Language Structure and Language Acquisition: Grammatical Categorization Using Phonological and Distributional Information

Michelle Christina St. Clair

Submitted for the degree of Doctorate of Philosophy

University of York
Department of Psychology

October 2007

Abstract

This thesis addresses the question of how words are grouped according to their grammatical categories during language acquisition. Over the past 20 years a general consensus has developed that distributional and phonological cues are important cues that language learners utilize in the grammatical categorization process (e.g., Kelly, 1992; Redington, Chater, & Finch, 1998). The combination of these cues was investigated with artificial language learning experiments, which combined two categories of phonologically coherent words with co-occurring distributional cues, and corpus analysis techniques. Experiments 1 to 4 indicated that both phonological and distribution cues are necessary for the categorization of high and low frequency words. Additionally, these experiments indicated that distributional information alone was sufficient to categorize high frequency words, but that phonological cues were necessary for low frequency words. It was also found that succeeding bigram distribution cues induced more grammatical categorization than the preceding bigram cues. This is explained by the Rescorla-Wagner (1972) model of associative learning; associations were stronger between the category words and succeeding cues as a single succeeding cue followed all category words. Associations were weaker with preceding cues as numerous category words followed the preceding cues. Experiments 5, 6 and 7 also found that the effectiveness of the distributional cues was influenced by prior linguistic experience, resulting in higher learning with distributional cues which were phonologically consistent with distributional cues found in the participants' native language (English). This thesis also investigated the debate as to what type of distributional cue is most useful in the categorization process, with some researchers advocating trigram cues (Mintz, 2002) while others advocate bigram cues (Monaghan & Christiansen, 2004; Valian & Coulson, 1988). The results of a corpus analysis and two experiments provided evidence that trigram cues (aXb) are very effective at categorization, but preliminary evidence suggests that this categorization may simply be due to the combined influence of the beginning and ending bigrams (aX and Xb). Overall, this thesis indicates that phonological and distributional cues are key to grammatical categorization, which occurs through associative learning principles; grammatical categorization progresses faster with succeeding cues; and bigram distribution cues may be the initial source of distributional information in the grammatical categorization process.

List of Tables

List of Figures

Acknowledgements

The first and foremost person who needs acknowledgement is my supervisor, Padraic Monaghan. For his patience throughout all three years, I must offer my thanks and gratitude. He has provided invaluable guidance in all aspects of this work – from the experimental design phase to the final drafts. But most of all, I must thank Padraic for his constant encouragement and faith that it would all come together in the end.

The members of my research committee, Gareth Gaskell and Graham Hitch, I thank for the numerous suggestions and discussion of this research that we have had over the past three years. Their contributions and ideas have helped shape the research presented within thesis, as have the ideas and suggestions from the members of the Psycholinguistic Research Group. Also of mention is Professor Michael Ramscar, who was kind enough to come to York to discuss the associative learning aspect of my research.

To Jakke Tamminen I thank for his numerous discussions with me on the intricacies of the International Phonetic Alphabet. To Liz Littlewood I thank for keeping me sane in the final few weeks (actually, throughout the entire three years). To Leesa Clarke and Victoria Williamson I thank for being my "native English speakers". To Liz, Jakke, Marilena, Leanne, Rachel, Vicky, and finally Mark thank you for making C224 the best office in the Department and ensuring I was never bored throughout all three years. To my family and friends I thank for their support and understanding throughout the entire process. To Alex Hancock I thank for always believing in me and never allowing me to loose sight of the light at the end of the tunnel. To Sebastian I thank for always greeting me when I came home far too late.

Declaration

This thesis contains original work completed solely by the author under the supervision of Dr. Padraic Monaghan.

The data reported in Chapter 3 was presented at the 2007 joint meeting of the Experimental Psychology Society and the Psychonomic Society:

St. Clair, M. C., Monaghan, P. & Rasmcar, M. (2007) *Language morphology for language learning.* Poster presented at the joint meeting of the Experimental Psychology Society and the Psychonomic Society, University of Edinburgh, Edingurgh, UK.

1. Review of language acquisition and grammatical categorization literature

This thesis is an investigation into the process of acquiring grammatical categories. Grammatical categories are a precursor to the constraints of grammar; grammatical phrases cannot be constructed or understood without knowledge of grammatical categories. The question that this thesis will address is how grammatical categories may initially be learned during language acquisition. Two main sources of linguistic information will be focused on: distributional cues that occur across phrases and phonological cues which are inherent in the category words.

Distributional information has long been thought to be of fundamental importance in the grammatical categorization process (Maratsos & Chalkley, 1980). For instance, the word *the* often indicates that the next word is a noun, and the inflection *-ing* indicates that a verb was the preceding word. However, distributional information alone has often been shown to be insufficient to induce categorization with artificial languages (Braine et al., 1990).

Phonological information within words has also been shown to differentiate grammatical categories (Kelly, 1992). Nouns and verbs, as well as open and closed class words, can reliably be grouped according to the phonological characteristics that define each category (Kelly, 1992, 2006; Monaghan, Chater, & Christiansen, 2005; Shi, Morgan, & Allopenna, 1998). One theme of this thesis will be to determine whether distributional and phonological cues combined can provide a more complete account of grammatical categorization than either cue individually.

In addition to focusing on how distributional and phonological cues may combine to influence the grammatical categorization process, the precise location of distributional cues within grammatical phrases will also be investigated. Whether learning progresses more easily from preceding cues, such as *the* in the above example, or from succeeding cues, such as *–ing*, was a second theme of this thesis.

There is no debate that distributional information is highly informative and necessary in the grammatical categorization process. However, there is debate as to what type of distributional cue is the fundamental cue behind this categorization. Specifically, Mintz (2002, 2003, 2006) advocates "frame" cues, or non-adjacent a_b dependency cues, as being fundamental to the categorization process. However, other have advocated that simple bigram cues, or one word distributional contexts, may be sufficient to underlie grammatical categorization (Monaghan, Chater, & Christiansen, 2005; Monaghan & Christiansen, 2004). This thesis will aim to provide new evidence on what type of distributional cue may be at the basis of grammatical categorization.

This chapter begins by briefly reviewing the main debate in language acquisition and outlining possible language learning mechanisms. After this overview, a more specific review of research related to statistical language learning will be discussed, with particular reference to the relevant linguistic structures that may aid in the acquisition of grammatical categories. Finally, the research directly investigating the process of grammatical categorization will be discussed, with particular emphasis on the role of distributional and phonological cues in this process.

1.1. Theories of language acquisition

Undoubtedly, the work and ideas of Noam Chomsky have been the most influential in shaping language acquisition research and fuelling debate as to the fundamental nature of how language is acquired. In his rebuttal to B.F. Skinner's 1957 "Verbal Behaviour" book, which advocated a behaviourist account of language learning, Chomsky advocated a nativist stance on language acquisition; according to this view, we are all born with innate knowledge that guides the acquisition of language (e.g., Chomsky, 1965, 1980). Over the following decades, Chomsky published many influential books detailing the arguments against language learning as the basis of language acquisition, one of the most influential being the Poverty of the Stimulus arguments. Additionally, Chomsky outlined Universal Grammar, which are exact "Principles and Parameters" of language proposed to be innate. These "Principles and Parameters" were theorized to account for language acquisition in all languages. The principles were a series of rules to aid in disambiguating linguistic structure believed to be impossible to learn, an example of which is discussed in detail below. Specific language parameters were also theorized to be innate; these parameters would be set to the individual native language of the child. For instance whether or not a language can drop the initial pronoun (i.e. whether "*she has a cat*" and "*has a cat*" are equally grammatical) would be determined from the language input the “language acquisition device”, or LAD, received; with sufficient exposure the pronoun-drop parameter would be set to the correct setting (Chomsky, 2006).

Central to the innateness theory was the premise that the linguistic input infants are exposed to is not sufficient to account for their rapid development of a sophisticated grammatical systems that approach adult capacity after only a few years. This is the Poverty of the Stimulus argument. Linguists have given many examples of grammatical structures that were proposed to be impossible to learn from linguistic input. One objection is how English interrogative sentences can be formed from declarative sentences (Crain & Pietroski, 2001). For example, “*The cat is black*” can form an interrogative by moving the verb to the front of the sentence: “*Is the cat black?*” Although the simple rule of transferring the first verb could account for the above sentence, more complicated sentences would become a problem, such as “*The cat that is sleeping is black*”. For this sentence the transition of the first “*is*” is not grammatical “**Is the cat that sleeping is black?*” but moving the second “*is*” creates a grammatical phrase “*Is the cat that is sleeping black?*” The innateness explanation is that this is acquired through a “head movement constraint” in which the verb only can be moved locally (Crain & Pietroski, 2001). Thus, the noun phrase “*the cat is sleeping*” disallows the verb in this phrase from moving to the beginning of the sentence, allowing only the second phrase to make this transition.

Whether or not this regularity can be learnt through linguistic input alone was put to the test by Perfors, Tenenbaum, and Regier (2006). Using Bayesian models they tested how well three different grammars were able to learn the grammatical structure of sentences from a child directed speech corpus. There were two types of linear grammars and one context free hierarchical grammar. When tested on strings not included in the training, the hierarchical grammar was the only grammar to show substantial generalization to new phrases. Furthermore, when tested on the critical “*Is the cat that is sleeping black?*” phrases, only the hierarchical grammar could parse these sentences even though they had never seen sentences of this structure in the input. These results indicated that children do not need to be equipped with innate constructs in order to learn this structure. Importantly, the authors make the point that children learn language “as part of a *system* of knowledge … most PoS [Poverty of Stimulus] arguments consider some isolated linguistic phenomenon and conclude that because there is not enough evidence for that phenomenon in isolation, it must be innate" (Perfors, Tenenbaum, & Regier, 2006, p. 667).

This argument can be applied to virtually all of the Poverty of Stimulus examples put forward by nativists. Over the last 20 years, research discounting the nativists claims and supporting the idea of language learning from linguistic input have been put forward leading to a decrease in the viability of innate language knowledge and of the language acquisition device (e.g., Bates et al., 1998; Christiansen & Chater, 1999; Lewis & Elman, 2002). The general pattern that is emerging is that it is not the exact nature of language, but our ability to learn that is innately specified.

This ability to learn may be specified within the genetic code in the form of the architecture of the brain. The development of the brain, and indeed the whole body, is carefully orchestrated by genetics and progresses through set stages during foetal development and after birth (Vasta, Haith, & Miller, 1999). By constraining the architecture of the brain, genetics is at the basis of language processes (as well as all cognitive processes) in that it sets up the brain in a certain way to be able to learn. This highly constrained brain architecture may indeed be the answer to a second main point of the nativists, which is that learning language would involve endless possible combinations of both relevant and irrelevant linguistic regularities. As Bates et al. (1998) state: "the channels used by human language are subject to universal constraints on information processing ... Under these circumstances, we should not be surprised to find that the class of solutions to the problem is quite limited, constituting a set of alternatives that might be referred to as Universal Grammar" (p.595). Thus, the solution to the problem of endless possible learning sources may indeed be constrained by innate forces; however, it is much more likely these innate forces are constraining the *way* in which we can process and learn language rather than the actual linguistic knowledge (as discussed by Saffran, 2002; Saffran, 2003).

This hypothesis would predict that the learning processes are domain general, rather than modular as most innateness accounts claim. As expected, there is evidence of cross modality learning in many of learning mechanisms used during language acquisition (e.g., Altmann, Dienes, & Goode, 1995; Conway & Christiansen, 2005; Saffran, 2002) as well as evidence of cross species learning (e.g., Conway & Christiansen, 2001; Hauser, Newport, & Aslin, 2001; Hauser, Weiss, & Marcus, 2002;

Herbranson & Shimp, 2003; Newport, Hauser, Spaepen, & Aslin, 2004). These studies point strongly to domain general learning processes that are shared with other species as the basis for language acquisition.

