CARPOIL
A VUSH
WARINKS
JOPPERIES
A GABIRATE
AN ANOMIVE
A ROTCH
CORTILE
MANIDOO

A MOLKINT
SHILLOCKS
DROXEL
DENK
ORKIBET
A NEBILTIC
A WOTLIC
EEDIES
GOMPER
JOVINARES
A GROCK
SLOVENTS
GRIBBICKS
A VORK
TWAIF
A CRINT
A CORF
FROLAGE
LIG
A FAND

BORSE
SWEEMS
BROCKERT
PLINDERS
A SADGE
A GIRD
FLOUSE
A YOG
A JABLE
NEAF
A SHUMPLE
A CHOB
A RINDLE
GLANTS
BINDLES
HAPER
NUSHES
SPRUNT
THORKS
A DICKET

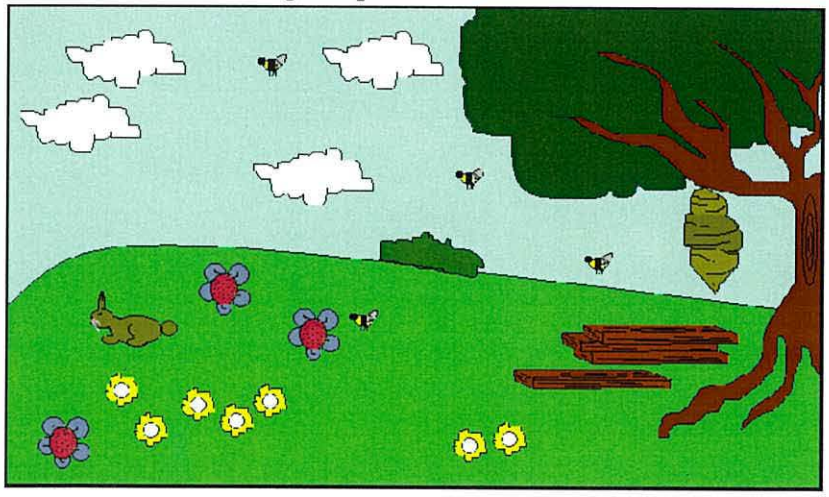
GIMBLE
BROG
PREET
RULL
PIRE
PLIN
LAGE
FLEEP