One language learning account compatible with these ideas was put forward by Winter and Reber (1994) who investigated the implicit nature of language acquisition. It was theorized that the foundational cognitive structures upon which language (and all of cognition) must be based are unconscious or implicit cognitive structures. The main findings about the nature of this implicit learning were summarised in four points by Winter and Reber:

> "(a) it is a generalised, unconscious inductive process that is sensitive to underlying regularities and rule structures in the environment; (b) knowledge derived from these regularities is represented tacitly and abstractly; (c) while this knowledge typically remains resistant to conscious inspection, it can be utilised to guide subsequent task related decisions and judgements; (d) as a generalised process of cognition it is not exclusive to any specific domain of knowledge of functioning and is considered to extend across sensory modalities." (p.118)

These general characteristics of implicit learning offer a new insight into a possible mechanism for language acquisition. The implicit learning mechanism described above operates independently of consciousness which has direct parallels with the fact that children do not consciously analysis their linguistic input in order to extract the relevant structures necessary to achieve an adult competency in language (Winter & Reber, 1994). This process happens naturally and just as easily for a child with an IQ of 80 as with a child with an IQ of 120. The process of unconscious, automatic language learning has the promise of providing a realistic language learning paradigm in which to describe language acquisition.

This type of implicit learning has been widely found within the artificial grammar learning field. When given a series of letters that conformed to an underlying grammar, participants were able to reject ungrammatical test strings and accept grammatical test string, but were unable to give explicit knowledge of the underlying rules of the grammar (Dienes, Altmann, Kwam, & Goode, 1995; Reber, 1967). The nature of this learning has been shown to be implicit and importantly participants are able to transfer this implicit knowledge from the trained stimuli to new stimuli that have the same underlying grammar (Reber, 1969). Additionally, this implicitly learned knowledge has been found to transfer across modalities (Altmann, Dienes, & Goode, 1995; Dienes, Altmann, & Gao, 1999). The transfer across modalities of implicitly learned structural information is particularly intriguing as it underlies the modality independence of this type of learning.

Implicit learning has also been shown to apply to the statistical regularities within language, which has shed light on the process of speech segmentation (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996) and of higher order grammatical regularities (Gómez, 2002; Onnis, Monaghan, Richmond, & Chater, 2005; Peña, Bonatti, Nespor, & Mehler, 2002) and indeed on the domain general and cross species nature of statistical learning (Hauser, Newport, & Aslin, 2001; Newport, Hauser, Spaepen, & Aslin, 2004). In addition, connectionist neural networks have also been very successful at modelling many different aspects of language development, including the previous discussed declarative to interrogative sentence transformations that have fuelled many nativists arguments (e.g., Cartwright & Brent, 1997; Christiansen & Chater, 2001; Lewis & Elman, 2002; Monaghan & Christiansen, 2004; Onnis & Christiansen, 2005; Reali, Christiansen, & Monaghan, 2003; Shi, Morgan, & Allopenna, 1998). However, all of these accounts have the same basis behind them; mainly the *associative* underlying nature of the learning. Statistical co-occurrence learning is intricately tied into the associations between the linguistic units under investigation. Connectionist models are also based on the associations between different units and linguistic constructs. The next section will review associative learning literature and discuss how the learning principles from this alternative field may aid in explaining language learning.

1.2. Associative (language) learning

The animal associative learning and human learning fields of inquiry diverged in the 1950s, partly as a response to Chomsky's review of "Verbal Behaviorism" (1959). The rise of cognitive theories dominated the field of human learning for decades; integration of associative learning to human learning contexts only began in the 1980s. In regards to language learning, research that takes into account general associative learning principles from animal research is only beginning to influence the field. However, the idea that general associative mechanisms may be the basis of our implicit learning abilities is intriguing which creates the possibility that some of language acquisition may be explained in terms of well known associative learning theories. Although the formal application of associative learning ideas to language learning is just beginning, the traces of ideas either originating in or paralleling the animal learning field can be found throughout a wide range of language learning literature.

Although many studies (e.g., Ellis, 2006b) directly refer to associative language learning, others only vaguely refer to associations between linguistic elements (Smith, 2000). The former approach is taken in this thesis; therefore a brief description of the Rescorla-Wagner (or R-W; 1972) model will be given below. Following description of the R-W model, associative learning theories about general categorization will also be discussed briefly.

The R-W theory of associative learning is a theory that focuses on "surprise"; if the occurrence of an event (food) immediately after a cuing event (a light) is surprising, the association between the light and food will increase (Rescorla & Wagner, 1972). If every time a light occurred, food also occurred the association would continue to increase until the co-occurrence of these events was no longer surprising. When this happened the associative strength would approach asymptote, or the level of association where no surprise occurs.

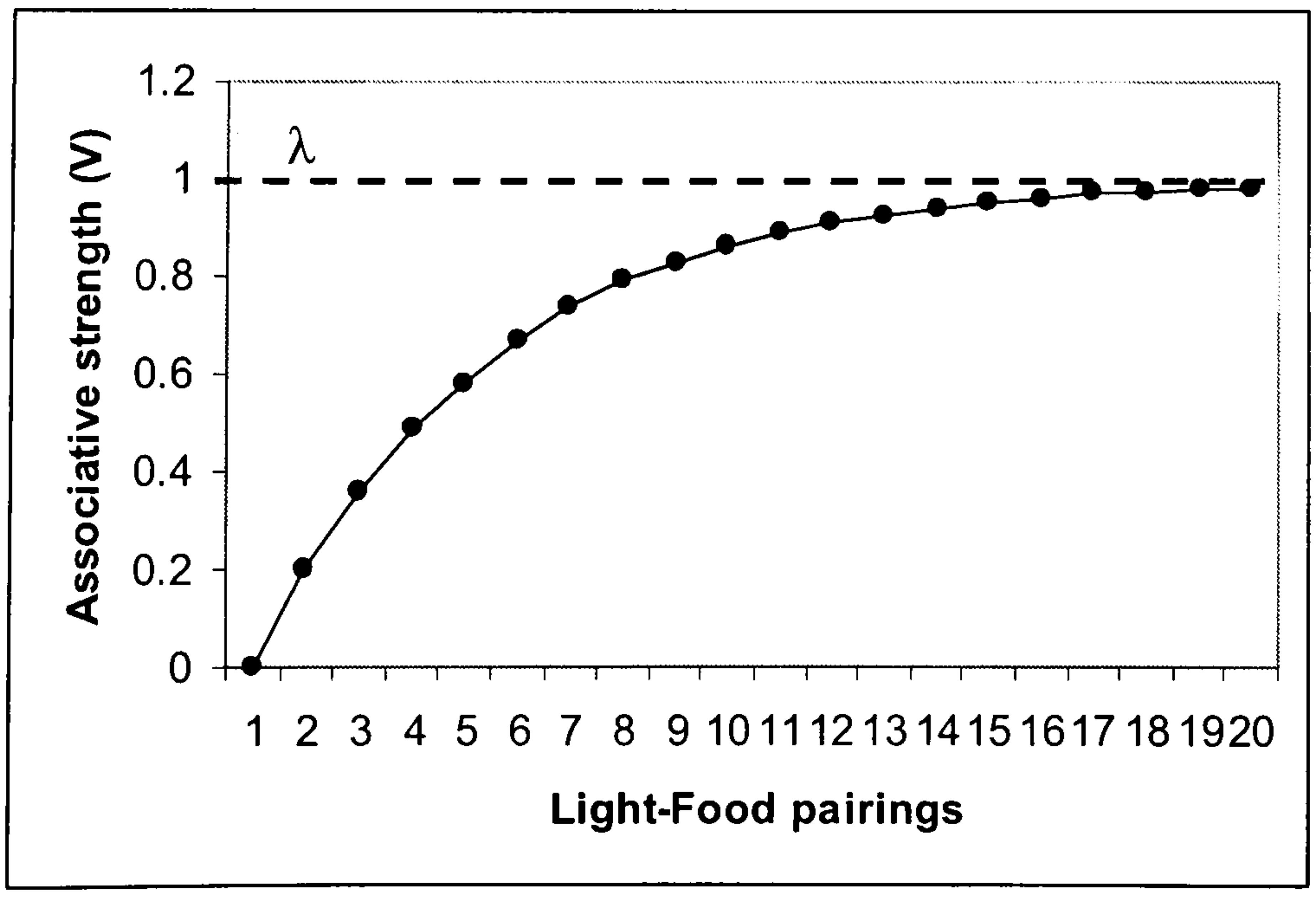

Figure 1.1 Associative strength growth as a function of the number of light-food pairings

Figure 1.1 shows the growth in associative strength between light and food across numerous pairings. As can be seen, there are larger increases in associative strength during the initial light-food pairings; eventually the light-food pairing is learned and associative strength approaches the asymptote level (λ) (Rescorla & Wagner, 1972). The change in associative strength (or V for predictive value) of each pairing of two stimuli can be expressed in the following equation form:

$$\Delta V = \alpha\beta(\lambda - V)$$

Where Δ means change, α and β are variables (always fractions) that relate to the salience of the two stimuli (light and food in the above example) and λ is the asymptote level. The term $(\lambda - V)$ is the key term in the equation; it has been described as the "surprisingness" of the unconditioned stimulus, or the food in the above example (Bouton, 2007). It is this term that determines whether any learning occurs on a particular trial. The α and β terms lead to different effects on the learning of the paired stimuli. β refers to the salience of the second stimuli (normally referred to as the Unconditioned Stimulus or US); when saliency of this factor increases, the overall asymptote level increases (Bouton, 2007). However, when the first stimulus (normally the Conditioned Stimulus; CS) has higher saliency, the learning of the relationship between the CS and US increases more quickly – in other words, fewer pairings need to occur to reach near asymptote level (Bouton, 2007).

The original context in which this theory was discussed was in terms of animal learning, in which rats or pigeons learned when a light comes on to expect a footshock or to receive a food pellet. Generally, the mechanisms of how we learn about the relationships between differing outcomes and predictors within our environment was effectively investigated based on the predictions this simple model generated. This model has aided in explaining the details of associative blocking and unblocking, extinction and inhibition and many other associative learning phenomenon (Bouton, 2007). The question for this thesis is how this model can aid in explaining the more complex learning situations which are found in language acquisition.

The original model postulated the two stimuli in the form of Conditioned Stimuli and Unconditioned Stimuli, but the principles of this learning apply to any relevant stimuli that have ordered relationships. Additionally, associative learning does not only occur between two stimuli; numerous stimuli can be associated in complicated, but predictable ways. The R-W model, as well as other associative learning models (e.g., Pearce-Hall model; Pearce & Hall, 1980), provide a framework for both predicting and interpreting results based on learning from many fields.

One phenomenon which R-W ideas have been influential in explaining is human and animal categorization. Categorization is, to our knowledge, a universal ability that has been found in every level of the animal kingdom (Pearce, 1997). In addition to differentiating objects found in their natural environment, more abstract categorising can be formed in many animals; for instance, the categorization of the character "Charlie Brown" by pigeons (Pearce, 1997). Several theories have been proposed to explain categorisation; including the prototype theory and feature theory. The prototype theory originated with experiments with humans and stated that exposure to category members resulted in a prototype of the category forming. This prototype can broadly be thought of as the central tendency of the category. The prototype becomes activated each time an individual category member is presented, but the response depends on how similar the individual category member is to the prototype (Pearce, 1997). This may relate to language in several ways; it has been found that prior linguistic experience impacts performance on many artificial language tasks (Lany, Gomez, & Gerken, 2007; see also Ziori & Dienes, in press for discussion of prior experience on implicit category learning). The possibility that linguistic prototypes may exist and that the amount of features or general similarity to the prototype may influence learning will be one line of inquiry in this thesis.

The feature theory is a more prevalent theory of categorisation which has been proposed for both human and animal categorisation. According to this theory categories are formed based on the shared features of the category members. The membership, or non-membership, of a particular item will depend on the extent to which the individual features match the general characteristics of the category. This is explained by the R-W model in terms of the associations between the individual features and the category outcome; for instance, if the category of "tree" is reinforced with food and any other picture is non-reinforced, the individual features that uniquely predict "tree" will increase in associative strength with each trial, while the features that are not unique to trees will not increase in associative strength. Categorisation of a "tree" category would occur when the associative strength of the "tree" features increased beyond the other, non-reinforced features. This would allow correct discrimination between "tree" and non-tree trials with both trained and novel presentations (Pearce, 1997).

The categorisation described in the above example was driven by the reinforcement – the food acted to increase the associative strength of the "tree" features to create the category. However, the task is much different within a language learning context; there is no direct reinforcement. However, Smith (2000) proposed an associative learning account of initial vocabulary acquisition that can be explained in terms of the R-W model in much the same way. This work aimed to explain the transition from slow acquisition of the first few words to the fast mapping stage of vocabulary acquisition, when learning occurs surprisingly rapidly. Very early word acquisition in young children displays a lack of specificity in the lexical categories, typically in the overgeneralizing of word meanings (e.g., calling bikes and aeroplanes "*cars*").

Children eventually overcome this initial problem by developing certain biases that aid in the learning of new words, specifically the shape bias for the acquisition of concrete noun names. Children learn to pay attention to the shape of the object, not the colour or texture, to aid in correctly specifying the objects in new lexical categories. Smith (2000) argued that the shape bias develops as the associative strength between new concrete noun words and the shape of the referent increases; this increase happens as children narrow down their overgeneralised lexical categories during the early stages of word learning. The shape of a car, not the colour or texture, is the most predictive feature in generalising the word for the vehicle their parents drive to all cars. Therefore, the associative strength between of the shape of concrete objects and the correct generalisation of the lexical label would increase beyond the other attributes. As the strength of this association increases, the child will begin to use the shape of a referent to determine what other objects should be labelled with the same word (Smith, 2000).

Smith (2000) tested this hypothesis by training 17 month old young children to pay attention to the shape of novel objects when determining which ones were a "zup" and which were "not a zup" (p. 72). The results were striking in that only young children trained to discriminate based on shape extended their knowledge of the shape bias in the experimental materials to their learning of real words. Children in the training condition showed a 166% increase in the number of concrete object nouns learned during the longitudinal experiment while the control children only showed a 73% increase. This indicated that Smith (2000) increased the associative strength between the shape of objects and their correct lexical categories via the training and this translated into the children using this useful bias to learn new words within the natural language environment.

This research indicated that the associative learning principles from the R-W model and from the associative learning accounts of categorization can aid in explaining language learning. The above account provides a convincing explanation for both the slow, effortful word learning at the start of vocabulary acquisition, but also the rapid increase in new word learning that occurs between 18 and 24 months, in which the amount of new words young children can learn increases exponentially. Once the associative strengths between the most useful features of the referent (shape for concrete nouns; texture for non-solid nouns) increases to a certain level, the young learners can then use this association to constrain further learning.

Although this language learning occurred on a much different time frame, the general principles between the "tree" and lexical categorical learning are very similar. Learners use the individual features of the referent objects that have the highest associative strength with correct categorisation to constrain the categories. The question of whether this same sort of explanation can account for more abstract grammatical categorisation will be investigated in this thesis.

Statistical learning has been proposed to be instrumental in the language learning process, for both the initial segmentation of the speech stream and also in the learning of higher order dependencies within language. Statistical learning and associative learning are not precisely the same thing; the main difference is that statistical learning models often postulate the involvement of cognitive processes in transforming the perceptual input into "a series of concepts" (Shanks, 2007, p. 297). The following section will review statistical language learning but will seek to integrate associative learning principles where possible.

1.3. Statistical Learning and Grammatical Categorization

Over the past ten years strong evidence has emerged in support of the view that children can learn language based on their environmental input. In particular, it now is quite clear that humans and infants are able to manipulate statistical regularities within language and can learn vast amount of information about language from these statistical regularities (e.g., Conway & Christiansen, 2005; Gómez, 2002; Saffran, 2001a, 2001b, 2003; Saffran, Newport, & Aslin, 1996). It appears that we are all, albeit implicitly, quite competent statisticians.

The first pivotal paper to demonstrate statistical language learning was Saffran, Aslin et al. (1996). This paper focused on the role of transitional probabilities between syllables in the word segmentation process. This process was illustrated with the sequence *pretty baby*; there is a high transitional probability between the syllables *pre* and *tty* and between *ba* and *by*. However, there is a far lower transition probability across the two words (*tty* and *ba*) which creates a natural boundary between the two words in statistical term. Within a continuous speech stream, this paper found that eight month old infants could differentiate "words", as defined by three syllable strings with high transitional probabilities, from part words, or strings of three syllables that occurred in the speech stream but crossed the word boundaries. This indicated that infants could make use of transitional probability information found within speech to segment words.

This experiment was monumental as this finding provided strong evidence that infants can learn from the statistical regularities of speech alone. The speech stream was synthesized so there were no prosodic cues that could aid in differentiating words; only transitional probabilities cues were available. Additionally, this learning can take place with as little input as 2 minutes exposure to the speech stream and without explicit instructions to listen to speech (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), indicating the robustness of the learning capacity of young infants. Although learning from the statistical regularities undoubtedly occurred, the possibility that this result is explained by higher associative strength between within word syllables and lower associative strength within between word syllables has not been investigated. However, the transitional probability and associative strength explanations make the same predictions; associative learning may simply be the method by which statistical learning occurs.

The demonstration that word segmentation can occur on the basis of statistical information led Peña et al. (2002) to investigate a statistical learning approach to the learning of nonadjacent transitional probabilities within speech streams. Adjacent dependencies aid in spotting word boundaries but nonadjacent dependencies relate to higher order grammatical structures, such as learning that *–ing* often comes a word or two after the word *am*. Peña et al. (2002) used speech streams embedded with trisyllabic units in which the nonadjacent transitional probability between the first and the third syllable was always 1, but the probability between the middle syllable and the first and third syllable was always .33. This structure can be denoted as A_iXC_i, in which A_i precisely predicted that C_i would follow after the intervening syllable. In this particular language there were three X syllables and three A_i - C_i combinations. The relationship between the syllables is illustrated in Figure 1.2.

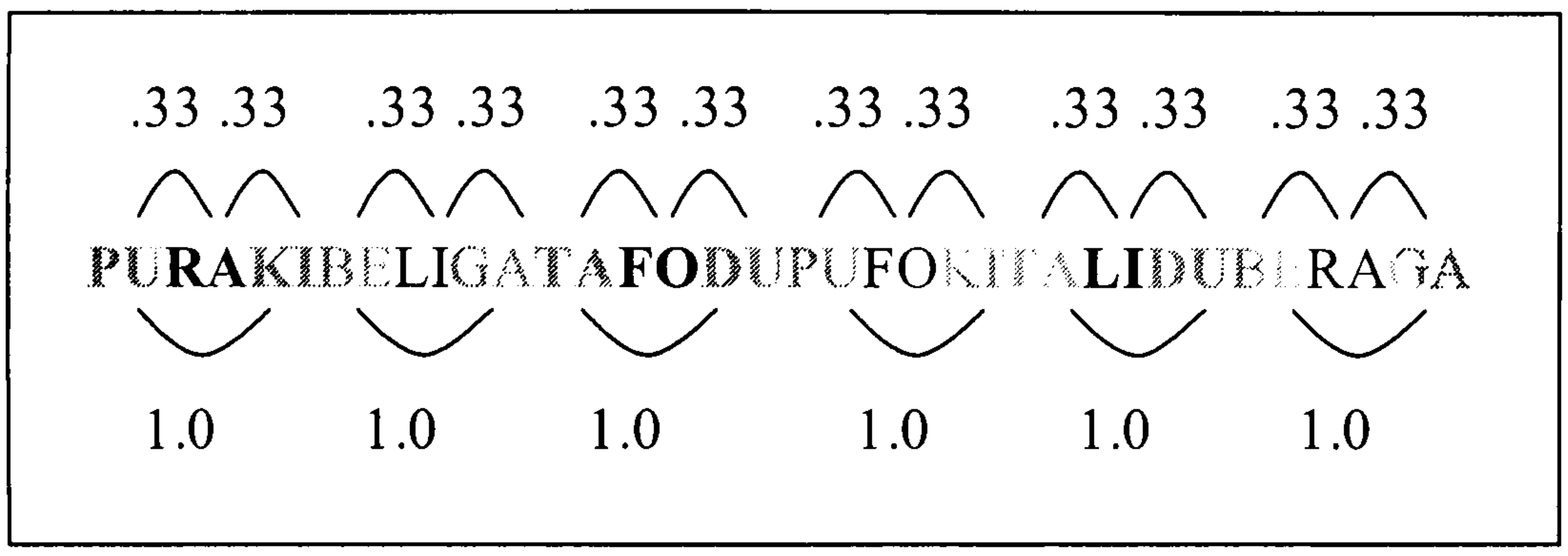

Figure 1.2 Transitional probabilities between syllables in Peña et al.'s training stimuli.

The results indicated that the participants were able to use the more distant transitional probabilities information to segment the speech stream into word like units, but were not, according to the authors, able to learn the more general structural information (Peña, Bonatti, Nespor, & Mehler, 2002). This was tested by comparing a "part word", or three syllables that cross word boundaries, with what was termed a "rule word", which had the correct A_i – C_i pairings, but a new X syllable (this syllable occurred in the training language, but never in the medial position). This inability to differentiate these two alternatives led the authors to conclude that "a computational mechanism sufficiently powerful to support segmentation on the basis of nonadjacent probabilities is insufficient to support the discovery of the underlying grammatical-like regularity..." (Peña, Bonatti, Nespor, & Mehler, 2002, p. 605).

To investigate this further, Pena et al. (2002) added subliminal gaps of 25-ms duration between the words on the premise that this would provide a strong segmentation cue and allow the overall grammatical structure to be learnt. Once the additional cues were added the participants were able to generalize the structure and preferred the rule word over the part word. The authors concluded that there are two different computational techniques at work; one is statistical in nature that can take advantage of transitional probabilities to segment speech, but cannot infer grammar-like regularities; the other an algebraic computational technique that is oriented to discover underlying regularities in language. This statement has provoked a fierce debate that is too detailed to cover in depth (e.g., Bonatti, Nespor, Peña, & Mehler, 2006; Bonatti, Peña, Nespor, & Mehler, 2005; Garzón, 2005; Onnis, Monaghan, Richmond, & Chater, 2005; Perruchet, Tyler, Galland, & Peereman, 2004)

What was not discussed by Peña et al. (2002) was the role of the middle syllables. What were the participants treating these middle syllables as? One possibility was that they were grouped together in a class of middle elements. The by-product of learning grammatical regularities within language could be learning what types of words can be integrated within these regularities. In regards to natural language, words can be grouped together into different categories depending on the words that regularly co-occur with them. For example, in the phrase *the cat is* the word *cat* co-occurs with the words *the* and *is*. However, many other words can occur between these two words, and the majority would share the same grammatical category; primarily they would be nouns. Even using just one of the framing words (i.e. *the cat*, or *cat is*) may be enough to group *cat* together with many other nouns. These grammatical category groupings would allow young language learners to generalise from knowledge of how one particular word can be used in certain grammatical contexts to other words of the same grammatical category. For instance, if a young language learner were to hear the sentence *The dog ran down the street* she should be able to infer that since she knows *dog* and *cat* are the same "type" of word the sentence *The cat ran down the street* is equally likely and plausible.

Gómez (2002) conducted two experiments, with adults and infants, which looked at how variance within these medial elements influenced the learning of nonadjacent dependencies. Gómez attempted to determine what strategy was used for learning nonadjacent dependencies. One possibility was that the higher order dependencies were embedded within the lower order dependencies. This means that as adjacent dependencies, such as A_1 - X_1 and X_1 - C_1, become grouped together higher order dependencies emerge by linking the adjacent pairs together ($([A_1 - \{ X_1] - C_1\})$). This predicted that increased learning would take place when the medial element had very low variability, as the higher the variability in the X elements the lower the individual transitional probabilities would be within the adjacent pairs. An alternative hypothesis was that high variability within the medial element will lead learners to focus on the invariant regularities, leading to higher learning when there was high variability in the medial element.

Speech streams were not used in these experiments; rather individual three nonword A_iXC_i sentences were used. There were three groupings of the invariant nonadjacent elements and critically, the manipulation of X variability included conditions with 2, 6, 12, or 24 possible X elements. After an 18 minute exposure phase the participants were given test strings, half of which were included in the training phase and half of which violated the legal word order. The participants had to decide which strings violated the word order and which conformed to it.