Stimuli used in Study 4:Cognition of pictoral scenes

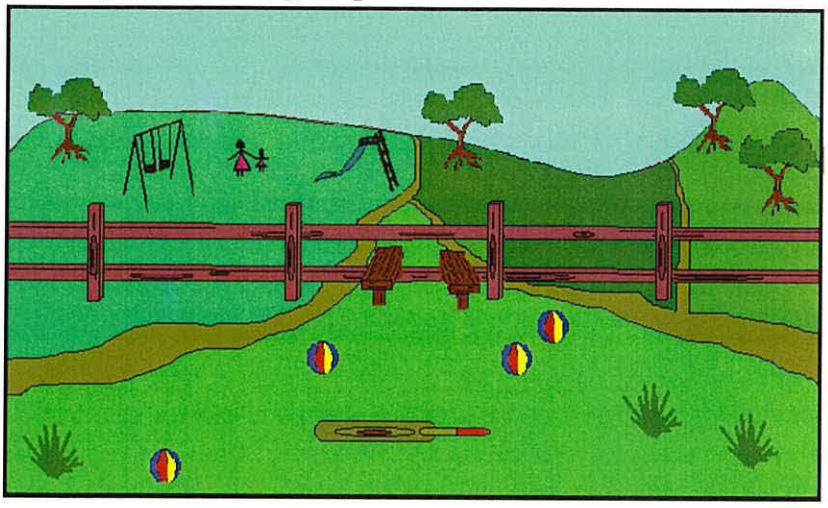
Task 1 beach scene: original picture

Task 1 kitchen scene: original picture

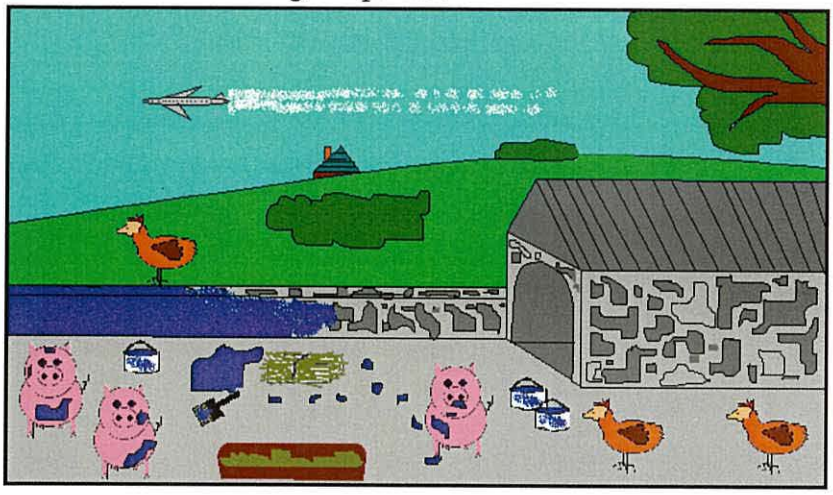
Task 1 bees scene: original picture



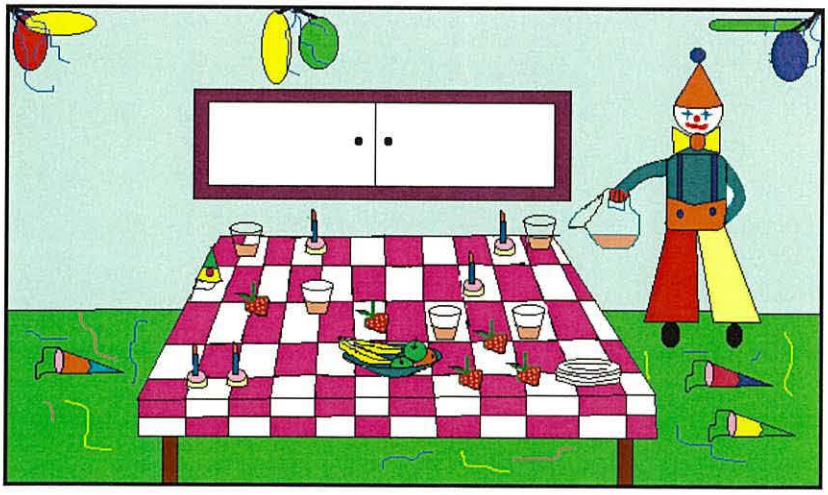
Task 2 park scene: original picture



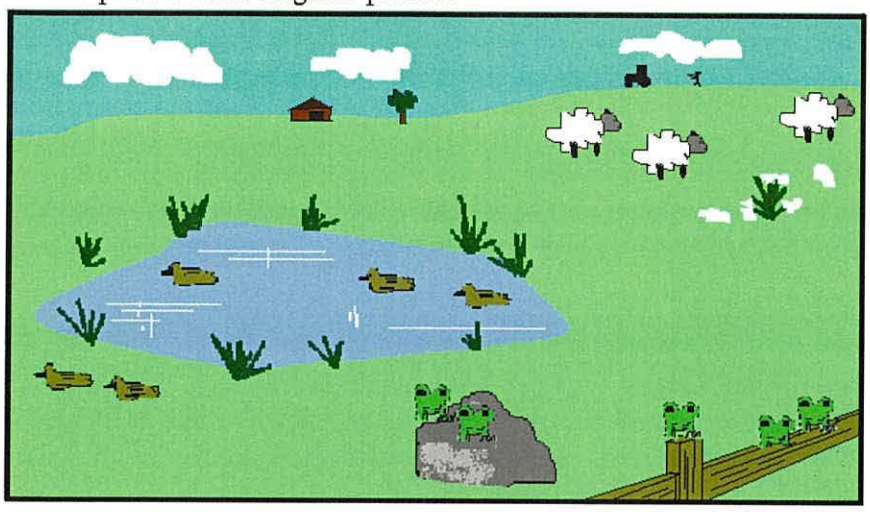
Task 2 farm scene: original picture



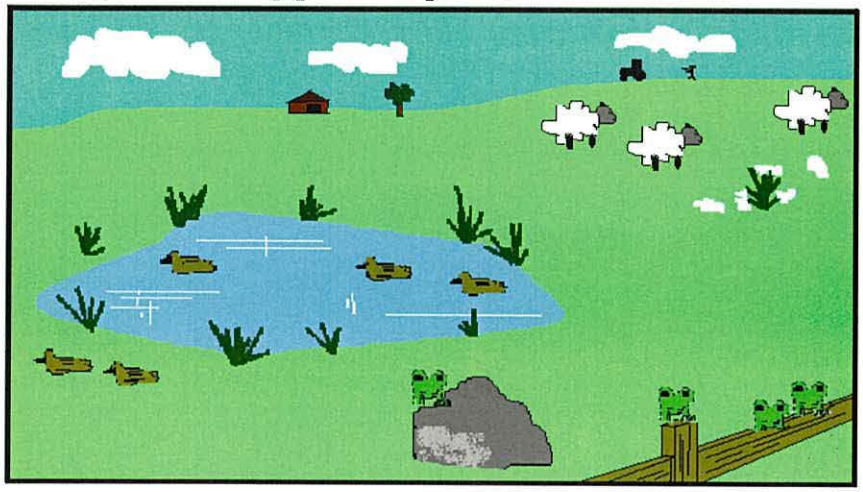
Task 2 party scene: original picture