As expected the adult participants in all X set sizes displayed a preference for the trained items. However, there was a significant increase in correct responding with the set size 24 over the other X set sizes (Gómez, 2002). There was no difference in overall accuracy between the lower X set sizes. A second experiment exposed infants to the same stimuli for three minutes. During the test phase infants showed a novelty effect for the untrained test items in set size 24, but not for the lower set size numbers. This indicated that they could differentiate between the test items based on the non-adjacent structure only when there was a very inconsistent medial value. These results were consistent with the adult pattern and also indicated that statistical learning was automatically activated and flexible given the short length of exposure (as also found in Saffran, Aslin, & Newport, 1996).

These results do not support the idea that the nonadjacent dependencies were embedded within adjacent dependencies as learning would have been highest at the lowest set sizes, where the adjacent dependencies were most informative. However, the alternative hypothesis predicted that learning would gradually increase with increased X variability; this would allow the non-adjacent regularities to become more apparent. The results showed a rapid increase in learning with set size of 24, with no significant differences between the lower set sizes. Gómez (2002) proposed that only when the variability reached a critical level did the adjacent dependencies become a very poor source of information; once this critical variability level was reached learners would switch attention to the more meaningful nonadjacent regularities. This result was consistent with ideas from Elman (1993) who stated that much of language learning proceeds by the "starting small" principle; the learner first focuses on the smaller components of the language and gradually proceeds to focus on the more complex linguistic constructs. Furthermore, this ability takes time to develop: 12 months olds cannot track this non-adjacent dependency but by 15 months old this ability has developed (Gómez & Maye, 2005).

This study indicates that our natural inclination is to look towards adjacent dependencies for statistical information and only when this analysis becomes unreliable do we analyze higher order nonadjacent dependencies. In the larger picture, these results support the idea that language learners are not passive learners but are active learners who search for the analysis that provides the most useful information. However, "active learners" does not necessarily imply explicit learners. As with the segmentation findings, the above finding may be interpreted as only when the associative strength between adjacent elements becomes significantly less information than the associative strength between non-adjacent elements will focus be given to the non-adjacent dependencies. However, the conceptualization is the only difference when interpreting the results in terms of associative learning versus transitional probabilities.

These experiments provide more support for the idea that grammatical regularities can be used to infer grammatical categories within words. Gómez (2002) only found learning of nonadjacent regularities when there were numerous medial elements in the language. This result leads to two complementary implications for language learning and grammatical categorisation. The first is that in order for nonadjacent regularities such as *am* runn*ing* to be noticed many different verbs need to be used within the *am* and *ing*, which is seen in natural language. The second implication is that these regularities sort words into grammatical categories. Once words are grouped according to the syntactic regularities into grammatical categories, such as nouns and verbs, generalization can occur. This would allow the language learner to generalize all words in a particular category to all syntactic contexts in which they have consistently heard words of that category, even if they had not heard the specific word/syntactic regularity paired together.

Onnis, Monaghan, Christiansen, and Chater (2004) extended this work by investigating how nonadjacent regularities were learned when there was zero variation in the middle elements. According to their Variability Hypothesis, nonadjacent invariant structure should be easy to learn both when there is no variation and when there is large variation in the medial elements. In the former case, the nonadjacent regularities would vary against the invariant middle item whereas in the latter case, the nonadjacent regularities stand out against the ever changing middle item. These are both cases of noticing change vs. no change. Nonadjacent regularities should be much harder to learn when the number of medial elements is somewhere in between these two extremes.

The authors tested this hypothesis using a language similar to Gómez (2002) with three variability conditions: large (24 middle items), small (two middle items) and zero variability (Onnis, Monaghan, Christiansen, & Chater, 2004). There were three sets of the nonadjacent invariant elements. The test items were created so that the adjacent transitional probabilities were equal between grammatical and ungrammatical strings; grammaticality could only be differentiated by looking at the nonadjacent regularities. In addition, half of the test items contained new X words as a test for generalization. As can be seen in Figure 1.3 learning was highest in the zero variability condition whereas learning dropped in the small variability condition and increased again in the high variability condition, in line with the hypothesis. These results not only showed that nonadjacent dependencies can be detected in both conditions of zero and very high variability, but also that generalization to new X elements can occur.

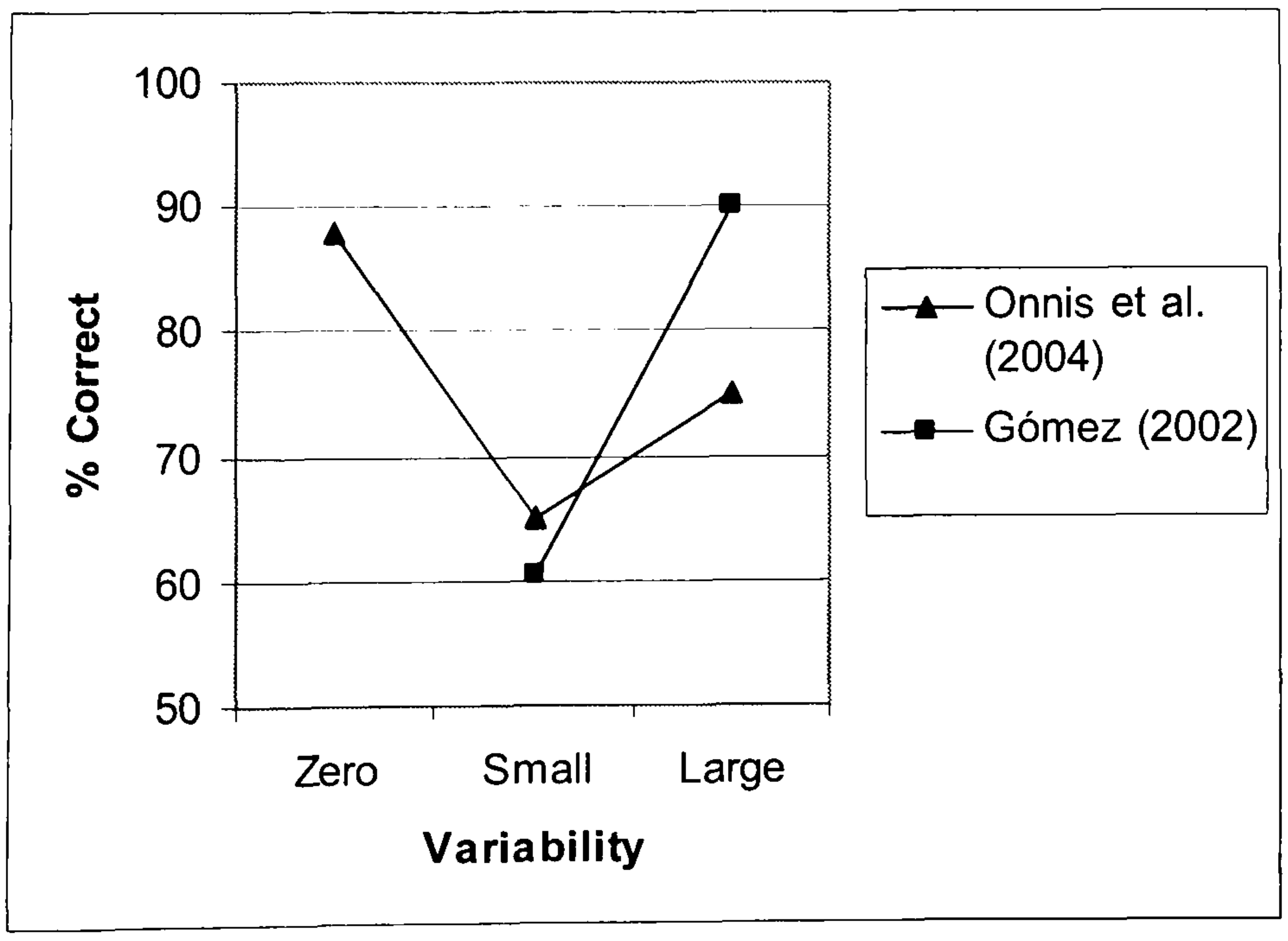

Figure 1.3 Graph from Onnis et al. (2004) including results from Gómez (2002).

In a second experiment, Onnis et al. (2004) tested whether this effect is present in continuous speech. Peña et al. (2002) concluded that speech segmentation and generalization of nonadjacent regularities could not be done via the same statistical learning mechanism. However, Peña et al. used three A_i - C_i non-adjacent dependencies and three X values in their language. According to the Variability Hypothesis, this language is predicted to be hard to learn, as there was no increased variability in either the X elements or the A_i - C_i dependencies to aid learning. Onnis et al. used a similar speech stream as in Peña et al. but with either three or 24 X items. Their results supported the previous finding in that there was no evidence of generalization in the low variability condition (41.9%). However, with the high variability condition participants preferred the rule word over the part word 63.3% of the time, which was significantly higher than chance level and the low variability condition.

These results again highlight the importance of variability in the process of both generalizing words to grammatical categories and learning linguistic structural regularities. The finding indicated that generalization can occur in unsegmented speech apparently with the same statistical computation mechanism as is used to segment speech. The key is simply to provide enough variability for the invariant structural regularities, such as *am* go*ing*, to be perceived. This then allows *go* to be categorized with all other verbs heard in between *am –ing*, such as *jog, cook, eat, write, read, climb, drink, close, open,* etc.

This finding raised an interesting question about the Peña et al. (2002) result; specifically, how the non-adjacent structure was learnt when subliminal segmentation gaps were inserted into the language. Onnis et al. (2005) investigated whether phonological regularities within the Peña et al. language could have influenced this learning. A series of experiments were conducted that investigated the phonotactic information within the language. Specifically, the beginning and ending syllable in Peña et al. experiments all began with plosives (e.g. /b/, /p/, /k/, /g/, and /d/), whereas the middle syllables always began with a continuant (e.g. /f/, /l/, and /ɹ/). When comparing *plosive-continuant-plosive* words to part words that either began with a plosive or with a continuant, there was a larger advantage for words when compared to the continuant part words. This indicated that participants were less able to differentiate the words and part words when the part word beginning syllable was a plosive.

When the phonological structure of the words were changed to *continuant-plosive-plosive* the participants preferred part words than the words; all part words now began with a plosive. This finding indicated that phonological properties may have more influence than statistical regularities. Furthermore, when the *plosive-continuant-plosive* pattern remained in the language, but the non-adjacent structure was broken the participants still had a preference for the words over part words when the part words began with a continuant. When a similar *continuant-plosive-continuant* pattern was examined, there was no significant preference for words. In other words, participants could differentiate words based on the phonological pattern alone when the words began and ended with a plosive syllable, but not when the words began and ended with a continuant syllable.

These results point to the importance of phonology when investigating non-adjacent dependency learning, and account the lack of non-adjacent dependency learning found in Newport and Aslin (2004) when plosives began each syllable. The learning of adjacent dependencies can progress relatively smoothly in a variety of modalities (Creel, Newport, & Aslin, 2004; Kirkham, Slemmer, & Johnson, 2002; Saffran, Johnson, Aslin, & Newport, 1999; Saffran, Newport, & Aslin, 1996) but these results indicated that non-adjacent dependencies may be modulated by the phonological patterns of the individual components of the dependencies. More generally, these results are indicative of how statistical learning interacts with other linguistic cues, specifically phonology in this case. Although not specifically investigated in this study, the role of phonology in the learning of word classes will be discussed in detail below.

The above evidence provides the underlying support for the idea that grammatical categories can be acquired via grammatical structural information. It clearly outlines that humans have the analytic capability to use the regularities within language to aid in the learning of language. The literature investigating exactly how syntactic regularities, or distributional cues, have been found to be used in the acquisition of grammatical categories will be discussed next.