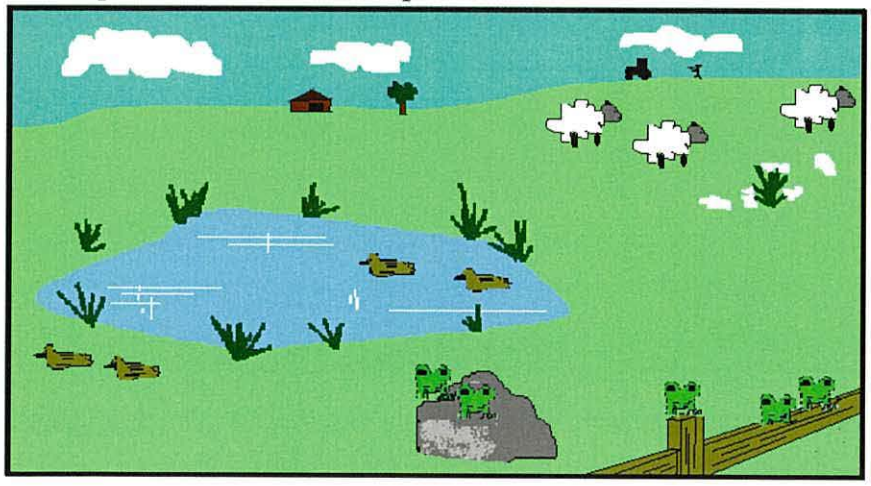
Task 3 pond scene: original picture



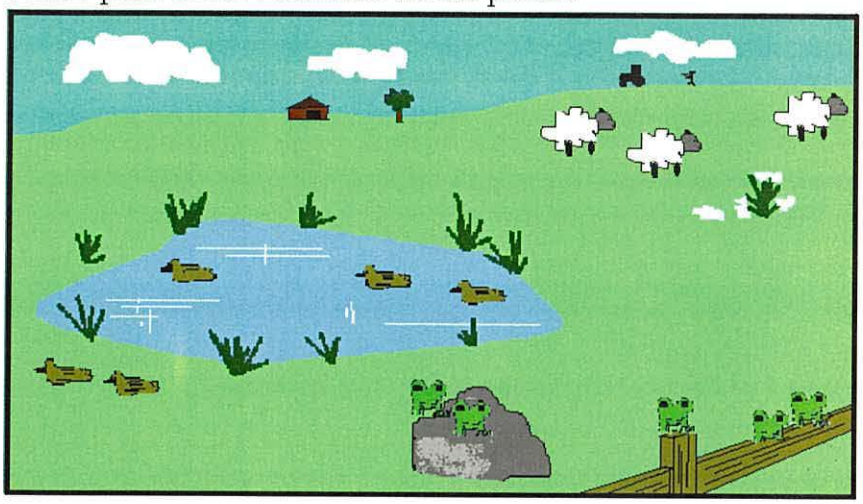
Task 3 pond scene: sg/pl variant picture



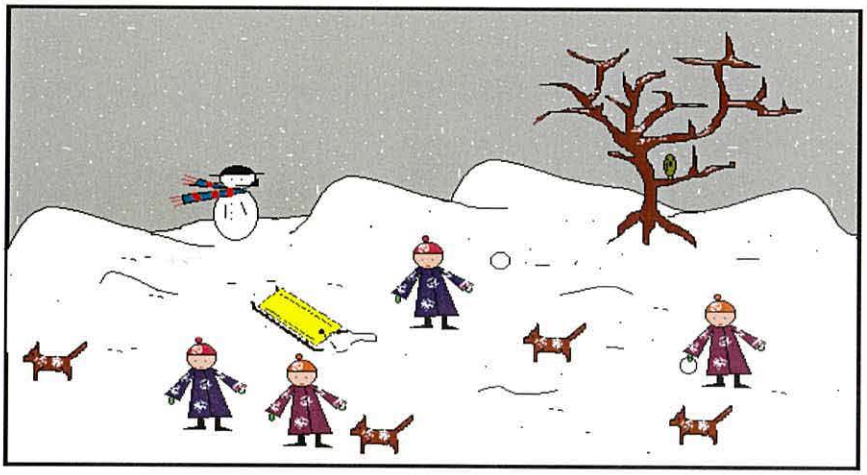
Task 3 pond scene: c/u variant picture



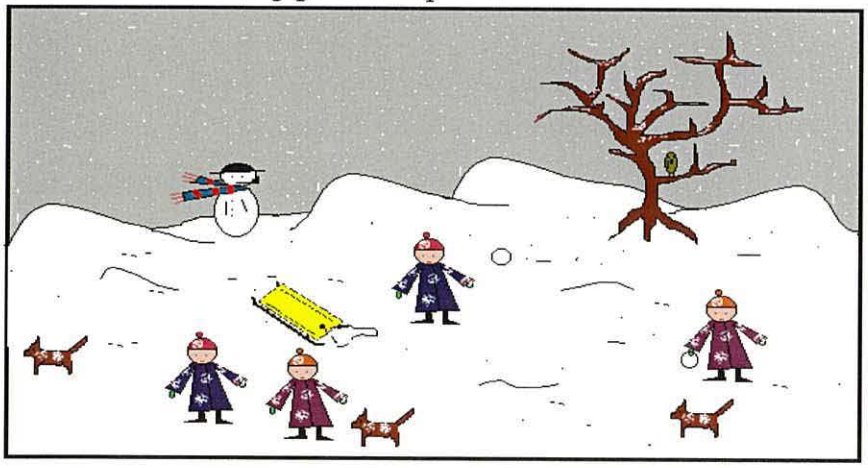
Task 3 pond scene: substance variant picture



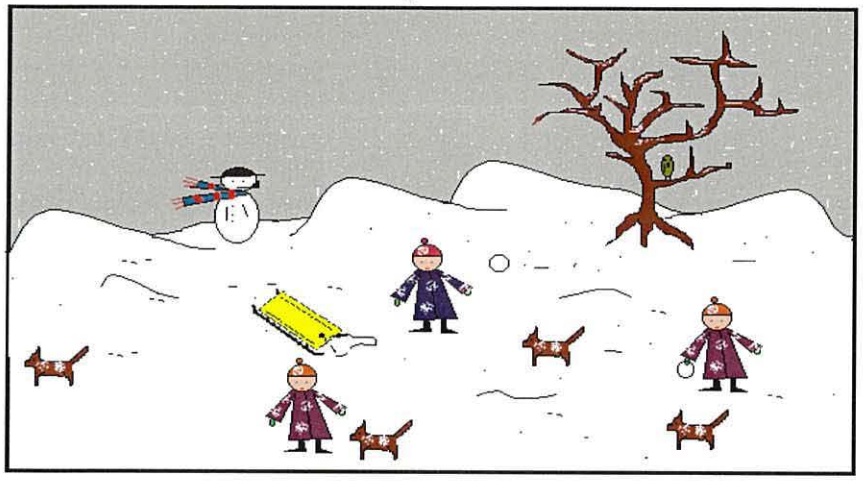
Task 3 snow scene: original picture



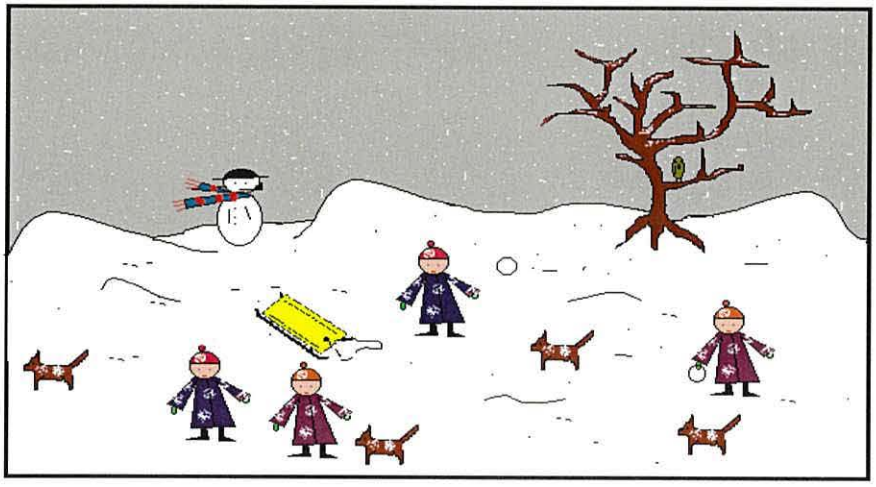
Task 3 snow scene: sg/pl variant picture



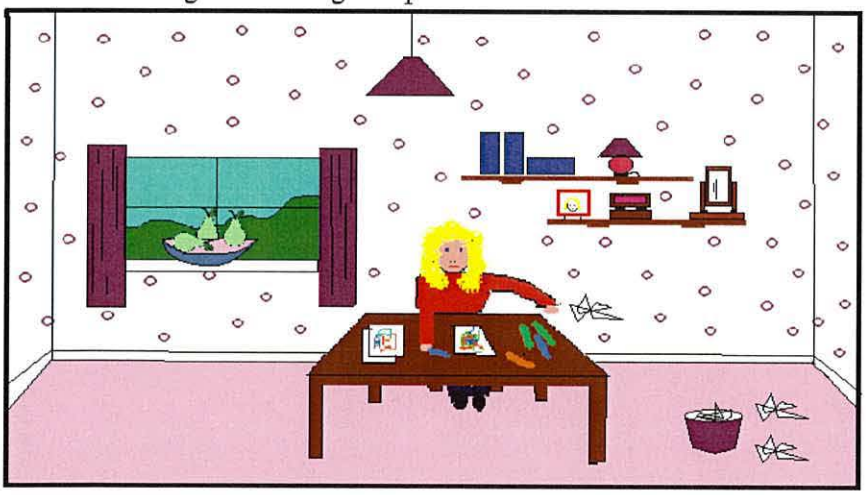
Task 3 snow scene: c/u variant picture



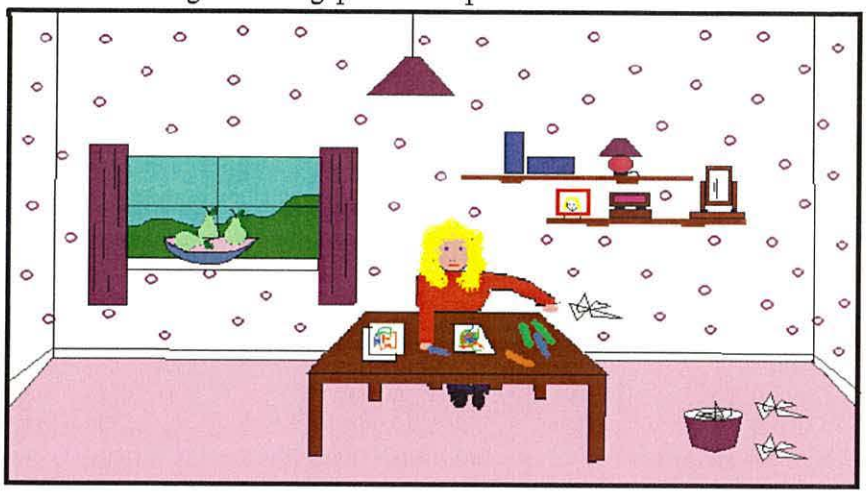
Task 3 snow scene: substance variant picture



Task 3 drawing scene: original picture

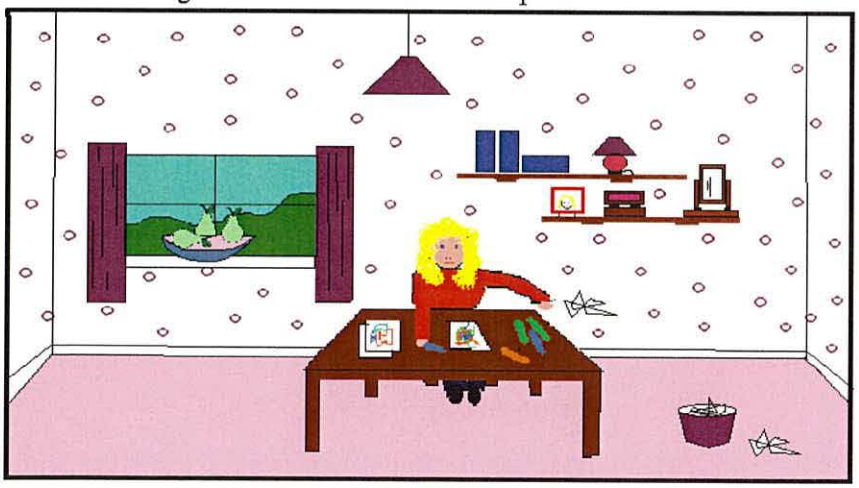


Task 3 drawing scene: sg/pl variant picture



Task 3 drawing scene: c/u variant picture

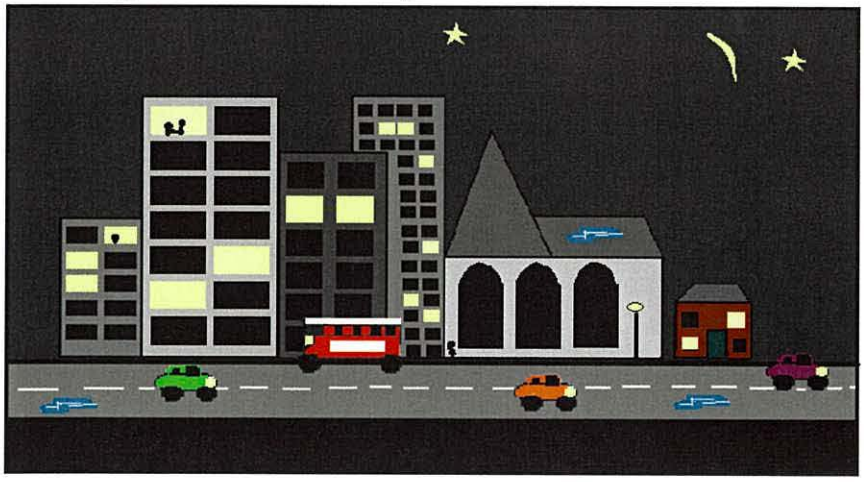
Task 3 drawing scene: substance variant picture