1.4. Distributional Information and Grammatical Categorization

The importance of "marker elements", or distributional cues, in language has long been established (Braine, 1966; Green, 1979). Artificial languages without these marker elements are virtually impossible to learn, but when there are consistent marker elements present, languages are readily learnt. The importance of distributional cues in the acquisition of grammatical categories was advocated by Maratsos and Chalkley in 1980. This paper put forward the theory that grammatical categories are learnt by the intercorrelations between the distributional contexts in which they are used. More specifically, they state "form [grammatical] classes and other syntactic classes such as gender class are not, at least for awhile, primitives on which the child bases further analyses, but an *outcome* of the analysis of more primitive sequences and semantic properties" (Maratsos & Chalkley, 1980, p. 132). For example, if we hear a new word in a certain combination, such as "John glixed Mary" (p. 133) then numerous other possible sentences can be predicted from this one sentence such as "Mary was glixed by John, for John to glix Mary was absurd, the glixing of Mary by John is commendable" (p. 133). As stated above, the idea that grammatical categories were an outcome of the correlations between different grammatical combinations was advocated but this assertion was only supported by secondary evidence based on the types of errors children make during language acquisition. Since this paper was published there have been numerous papers published that provide evidence both that distributional information is useful and usable in language acquisition. There are two main types of evidence for the usability of distributional information in language acquisition: corpus analyses and experimental evidence.

1.4.1. Evaluating the usefulness of distributional information with corpus analyses

Analysis of child directed speech has become a powerful tool in language acquisition research; the fine grain details of languages can now be analysed to determine what types of information are reliable and possibly important for language learning. This aids in determining the viable, reliable language regularities; in other words, what are *useful* cues within language. Corpus analysis has been widely used in investigating the types of cues from across grammatical phrases, termed distributional information, that are best suited to aid grammar learning. In particular, the types of distributional cues that may aid in the acquisition of grammatical categories have been investigated. For instance, the word *the* often indicates that a noun will follow. Generally, it is not a matter of whether distributional information aids grammatical category acquisition but what type of distributional information is the most informative.

Redington, Chater and Finch (1998) investigated several distributional contexts and evaluated their effectiveness in grouping words of the same grammatical category together. They used only child-directed speech from the CHILDES database which was comprised of approximately 2.5 million words from 3,000 different speakers. When looking at preceding word cues, it was found that the immediately preceding (aX, where a is the distributional cue and X are the grouped words) and the second preceding word (a_X) led to groupings that were more consistent in terms of the grammatical category of the X word than would be expected by chance. In other words, the aX and a_X distributional cues were able to group X words of the same grammatical together. When looking at succeeding word cues, only the immediately succeeding word (Xb, where b is the succeeding distributional cue) provided any relevant information that helped group X words of the same category together. However, the accuracy of categorizing words of the same grammatical category together was by improved by combining the immediate preceding and succeeding words, which was closely followed by the two preceding and two succeeding words. Overall, these results indicated that preceding and succeeding distributional information was useful in grammatical categorization, although the preceding information was more informative than succeeding information. Furthermore, the combination of preceding and succeeding information was found to be the most accurate cue in creating grammatical categories.

Further analyses were conducted by Mintz, Newport, and Bever (2002) using even larger distributional "windows" than Redington et al. (1998). They analysed four corpora and based their analyses on the 200 most frequent words in each corpus. They analysed three distributional "window" sizes; either just the immediately preceding and succeeding word, two words preceding and succeeding or eight words preceding and succeeding. Mintz et al. reported their results in terms of how nouns and verbs were categorized separately. The only distributional contexts within phrase boundaries were analysed, as research indicates that infants are able segment speech into different phrases (e.g.,Gerken, Jusczyk, & Mandel, 1994). For both nouns and verbs categorization was at above chance levels for all distributional contexts. For nouns, there were no great differences across the three distributional contexts, although the two word "window" categorized the nouns less accurately than the one and eight word distributional contexts. For the verbs, there were no great differences between the three distributional contexts. These results do not replicate the drop in categorization accuracy between one and two word distributional context found in Redington et al. (1998) which may have been due to differences in the procedure and analysis. The main conclusion from both studies was that the local context of the preceding and succeeding word provided the best cue to group words of the same grammatical category together.

The combination of the immediately preceding and succeeding words (aXb or "frequent frames") was further investigated by Mintz (2003) using six corpora. Mintz investigated the 45 most frequent frames in each corpus. An example of a frequent frame was "you_it"; all of the words that occurred between the "you" and "it" were grouped into one category. Categorization resulting from these frequent frames was found to be highly accurate at categorizing words of the same grammatical category together.

These three studies provide evidence that distributional frames are very good at categorizing words of the same grammatical category together. What should be noted, however, is exactly how much of the entire corpus was included in the analysis; the Mintz (2003) analysis only included an average of 6% of the corpora. Although these frames do produce quite accurate categories, they categorize a small amount of the total linguistic input. This reduces the viability that these frames could play a major role in the acquisition of grammatical categories. Additionally, it is not clear exactly where the distributional information comes from; whether it is mainly from the preceding word or from the succeeding word, or whether it is truly a combination from the two words.

This question was directly addressed by Monaghan and Christiansen (2004) who directly compared the Mintz frames analysis (aXb) with a bigram (aX) distributional analysis. They found comparable results to Mintz for the trigram frame analysis, although the accuracy of the grammatical category groupings was slightly lower. The bigram analysis (aX) had less accurate grammatical category groupings; however, this was due to more noise within the categories. This increased noise was due to the bigram analysis classifying more total words than the trigram frame analysis. In total, the trigram analysis classified 394 words whereas the bigram analysis classified 1930 words. Overall, the trigram analysis produced more accurate groupings but the bigram analysis was more proficient at classifying large amounts of words.

Monaghan and Christiansen (2004) next trained a neural network to classify words on the basis of trigram frame or bigram distributional information. The bigram network performed better than the random baseline and correctly classified 52.4% of the words. The network did particularly well on classifying verbs (83.9%) but was a bit lower at classifying nouns (66.3%). The trigram model performed at random baseline level; both the trigram and random baseline model were only able to classify words at the utterance boundary.

The lack of categorization ability with trigrams was investigated further by allowing the information from the initial and ending bigram within the frame (i.e. aX and Xb from the aXb frame) to be used separately as well as combined. They also tested a model that had information available from the two preceding words (ABx). It was found that both models performed above the baseline measure with the composition trigram model categorized 69.4% of the words correctly while the ABx model categorized 56.3% correctly (Monaghan & Christiansen, 2004). These models indicated that trigram information provided the most useful cue to grammatical categorization but only when beginning and ending bigram information was also available. Therefore, the advantage for "frequent frames" was due to the initial and ending bigram occurring at the same time to predict the middle word.

Monaghan, Chater and Christiansen (2005) further investigated the use of bigrams in grammatical classification. They took the 20 highest frequency words from the CHILDES database and analysed the likelihood of co-occurrence with the 5000 most frequent words, with the context word (or distributional cue) preceding the target words (aX). The context words were divided according to the highest proportion of grammatical category. When seven noun context words and five verb context words were entered into a discriminant analysis it was found that nouns were classified correctly 93.7% of the time and verbs 31.1% of the time. This interesting discrepancy between nouns and verbs was replicated in Christiansen and Monaghan (2006). This noun/verb difference may be due to the increased variability of local contexts in which verbs may occur (Monaghan, Chater, & Christiansen, 2005). For instance, verbs usually occur after a variety of nouns, whereas nouns typically occur after determiners; this would lead to more reliable preceding distributional contexts for nouns rather than verbs.

Further analysis was conducted with frequency taken into account; it was apparent that classification differed drastically for the different frequency groups. For the 1000 most frequent words 85.8% of the words were correctly classified which contrasts with only 69.9% of the words correctly classified for the 1001-2000th group; classification further decreased to 53.5%, 52.4%, and 50.4% for the 2001-3000th, 3001-4000th, and 4001-5000th most frequent group (Monaghan, Chater, & Christiansen, 2005). This analysis indicated that the usefulness of bigram distributional information in grammatical categorization reduced with lower frequency words. This makes intuitive sense as higher frequency words were more frequently paired with distributional cues; the more often the X words/distributional cue were paired, the higher the likelihood the X words were grouped with words of the same grammatical category.

The above corpus analyses indicate that distributional information is a highly informative and useful cue for sorting words into grammatical categories. However, these studies do leave several issues open; the first of which is whether trigram or bigram distributional information are the most useful in categorization. It is certain that trigram analyses provide more accurate categorization but they do so at the cost of coverage. Bigram analyses are less accurate but are able to categorize more total words than trigram analyses. The finding that trigram distributional cues are only useful if the two bigrams can be considered separately brings up the possibility that the trigram distributional cues are only two bigrams considered simultaneously. To sum up, the adjacent co-occurrences within the bigram analyses appear to be at the basis of grammatical categorization from distributional cues.

These corpus analyses have investigated specific word cues; a slightly different approach was used by sentential minimal pairs in simulations (Cartwright & Brent, 1997) with similar results. Whether from word or sentential context it has been shown that distributional information is useful in grammatical categorization, but not that language learners are able to use this information. The research discussed below outlines the literature that shows that we can use distributional information within language learning tasks, which has strong implications that distributional information is an important grammatical category cue to the young language learner.

1.4.2. Evaluating the usability of distributional information with experimental evidence

The distributional cues described in the corpus analyses above can be thought of as either preceding or succeeding marker cues. However, many other variables have been found to be influential in categorical learning (e.g., Mori & Moeser, 1983). The role of frequency within marker elements was investigated by Valian and Coulson (1988) by looking at how very frequent marker words aid grammatical categorization. Two main properties were set out as being vital in order for distributional cues to be effective; the first was that the marker must be reliably associated with the paired structure (in this case a specific grammatical category) and the second was that it must be of sufficient frequency (Valian & Coulson, 1988).

In the artificial language each sentence had four words that were split into two phrases that corresponded to two word categories (i.e. *aAbB* with *a* and *b* as the marker words and *A* and *B* as the category words). The marker words were manipulated based on frequency; there were two marker elements (one per category) and 12 content words (six per category) for the high frequency dialect but four marker elements (two per category) and six content words (three per category) for the low frequency dialect. There were 24 test sentences; 12 of these sentences contained one of four types of errors. Types 1, 2 and 3 had obvious syntactic violations (e.g. *AabB*, *aABB* and *aBbB*). The Type 4 violation had incorrect marker/category word pairings in both the phrases (bAaB) (Valian & Coulson, 1988). The participants had to decide whether the test sentences were similar or dissimilar to the training sentences.

The results indicated that the participants who received the high frequency marker language made far fewer errors whereas participants in the low frequency language only showed a small decline in error rate. However, closer inspection of the errors indicated that the participants were able to tell that Type 1 and 2 sentences were dissimilar; this can be explained due to a violation in the *marker-content-marker-content* structure that characterized all of the training items. The participants in the low frequency group could not respond correctly to the Type 3 and Type 4 error test sentences, whereas the high frequency marker word sentences were able to differentiate the correct test sentences from the Type 3 and 4 test sentences. In order to reject Type 3 and 4 test sentences the participants must have linked the marker word with the content words, and thus formed separate categories of A and B words. This was only possible with the high frequency marker words. This result is in line with the Variability Hypothesis discussed earlier – only when there is sufficiently high variability between marker elements and the content element should there be any category learning.

While Valian and Coulson (1988) investigated categorization using bigram distributional cues, Mintz (2002) followed up his corpus investigations of "frequent frames" with a trigram artificial language study. The language had an aXb structure, and there were three a_b frames and four X words. Each of the X words occurred within all three a_b frames; categorization of the X words should occur due to the trigram a_b distributional cues. There were two other a_b frames; the partial paradigm a_b frame contained only three of the four X words and the alternate paradigm a_b frame contained three new medial words. Categorization was tested by using the remaining category word within both the partial and alternate paradigm a_b frames; the partial paradigm test sentence should be endorsed more frequently as similar to the artificial language as the other three category words occurred within the partial paradigm distributional context. This is precisely what Mintz found; the participants were able to use the trigram distributional context to categorize the four words together. This result provides evidence that trigram distributional cues can indeed induce categorization, but it does not separate whether the learning occurred through the a_b unit, or from the separate aX and Xb bigrams.