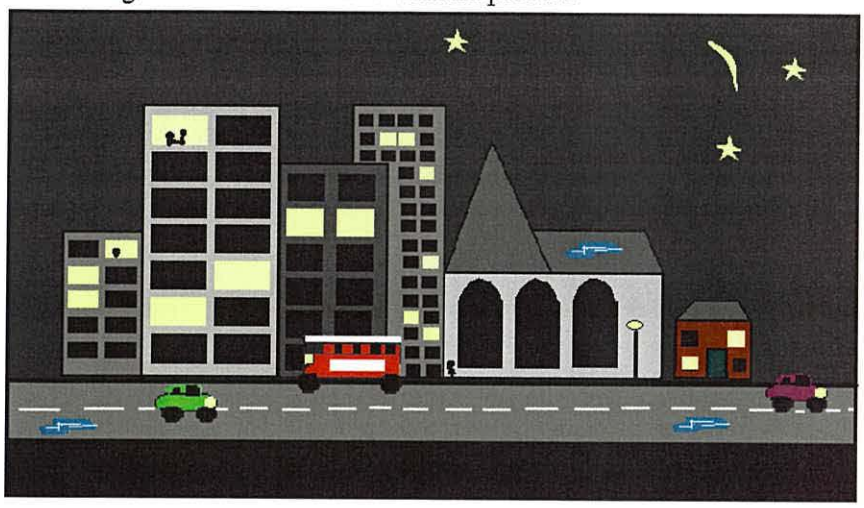
Task 4 night scene: original picture

Task 4 night scene: sg/pl variant picture

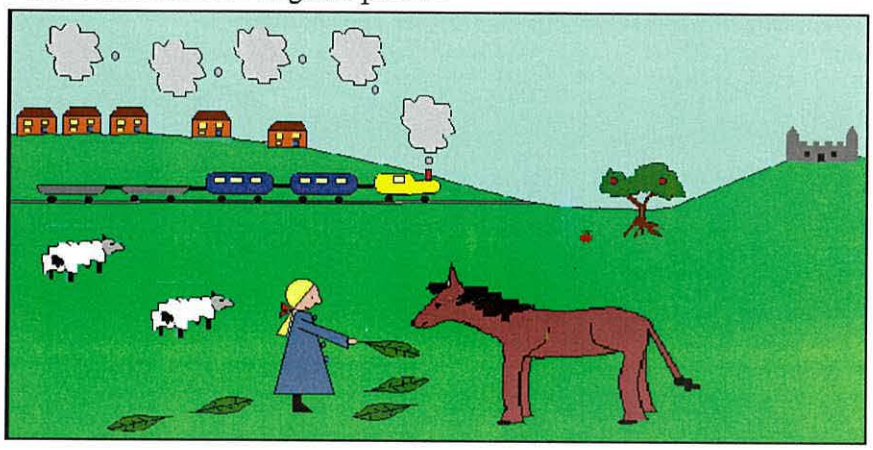
Task 4 night scene: c/u variant picture



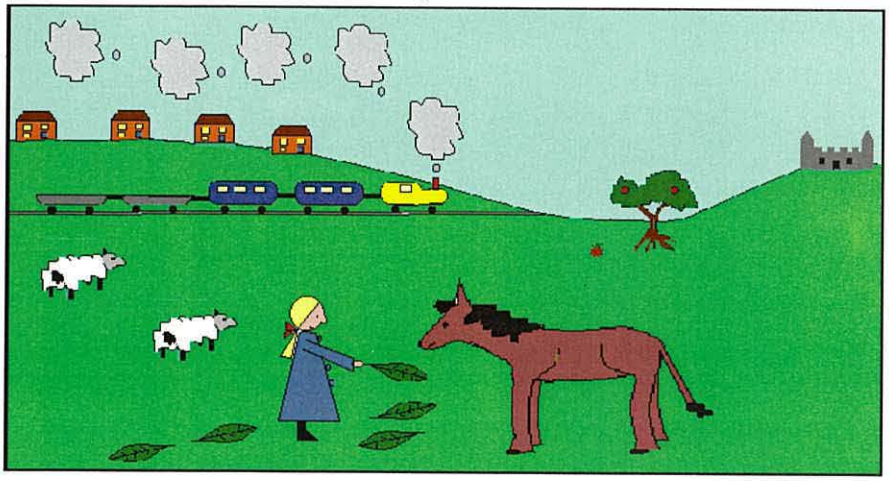
Task 4 night scene: substance variant picture



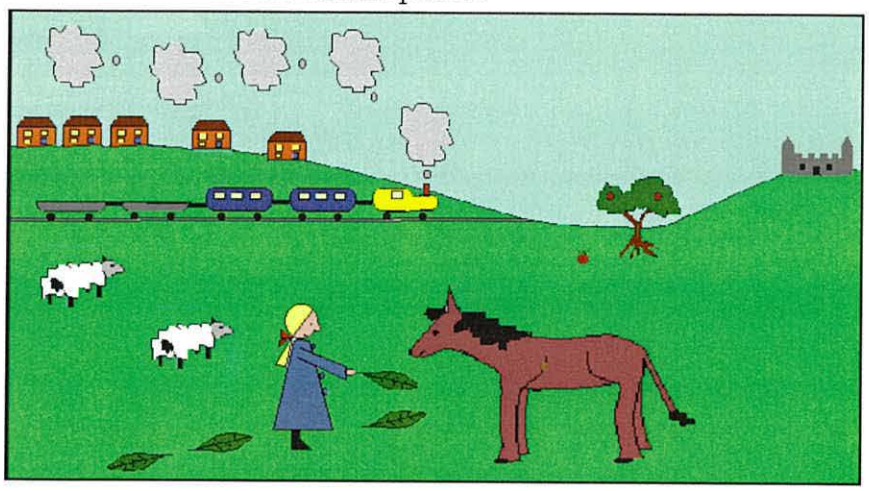
Task 4 train scene: original picture



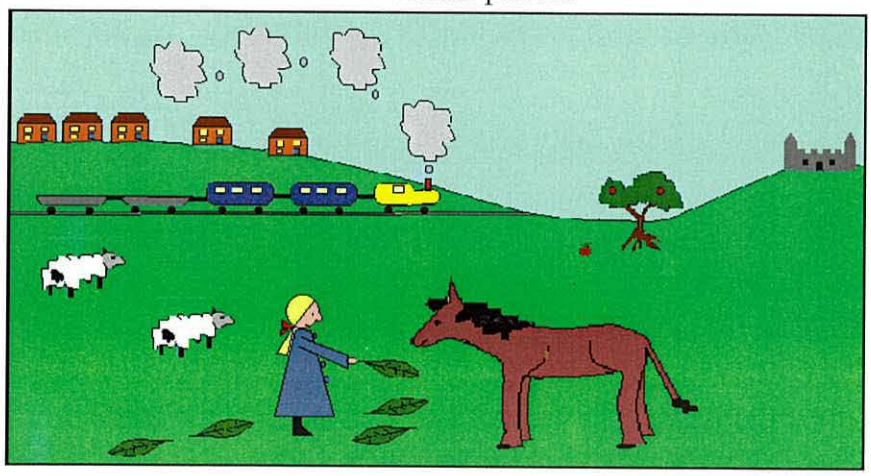
Task 4 train scene: sg/pl variant picture



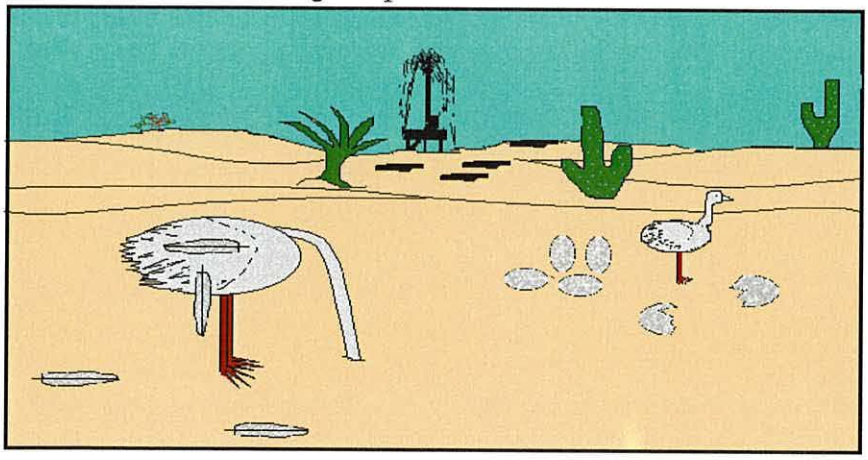
Task 4 train scene: c/u variant picture



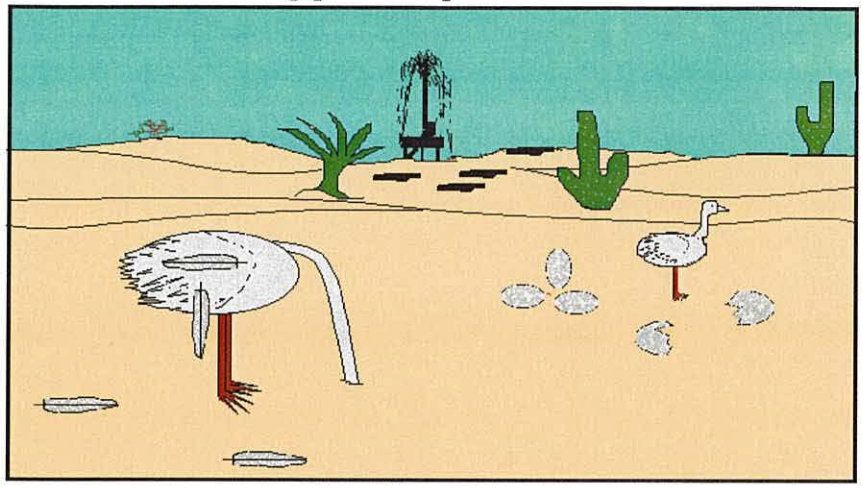
Task 4 train scene: substance variant picture