Most studies discussed thus far have used adult participants and artificial languages. While this method is invaluable in determining what kinds of information within the linguistic environment adults are able to learn from, it does not address whether infants are able to learn from the same information. Gómez and Lakusta (2004) investigated this issue by exposing infants to an *aXbY* language, in which they had to learn that *a* was associated with disyllabic *X* words and *b* was associated with monosyllabic *Y* words. Different disyllabic *X* and monosyllabic *Y* words were used during the test. Using a head turn preference it was determined that the infants could differentiate between test strings that matched the correct and incorrect marker word/category word pairings. This result showed not only an ability to abstract the general features of the category words and associate them with the markers, but the ability to extend this knowledge to novel instances; in other words, to generalize to new category members.

Gerken, Wilson and Lewis (2005) further investigated categorization in infants using masculine and feminine Russian nouns. The noun subgroups were based on distributional information in the form of case markings. There were two conditions: the first condition only contained the case marking cues to indicate gender status and the second in which half the words also contained a productive diminutive inflection, thus making these nouns "double marked" (p.256). In both conditions, there were two different case markings for each gender category; *-oj* and *–u* for feminine nouns and *–ya* and *–yem* for masculine nouns.

After familiarization, the infants were tested using head turn preference on grammatical and ungrammatical nouns. In the double marked condition, these test items only contained the case marking cue to gender status. Results indicated that only when there were two cues to the category membership in the familiarization material were the infants able to categorize the nouns into feminine and masculine groups. This supports previous research using Russian gender categories in adults, in which it was found that learning only occurred when there were partially correlated cues in the familiarization language, as was found in this study (Gerken, Wilson, Gómez, & Nurmsoo, 2002 as cited in Gerken et al., 2005). More importantly, this study demonstrated that infants could make use of naturally occurring distributional cue to separate naturally occurring masculine and feminine noun subcategories.

The last two studies highlight the possible importance of associative learning in the category learning process. Both studies show an ability to use associated cues to separate words into two categories. This is especially striking in Gerken et al. (2005) which shows that cues amongst only some of the category words was sufficient to induce categorization that included all the category members. These addition cues increased the associated strength of the marked nouns with case markings; this increased associative strength amongst a subset led to quicker learning of the whole category.

Billman (1989) investigated how more complex correlated cues could influence categorical learning. The "structured" language consisted of correlated cues between the beginning and end vowels (in *cvcvc* nonwords) and referent shapes using a written language. Each phrase consisted of four words; three of the words referred to corresponding referent shapes. In the fully correlated language the beginning vowel predicted the end vowel, the end vowel predicted the type of shape the word referred to, and the shape of the referent predicted the stem vowel (Billman, 1989). There were three other conditions in which each of these separate correlated cues was isolated. One object in the referent field showed an action in relation to the other objects in each scene. The last phoneme of the words indicated whether the word was referring to the agent, object or bystander in the scenes.

The participants were tested on the items they had learned during training, new items and partial presentation judgements (the pictorial representation, stem or ending of the words was missing) in both correct and incorrect correlated cue contexts. The task was to judge whether the sentences were "appropriate and fitted in" with the training sentences (Billman, 1989, p. 138). It was found that all conditions performed quite well on the old items but the correlated cues condition had a large advantage on the novel item and partial presentation judgements. Therefore, the acquisition of the trained nouns was aided in the correlated cues condition as these participants were able to use the associations between all three cues to learn about the general structure of the nouns which aided performance in the partial presentation decisions. Additionally, the increased ability with the novel items indicated that the correlated cues aided the integration of new nouns into this language.

This relates back to the proposal put forward by Maratsos and Chalkley (1980) that knowledge of a new word in one context can lead to generalization to other contexts and ultimately inclusion in a category of words that can be used in all correlated linguistic contexts. These results support this view that by providing evidence that correlated distributional contexts can lead to generalization of new category members. More broadly, the reviewed studies indicate that distributional information is both a *useful* and *usable* cue to aid in the grammatical categorization process. Although distributional information plays an important role in this process, the role phonological information plays in grammatical categorisation will be considered next.

1.5. Phonological Information and Grammatical Categorization

In a review of the role of phonological information in grammatical categorisation Kelly (1992) came to the conclusion that the virtual exclusion of this rich source of information from previous accounts of grammatical category acquisition was unwarranted. Previous research had focused almost entirely on syntactic and semantic information while downplaying the possible effect of phonological cues due to the belief that the nature of this information is arbitrary (e.g., Maratsos & Chalkley, 1980). Contrary to this, Kelly demonstrated that phonological cues are not arbitrary but that many cues systematically differentiate classes of words, such as the open and closed class word distinction as well as differences between nouns and verbs.

Two main examples of how phonological cues have been shown to be good indicators of grammatical category are the placement of stress[1] and number of syllables in words (Kelly, 1992; see Kelly & Bock, 1988 for specific evidence of stress differences). Generally, nouns have trochaic stress and more syllables than verbs, which tend to have iambic stress. Word duration of monosyllabic words also differentiates nouns and verbs with nouns tending to be longer than verbs. This is possibly due to nouns occurring in utterance final positions more often than verbs. Adjectives and verbs are often differentiated on the vowel epenthesis in the –ed inflection. Verbs do not usually have the vowel pronounced whereas adjectives usual do, which creates an extra syllable at the end of the word for adjectives but not verbs. Voicing of consonant phonemes also differentiates open and closed class words. For example, the word initial *th* is generally pronounced voiced with closed class words (i.e. *the, that,* etc.) but unvoiced with open class words (i.e. *thing, think,* etc.). There is also some initial evidence that related noun-verb homonym pairs often differ only on the voicing of the final phoneme. See Table 1.1 for a summary and examples of these phonological cues.

---

[1] Stress is a cue that can be considered either as prosodic or phonological. Some authors refer to it as a prosodic cue, but others, such as Kelly refer to it as a phonological cue. For the current purposes, the status of stress as a phonological or prosodic cue will depend on the opinion the authors of the literature currently being reviewed.

Table 1.1 Phonological cues relevant to grammatical category differentiation and examples.

Phonological Cue	*Example*
Stress	Fructose (1st syllable stress) Betray (2nd syllable stress)
Syllable Number	Evolution (4 syllables) Learn (1 syllable)
Phoneme Number	Handkerchief (9 phonemes) Freeze (4 phonemes)
Word Duration	Coach (Open class noun – longer duration)* Coach (Open class verb – shorter duration)
Vowel Epenthesis in –ed inflection	Learned (Verb) Learned (Adjective)
Voicing	Teeth (Noun – voiceless "*th*")* Teethe (Verb – voiced "*th*")
Vowel Type	Spa (Back vowel) Meet (Front vowel)

* Noun – Verb homonyms often have different word duration with the noun version being longer.
Some of these examples were taken from Kelly (1992).

Although these grammatical category differences are documented, it must be shown that language learners and users are able to manipulate these differences to aid in the learning and processing of language. In particular, English speakers have been shown to be sensitive to a difference in vowel types within high frequency words (Kelly, 1992). In particular, nouns are categorised faster when they contain back vowels (which are typical of high frequency nouns) than when they contained front vowels. High frequency English verbs are more likely to contain front vowels than back vowels, and are subsequently categorised faster when they contain front rather than back vowels. However, these differences disappear with lower frequency words indicating English speakers are not only sensitive to this noun/verb difference but are aware that it only occurs in high frequency words. It has also been found that both syllable number and the placement of stress within a word influence how acceptable the word is in a double object structure; monosyllabic words are more acceptable than polysyllabic words, but polysyllabic words are more acceptable when they have stress on the first syllable (Kelly, 1992).

The evidence that Kelly (1992) reviewed clearly indicates that phonological cues are a rich sort of information for differentiating different types of words and possibly categorising similar sounding words together, thus greatly weakening the claim that the phonological aspects of language are arbitrary. There is even evidence that male and female names can be differentiated on the basis of phonology (Cassidy, Kelly, & Sharoni, 1999). Further evidence in support of the importance of phonological cues comes from Monaghan et al. (2005) who investigated how well these phonological cues differentiated open and closed class words as well as nouns and verbs within the CHILDES corpus. The phonological cues used in the analysis were shown, in previous literature, to have some promise in possibly differentiating either grammatical classes (nouns and verbs) or open and closed class words. The specific phonological cues used are listed in Table 1.2. The 5000 most frequent words were analysed.

Table 1.2 Phonological cues and differentiation status; "+" means there was a significant difference between the classes and "-" indicated there was not significant effect (Monaghan et al., 2005).

Phonological Cues	*Open/Closed Differentiation*	*Noun/Verb Differentiation*
Phoneme length	+	-
Syllable length	+	+
Presence of stress	+	-
Stress position	+	-
Onset complexity	+	+
Syllabic complexity	+	+
Reduced vowels	+	+
Reduced 1st vowel	+	-
-ed inflection	-	+
Coronal	-	+
Initial dental	+	-
Final voicing	-	-
Nasal	-	+
Stressed vowel position	-	-
Vowel position	-	+
Vowel height	-	+

It was found that numerous cues differentiated open and closed class words (see Table 1.2 for details) but that the individual cues were not very reliable (Monaghan, Chater, & Christiansen, 2005). However, when all cues were entered into a discriminant analysis 76.4% of the words were correctly classified, showing good discrimination of open and closed class words. Similarly, nouns and verbs were differentiated by many of these phonological cues (see Table 1.2). When all cues were entered into a discriminant analysis 63.4% of the nouns and verbs were correctly classified. However, when the words were broken down by frequency it appeared the phonological cues could more accurately predict the less frequent words than the most frequent words. For the 1000 most frequent words classification was at 60.1% but for the least frequent words classification was at 67.3%. This indicated that within category phonological cues were more consistent with lower frequency words than high frequency words. This was perhaps due to the lack of phonological reduction that occurs with higher frequency words over time. Without this phonological distortion low frequency words were better exemplars of the phonological characteristics of a category.

This study demonstrated that phonological information is useful in differentiating grammatical categories. Although Monaghan et al. (2005) and Kelly (1992) have shown that these differences exist within languages there are two interrelated issues which are not fully resolved. The first is whether language learners are sensitive to all of the cues discussed above and the second is whether these phonological cues are used to aid in the grammatical categorization process during language acquisition.

Shi, Werker, and Morgan (1999) looked at the perceptual abilities of newborns to differentiating lexical (nouns, verbs, adjectives, and adverbs) and grammatical (articles, prepositions, and auxiliaries) words. These two overarching categories would need to occur before finer grain differentiation could occur. Lexical and grammatical words had been found to differ on numerous intercorrelated acoustic and phonological measures across many languages (Shi, Morgan, & Allopenna, 1998). For instance, grammatical words tend to be acoustically and phonologically reduced in comparison to lexical words.

Newborn infants (1-3 days old) were presented with lists of words composed of either lexical or grammatical words. After habituation the newborns were presented with a new list which was either lexical or grammatical words. It was found that the infants were significantly more interested in the category of words they were not exposed to during the familiarization phase indicating they could differentiate the two categories of words. This result was obtained in infants whose mothers spoke a range of languages indicating that the ability to perceptually differentiate these two categories of words was not dependent on prenatal language exposure (Shi, Werker, & Morgan, 1999).

The above finding provides evidence of the underlying basis for grammatical category differentiation but further research was lacking in determining how phonological information could be used in grammatical categorisation. Kelly (1988) investigated how patterns of stress differentiated nouns and verbs. As mentioned above, nouns are more likely than verbs to have stress on the first syllable (e.g. parcel; trochaic pattern) and verbs are more likely to display an iambic pattern (e.g., convince; second syllable stress). Kelly investigated whether stress differences influence which words shift from one grammatical category to another grammatical category. For example, a noun that has an iambic pattern of stress may be more prone to become a verb, such as the word *canoe*, which originated as a noun with an iambic stress but subsequently was extended to verb usage (Kelly, 1988). Conversely, the verb *challenge* originated as a verb with a trochaic stress but was extended to noun usage. Historically, trochaic verbs are more likely to be extended to noun usage and iambic nouns to verb usage (Kelly, 1988).

Furthermore, Kelly (1988) demonstrated that people are sensitive to the stress differences between nouns and verbs. When people were told to use novel disyllabic nonwords in a sentence it is found that they are more likely to use the iambic stressed nonwords as verbs (48% of iambic used as verbs, compared to only 31% of the trochaic nonwords). However, the use of nonsense words did not allow a direct conclusion to be drawn about how words are extended into new grammatical categories.