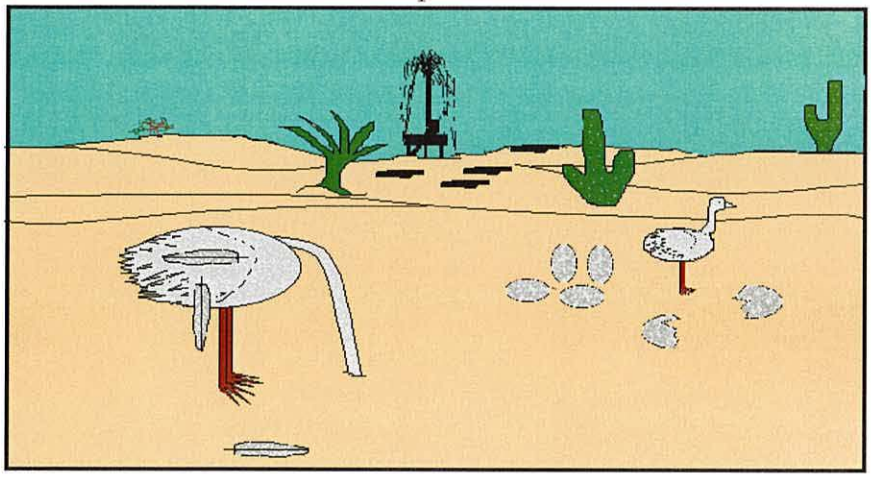
Task 4 ostrich scene: original picture



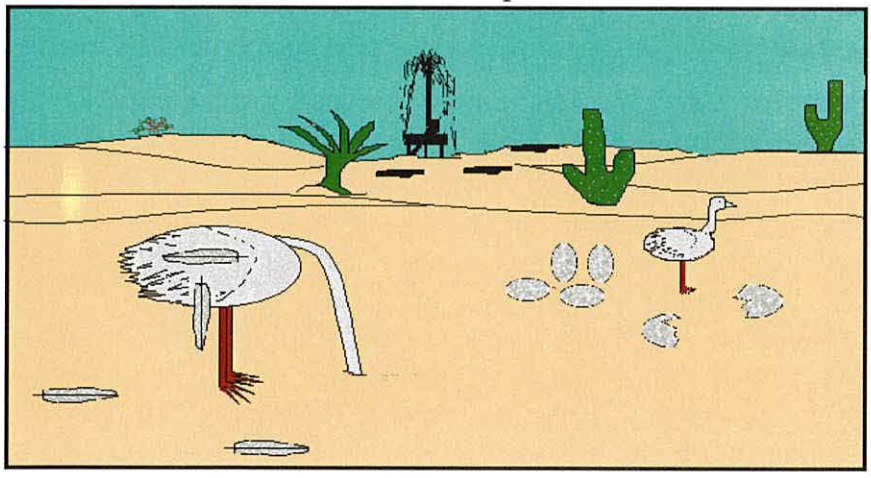
Task 4 ostrich scene: sg/pl variant picture



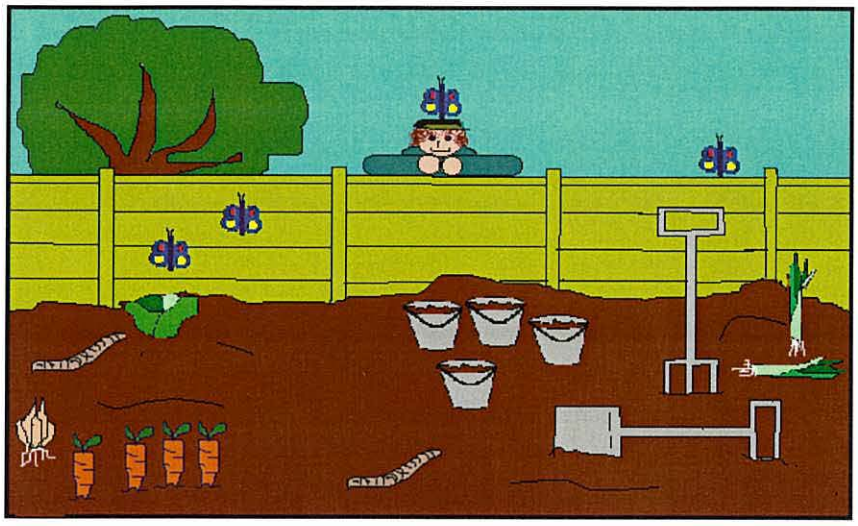
Task 4 ostrich scene: c/u variant picture



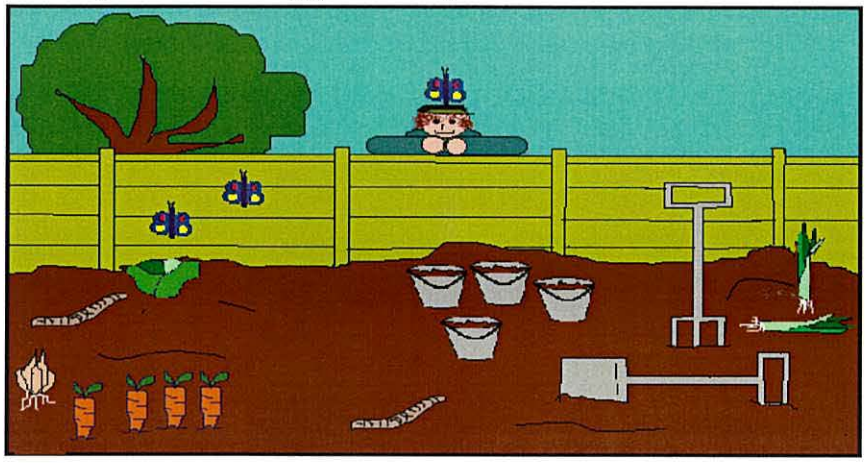
Task 4 ostrich scene: substance variant picture