A further study was conducted that addressed this issue. The participants were presented with two semantically related words (either two nouns or two verbs) in which one had trochaic stress and one iambic stress. The participants were informed about grammatical category shifting and had to pick one of the two words to use either as a verb when the target words were nouns, or as a noun when the target words were verbs. The hypothesis was that the iambically stressed nouns would be more likely to be chosen for verb usage and the trochaically stressed verbs would be used more as nouns. This is exactly what was found; 60% of the iambic nouns were used as verbs and about 66% of trochaic verbs were used as nouns (Kelly, 1988).

This study provided evidence participants were sensitive to the stress differences between disyllabic nouns and verbs. Many participants reported making their decision on which word "sounds best" but could not say directly that this had anything to do with stress patterns within the words. The authors translate this simple strategy to "sounds like other exemplars of the new grammatical category and/or an abstract prototype of that category" (Kelly, 1988, p. 354). This explanation is surprisingly similar to the prototype theory of associative categorization discussed earlier. However, another explanation may be that the stronger associative strength in general between the differing stress pattern feature and the category status directly influenced participants' judgements. Unfortunately, this study does not demonstrate that this cue is usable in forming grammatical categories during language acquisition, although the results suggest that this is plausible, specifically as stress patterns are known to aid infants in segmentation from an early age (Echols, 1993).

Another prominent phonological cue, syllable number, was investigated by Cassidy and Kelly (1991). It was found that both children and adults were sensitive to the syllable number difference in nouns and verbs, where one syllable words were more likely to be used as verbs than nouns; the percentage of verb usage decreased when the syllable number increased. Cassidy and Kelly (2001) further investigated the sensitivity children demonstrate a syllable number differences between nouns and verbs. There were three conditions: *consistent,* which paired one syllable nonwords with action referents and three syllable nonwords with object referents; *inconsistent,* which had the reverse prediction than in the consistent condition; and *independent*, in which syllable number was unrelated to the word reference type. The results indicated that accuracy generally improved over the three learning blocks and also that the consistent group was significantly better than the inconsistent and independent conditions. Crucially, this difference between the conditions was present in the first learning block, indicating that the children had knowledge of how nouns and verbs differ in terms of syllable number and used this knowledge to guide their decisions. Furthermore, children in the consistent condition learned to respond correctly to 70% of the words in the third block that they initially got wrong in the first block, whereas the inconsistent and independent only corrected 58% and 52% of the erroneous answers in the first block.

These findings provide evidence that children are able to exploit their knowledge of the syllable number difference between nouns and verbs in learning new words. The fact that the children in the *inconsistent* and *independent* conditions learned to correct fewer words that were incorrect in the first block compared to the *consistent* group indicated that they were not able to fully overcome the influence of the pattern within their native language. This result combined with the previous literature shows both sensitivity and manipulation of phonological cues in differentiating grammatical categories. As the authors point out these phonological cues "may increase the speed and accuracy of children's learning the meaning of novel words" by aiding the process of applying grammatical categories (or grammatical function) to novel words (Cassidy & Kelly, 2001). These results indeed provide complementary evidence to the associative learning account by Smith (2000) discussed earlier; the more associated cues are present in the linguistic input the easier learning new words and creating grammatical categories will become.

Bijeljac-Babic, Bertoncini and Mehler (1993) provided further evidence that syllable number can be used to differentiate words. French three day old newborns were exposed to either two or three syllable nonwords. When the infants reached habituation they were presented with a new set of nonwords either identical in syllable number to the habituation nonwords or differing in syllable number. It was found that newborns could discriminate between two and three syllable nonwords, even when the duration of the nonwords was controlled. However, the infants were not able to discriminate two syllable nonwords that contained either four or six phonemic units. Although this does not prove that infants are categorizing words based on syllable number it does support the previous research (Cassidy & Kelly, 1991, 2001; Shi, Werker, & Morgan, 1999) by demonstrating that from an early age we have the perceptual ability to differentiate words based on phonological information.

This tentative conclusion is also supported by Shi, Morgan and Allopenna (1998) who looked at how multiple phonological, distributional and acoustic cues can differentiate lexical and functional (or grammatical) words in Mandarin Chinese and Turkish. Some of these cues were common to both languages and some unique to each language; see Table 1.3 for a summary of the measures used in both the Mandarin Chinese and Turkish sample. Child directed speech to infants (age range 0;11.25 to 1;8.3) was sampled from two mothers (primary care givers) in each language.

Table 1.3 Distributional, phonological and acoustic measures present in the Mandarin Chinese and Turkish sample and whether the measure was useful at differentiating lexical and function words.

Measures	*Mandarin Chinese*	*Turkish*	*Significant at differentiation?*
Type frequency	+	+	+
Rough utterance position	+	+	+
Number of syllables	+	+	+
Presence of complex syllable	+	-	+
Presence of syllable coda	+	+	+
Syllable duration	+	-	+
Vowel duration	-	+	+
Relative amplitude	+	+	+
Pitch change normalized for duration	+	+	-
Syllable reduplication	+	-	+
Presence of marked tone	+	-	+
Vowel harmony	-	+	+

Five percent of the words from the transcript were randomly selected for analysis. These words were divided into lexical words (nouns, verbs, adjectives, and adverbs) and functional words (auxiliary verbs, case markers, complementizers, determiners, prepositions and pro-forms). For both language samples there were significant differences for each cue across lexical and functional words with the exception of acoustic cue of pitch change, which was not significant. Although the individual cues were found to differentiate lexical and functional categories Shi et al. (1998) maintained that each cue alone was not sufficient to aid initial grammatical categorization and that the combination of several perceptual cues could more accurately categorize words into these broad categories. Accordingly, they trained a neural network to classify the lexical and functional words given sets the cues described above for each language. For both Mandarin and Turkish there was a high rate of classification for both lexical and functional items (with classification accuracy at approximately 80%) with all cues included (Shi, Morgan, & Allopenna, 1998).

This research provided cross linguistic evidence that child directed speech contains multiple phonological cues that can aid in the initial categorization of words into superordinate categories. The authors point out that these initial categories may soon be refined by more sophisticated sources of information (lexical, semantic, distributional, and syntactic) that become accessible after the initial categorization (Shi, Morgan, & Allopenna, 1998). However, it may be premature to assume that the types of cues discussed above are not of importance in later categorization, such as differentiating words based on their grammatical usage (i.e. nouns, verbs, etc.). As has been discussed, phonological cues can be used to differentiate nouns from verbs (i.e., Cassidy & Kelly, 1991, 2001; Kelly, 1992, 1996; Monaghan, Chater, & Christiansen, 2005) thus the usefulness of phonological cues in later grammatical categorization should not be discounted.

Onnis and Christiansen (2005; see also Christiansen, Hockema, & Onnis, 2006) conducted an investigated with four languages about whether the word initial and final phonemes of words can predict the grammatical class of the words. The findings were that across all languages word edge cues could reliably classify words into their noun, verb and "other" categories above baseline levels. This indicated that highly detailed phonemic information of beginning and ending phonemes varied based on grammatical class and may be useable to language learners (Onnis & Christiansen, 2005). This result highlights the redundancy that is inherent within language; the associative structure between these redundant cues may provide a powerful learning system for young language learners.

As reviewed above there have been many sources of information from which grammatical categories could be acquired, from innate constructs to distributional analysis to phonological cues, and so on. The literature reviewed provides a convincing argument as to how grammatical categories can be acquired but the acquisition of gendered categories must also be explained. Many languages, such as French, Italian, Spanish, Russian and German have subclasses of masculine, feminine and sometimes neutral nouns. Explaining how seemingly arbitrary categories are leant was without an adequate explanation for many years, but the general trend in the current literature is pointing to language internal information as the key to the acquisition of gender subclasses, mainly phonology.

1.5.1. Gender Subclasses and Phonological Information

Categorisation of words as either feminine or masculine shares many of the same characteristics of grammatical classification, with the added difficulty in that both masculine and feminine words are usually used in the same syntactic and semantic context. This added difficulty helped make gender subclasses somewhat mysterious as it was thought that gender subclasses were completely arbitrary. However, research in recent years has shown that gender subclasses are far from arbitrary. Gender subclasses, like grammatical categories, have numerous interrelated phonological cues. In every studied language phonological cues correlate with gender subclasses, such as in Spanish and Italian feminine nouns often ending in "-a" and masculine nouns in "-o" (Brooks, Braine, Catalano, Brody, & Sudhalter, 1993). Generally, these cues are only found in a subset of all nouns in the subclass creating a situation where not all words have gender cues yet language users are still able to categorize all nouns into their respective gender. The process of learning category subclasses based on phonological information embedded within only a subset of the subclass members has been the focus of many investigations.

Brooks et al. (1993) used an artificial language to investigate this phenomenon with both undergraduate participants and children aged nine to ten years old. They used 90 event cards in their experiment which contained 30 objects and an actor subject (which was usually Frippy, the monkey). The relation between the actor and the object was the key factor; either the actor was walking towards, away or was at (usually handling) the object. Their language consisted of 30 nouns, broken down into two subclasses. There were consistent phonological markers to signal subclass status in 60% of the nouns in each subclass (the word ending "oik" for subclass one and "oo" for subclass two). For each subclass there was a set of suffixes that indicated the relation the Frippy was to the object; this effectively was a miniature case marking system. For subclass 1 "eef," "rog," and "ast" meant walking towards, walking away and "at" the object. For subclass 2, the suffixes "foo," "ilg," and "tev" indicated the same meanings. A typical phrase in this language would be the actor name followed by the noun and the appropriate suffix ("Frippy choik-eef") (Brooks, Braine, Catalano, Brody, & Sudhalter, 1993, p. 79). In the control condition, the 18 words with the phonological markers were randomly assigned to either of the two subclasses, so that the phonological form of the words no longer aided suffix learning. Only one third of the possible sentences were used during the training phase; the remainder were withheld for the test phase. The participants were shown event cards during test and had to state the corresponding phrase.

Results showed that the experimental group, who received phonological cues to the subclasses performed significantly better than the control group on all tests. The experimental group performed near ceiling for all test items which occurred during the training session, while the control group performed far lower. For the test items that did not occur during the training phase, the experimental group named the correct phrase significantly more then the control group, although the difference was somewhat less on the unmarked nouns (Brooks, Braine, Catalano, Brody, & Sudhalter, 1993). On the unmarked nouns, the experimental group were poorer than for the marked nouns while the control group remained the same. Most importantly, the experimental group was able to generalize their knowledge of the subclasses to both new marked and unmarked nouns. This indicated that categorisation and generalisation of the subclasses occurred when only a proportion of all nouns containing phonological cues. This result is supported by earlier word by Brain et al. (1990) in which arbitrary subclasses with only case marking were not learned and also the Gerken et al. (2005) experiment discussed earlier (although see Taraban, 2004 for a discussion on the role of attention in learning from case markings).

Rote learning was the only type of learning evident when only the case marking indicated subclass status. However, in these experiments the phonological information was quite prominent and explicit whereas the explicitness of gender subclass phonological cues varies across languages. For example, German gender appears to be hard to learn and is not fully acquired until 8-9 years of age. However, it has been found that German contains numerous interrelated cues on many different levels (phonological, semantic, and morphological) but that each individual predictive power is quite weak. However, when these cues are combined gender membership can be accurately predicted (Köpcke, 1982, as cited in Frigo & McDonald, 1998). The general complexity and subtlety of these cues contributes to the difficulty in mastering gender classification whereas French and Italian contain more reliable and explicit cues to the gender subclasses, thus children display a mastery of gender classification at a much younger age.