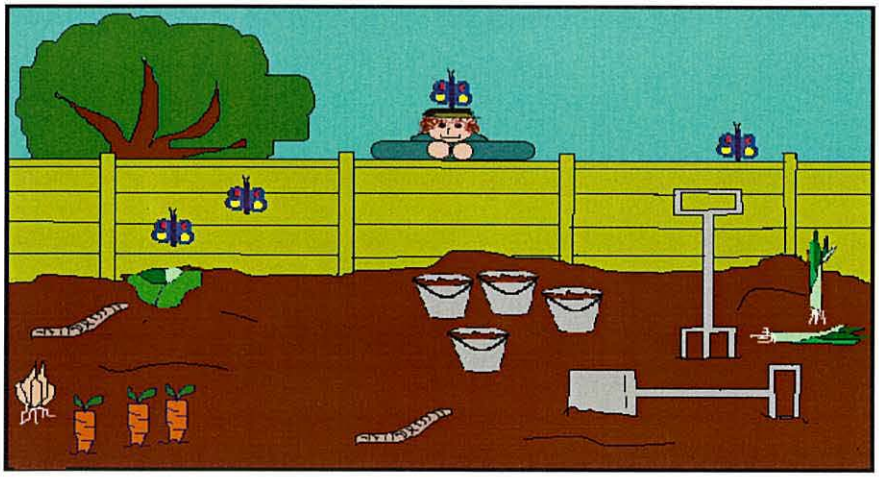
Task 5 garden scene: original picture



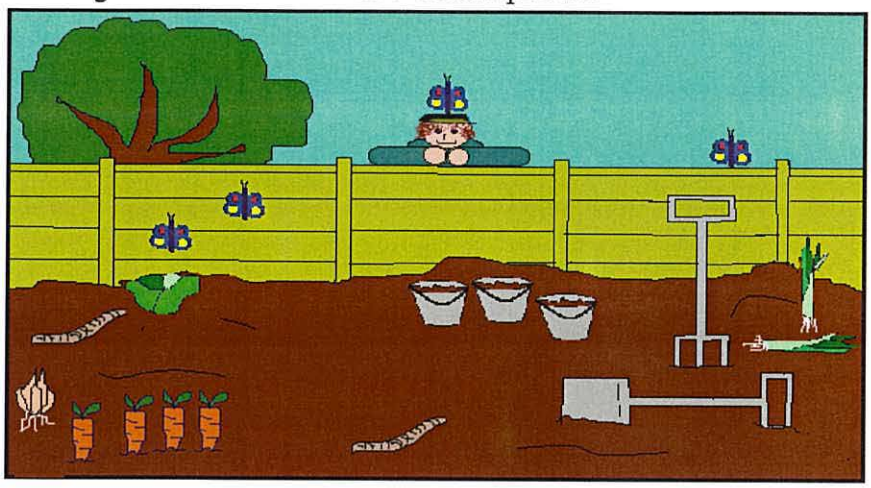
Task 5 garden scene: sg/pl variant picture



Task 5 garden scene: c/u variant picture



Task 5 garden scene: substance variant picture



Task 5 ants scene: original picture

Task 5 ants scene: sg variant picture

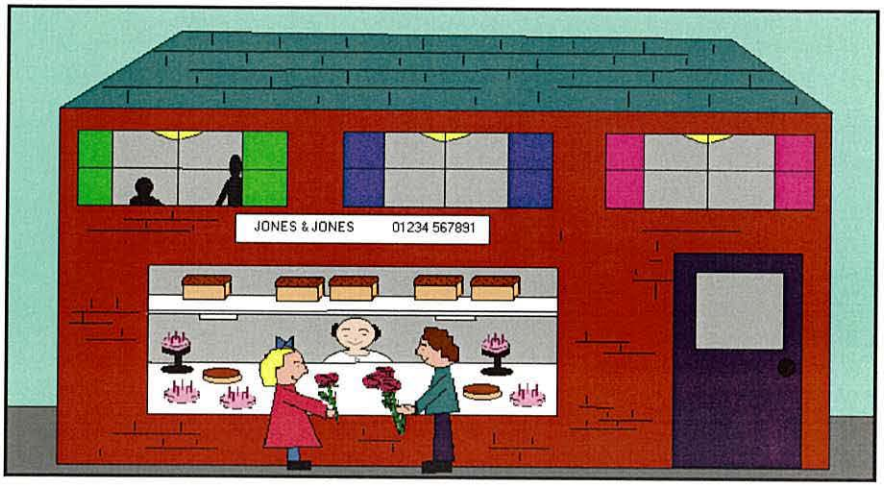
Task 5 ants scene: c/u variant picture

Task 5 ants scene: substance variant picture

Task 5 bakery scene: original picture

Task 5 bakery scene: sg/pl variant picture

Task 5 bakery scene: c/u variant picture



Task 5 bakery scene: substance variant picture

Scoring sheet

Name..... School.....
Date..... Year.....

Task					
1	Notes				
2					
3		V 1	V 2	V 3	
		V 1	V 2	V 3	
		V 1	V 2	V 3	
4	Original	V 1	V 2	V 3	
	Original	V 1	V 2	V 3	
	Original	V 1	V 2	V 3	
5	Original	V 1	V 2	V 3	
	Original	V 1	V 2	V 3	
	Original	V 1	V 2	V 3	

Summary table of alternate nouns used in tasks 1 and 2 of study 4

Language	Task 1 Target items								
	shells	**birds**	**sand**	**clouds**	**bees**	**wood**	**cups**	**mice**	**flour**
Welsh	anifeiliaid Pysgod/yn Pryfaid Pethau 'Monsters' Cranc/iau 'Caterpillar' Cerrig 'Slug' 'Crabs'	Gwylan/od 'Seagulls'	Swnd 'Sand'			Prenniau Priciau Coed 'Boards'			Llwch Pwdwr Siwgr Baw Bwyd 'Dirt'
English	Ice-cream Worms Lollies Pink things Fish	seagulls				logs	mugs	rats	Sugar Rubbish Sand Dust Powder Salt
	Task 2 Target items								
	balls	**trees**	**grass**	**hens**	**pigs**	**paint**	**cakes**	**strawberries**	**juice**
Welsh		Coedwig	Gwair Gwellt	Chwiaid Ceiliog			Teisennau	Ffrwythau 'Strawberries'	Oren Diod Dŵr Jŵs
English	Football			Chick/s Ducks Cockerel/s				Fruit Cherries	Orange Drinks Squash Pop