Frigo and McDonald (1998) extended the work of Brooks et al. (1993) by systematically looking at several factors, including marker frequency, marker salience, marker position, and marker redundancy. The basis of this experiment was that participants had to learn how to say "Good day" and "Good evening" in a new language, but the type of greeting depended on the social class of the person they were to greet. "Good day" could be expressed with the word "jai" or "fow" and "Good evening" corresponded to either "quo" or "mih." This again is roughly equivalent to case marking in natural languages. Twenty nonsense words were created to represent the names; ten in each subcategory. Six of these names in each subclass ended with a phonological marker; the frequency of the marker was manipulated with one marker per subclass for the high frequency condition and two markers for the low frequency condition. The markers also varied on saliency: highly salient markers were a syllable long and low saliency markers were a phoneme long. See Table 1.4 for examples of these manipulations.

Table 1.4 Example of the phonological marker nonwords used in the Frigo and McDonald (1998) paper with marker frequency and saliency manipulations examples.

	High frequency markers		*Low frequency markers*	
	Subclass 1	*Subclass 2*	*Subclass 1*	*Subclass 2*
High salience	ersum*glot*	halam*rish*	ersum*thope*	halam*rish*
	lauf*glot*	sam*rish*	lauf*glot*	sam*wek*
Low Salience	ersum*t*	halam*sh*	ersum*p*	halam*sh*
	lauf*t*	sam*sh*	lauf*t*	sam*k*

Four of the six names with phonological cues were paired with both the greetings during training; the remaining two were only paired with one greeting. Likewise only two of the four unmarked names in each subclass were paired with both greetings; the remaining two were only paired with one greeting. The remaining names/greeting combinations not used during training were used during the test phase. An unsystematically marked condition was created by exchanging half of the marked names from subclass one with the half of the marked names from subclass two.

The results indicated that the highly salient and high marker frequency language led to an ability to generalize to new marked names (Frigo & McDonald, 1998).

However, all participants were unable to correctly supply the correct subclass greeting when provided with the other greeting on unmarked names. Therefore, the participants were able to use the phonological information to categorise the words according to the subclasses but were not able to link this knowledge with the subclass indicator (or case marking). The same results were found when the phonological markers occurred at the beginning of words. However, learning from both the case markers and the phonological cues was found when the language included highly salient markers at both the beginning and end of the words. This language allowed the participants to supply the correct greeting for new phonologically marked names and unmarked names (given one greeting is provided).

More generally, this study indicated we have the ability to very quickly (within a short experiment) categorise nonsense words and link them to category indicators, provided the phonological cues to category membership are redundant, salient and very frequent. However, these phonological cues do not need to be present in all category members; a subsection of the category members are sufficient to categorise all category members. This provides further evidence that phonological cues may be a critical component in the acquisition of gender subclasses. The combination of phonological and case marking cues in these experiments mimics the cues available for many gendered subclasses in natural languages. However, there may also be other semantic, syntactic, and morphological cues to aid in the acquisition of gender subclasses.

The above experiments outline that phonological information within words is both useful and usable in language learning. This research, along with many other studies, clearly demonstrates that phonological information is important in the language acquisition process in general, and specifically in the acquisition of grammatical categories. However, the importance of phonological information may be enhanced when considered in parallel with distributional information.

1.6. The combined effects of distributional and phonological information

Two possible sources of information to aid grammatical categorization during language acquisition have been outlined. Distributional information has long been known to be an important and reliable cue to grammatical categories. Phonological information has also been shown to be an important cue in grammatical categorization as well. The majority of the reviewed studies have isolated either distributional or phonological cues in order to determine whether these cues are both *useful* and *usable* in language learning. Isolating these cues is a very important step in determining what information is relevant and what is irrelevant in regards to language learning. However, once isolated cues have been found to be useful in language learning, there must also be research that investigates how combined cues influence the language acquisition process. This important step brings us closer to understanding the learning process as it happens in early childhood. Children do not attend to one type of information; they take advantage of all types of information within their environment. It would not be advantageous to ignore information which aids learning. This hypothesis has been investigated across a variety of language learning problems (Christiansen & Monaghan, 2006; Curtin, Mintz, & Christiansen, 2005; Gerken, Jusczyk, & Mandel, 1994; Johnson & Jusczyk, 2001; Jusczyk, 1999; Mattys & Jusczyk, 2001; Monaghan, Chater, & Christiansen, 2005; Monaghan, Onnis, Merkx, & Chater, Submitted; Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara, 2006; Onnis, Monaghan, Richmond, & Chater, 2005; Reali, Christiansen, & Monaghan, 2003; Theissen, Hill, & Saffran, 2005; Theissen & Saffran, 2003).

Monaghan et al. (2005) extended their individual analysis of how phonological and distributional cues can separately induce categorization by conducting discriminant analyses which included both types of cues. For open and closed class words with all 20 distributional cues (described above) and all 16 phonological cues (also described above) 99.0% of the closed class words were correctly classified while 52.7% of the open classed words were correctly categorized. Similarly, 67.0% of the nouns and 71.4% of the verbs were correctly categorized, giving overall categorization of 69.2%. Furthermore, when broken down by frequency 79.7% of 1000 highest frequency words and 67.4% of the lowest frequency words were correctly categorized. It should be recalled that when only distributional cues were considered the lower frequency words had more categorization errors whereas the higher frequency words had more errors with only phonological cues. One conclusion from these results was that for all frequencies to be categorized correctly (at an acceptable rate) both distributional and phonological cues must be available to the language learner.

These analyses have shown it is useful to have both types of cues available but it remained to be tested whether learning can actually occur from the combination of these cues. Monaghan et al. (2005) next tested this using an artificial language based on the Valian and Coulson (1988) bigram distributional language with two phrases *aA* and *bB*. There were two categories of words (*A* or *B)* and six words per category. The *a* and *b* words served as marker elements, determining which category of word would follow. In addition, half of all category words were of high frequency and occurred twice as frequently as the low frequency words. The training sentences consisted of both an *A* phrase and a *B* phrase (*aAbB* or *bBaA*). In addition to the distributional cues, there were consistent phonological cues within the category words for the phonologically coherent condition. Category 1 words had onset and offset consonant clusters, rounded low vowels and contained nasals and stops. Category 2 words had no consonant clusters, unrounded high vowels and did not contain nasals or stops, but only fricatives. For the phonologically incoherent condition the phonological cues were removed from the categories by exchanging the low frequency Category 2 words with the high frequency Category 1 words.

Half of the test sentences were compatible with the above language and half contained some sort of error. There were two types of incorrect test sentences; one type of incompatible test sentence contained the incorrect marker word in one of the phrases (*aAaB* or *bBbA*) and other incompatible test sentence contained the incorrect marker word for both of the phrases (*aBbA* or *bAaB*). The participants in the coherent condition were more able to discriminate between the compatible and incompatible sentences than participants in the incoherent condition. Participants also scored higher for high frequency sentences than for low frequency sentences, which was predicted as the higher frequency sentences inherently were paired with the distributional cues more often. There was also a frequency by phonological coherence interaction; there was no difference between the frequency conditions when there were both phonological and distributional cues, but the participants with only distributional cues were impaired on the low frequency test. This was expected from the corpus analyses; distributional information was more informative for high frequency words which lead to no differences between the coherency groups, whereas phonological information was more informative for low frequency words, which would allow the phonological cues in the coherent group to compensate for the lower frequency of the distributional cues.

These corpora analyses and experimental findings are preliminary evidence that the integration of phonological and distributional information is important and necessary for the process of acquiring grammatical categories. However, even the idea that these two sources of information are the only cues children use for this process is naïve, but it is necessary to show how these two cues work together before integrating other aspects of language, such as prosody or semantics. The first necessary step is to replicate the above study in the auditory modality; children do not learn from reading, they learn from hearing.

1.7. Chapter summary

The previous review has demonstrated that distributional and phonological information differentiates grammatical categories when investigated using corpus techniques. Critically, participants (adults and infants) are able to use these phonological and distributional cues to differentiate grammatical categories in highly controlled, experimental settings. The convergence of this evidence suggests that both types of cues are important in the early formation of grammatical categories during language acquisition. However, there has been relatively little work investigating how these two types of grammatical category cues may combined to provide a more complete picture of the acquisition process. This thesis will extend the work of Monaghan et al. (2005) by investigating learning from these combined cues within the auditory modality. Chapters 2 and 3 will investigate whether grammatical categorization will proceed with only distributional bigram cues, or whether additional phonological cues are necessary for categorical learning to occur.

A secondary aim of this thesis is to investigate the role of associative learning in the grammatical categorization process and in language learning in general. The feature theory and prototype theory of associative categorization will more particularly be investigated. In particular, the application of the prototype theory to language will be tested in Chapters 4 and 5 by investigating whether marker word phonological similarity to naturally occurring distributional information influences learning ability. Throughout the thesis the theoretical framework of the R-W model will guide any possible associative learning explanations of the data.

Chapters 6 and 7 investigated the difference between trigram and bigram distributional cues and attempted to differentiate the aXb frame explanation from the aX and Xb combined bigram explanation. Experiment 6 will replicate the aXb corpus analysis by Mintz (2003) with an aX and Xb comparison analyses using the same procedure and corpora. Chapter 7 directly tested whether learning from trigram distributional information occurs from the a_b structure of the combination of aX and Xb bigrams with two experiments; one with the non-adjacent dependency intact and the other with only bigram distributional cues.

2. Phonological and Bigram Distributional Information

The written artificial language experiment combining phonological and bigram distributional information in Monaghan et al. (2005) was extended to the auditory domain in St. Clair and Monaghan (2005). The auditory experiment was necessary to approximate more closely the acquisition process, as typically language learning occurs through auditory input. The structure of the language was not changed and the auditory stimuli were synthesized using Festival Speech Synthesizer (Black, Clark, Richmond, King, & Zen, 2004). Surprisingly, the results indicated that there was no difference between the phonologically coherent and incoherent conditions or high and low frequency sentences. However, on a category word sorting task which required the participants to sort the words into two equal piles, the coherent group was able to explicitly sort more category words together than the incoherent group, although this effect was only marginally significant. The incoherent group sorted significantly less cards together than expected by chance, which was suggestive that the mismatch between the phonological properties of the words and the categories interfered with their ability to correctly sort the cards based on the distributional cues to category membership.

The discrepancy between the auditory and written experiments was unexpected. St. Clair and Monaghan (2005) highlighted two possible explanations for the discrepancy. The first was a slight difference in exposure during the training; in the written version the training sentences were presented for ten seconds during which time the participants read the sentence aloud. This allowed time to repeatedly read and study the training items. In the auditory version, the participants heard two instances of each training sentence, separated by a one second interval, leaving approximately five seconds for the participant to repeat aloud the training sentence. Although the total time per sentence was equated additional cognitive resources were needed in the auditory version for the rehearsal of the training sentences which left fewer resources available for the extraction of categorical knowledge. A further problem occurred in matching the written words in the card sorting task to the auditory stimuli. This was more difficult in the auditory version as the participants had to map the orthography to their memory of the auditory words while trying to perform the task. However, even with these potential problems there was at least limited evidence that categorical learning was taking place.

A further experiment investigated whether the position of the distributional cue in relation to the category words would influence the role of phonological cues and overall categorization. All stimuli were changed from *aAbB* to *AaBb* sentences (St. Clair & Monaghan, 2005). When the category words preceded the marker words the participants were significantly more accurate at differentiating the grammatical and ungrammatical test sentences, but only when they had received a language that contained both phonological and distributional cues. Additionally, higher accuracy was found with high frequency test sentences, which indicated that more pairings with the succeeding distributional cues allowed more categorical knowledge. With regards to the card sorting task the phonologically coherent group did sort more cards than the incoherent group, although they were not significantly above chance level. The phonologically incoherent group was significantly below chance level, as in the previous study.

This finding was intriguing as it suggested the possibility that succeeding distributional cues facilitate the use of phonological information in grammatical categorization more than preceding distributional cues. This result was unexpected as it has been found in corpus analyses that preceding distributional cues are more informative than succeeding distributional cues in terms of grouping words of the same grammatical category (Redington, Chater, & Finch, 1998). Additionally, Frigo et al. (1998) found that preceding and succeeding cues (which could be thought of as morphological distributional cues) were insufficient to induce categorization. Only when preceding and succeeding cues were combined was evidence of categorization found in that study. Nonetheless, St. Clair and Monaghan (2005) provide evidence that succeeding (but not preceding) distributional cues allow categorization of word classes but only when there are numerous interrelated phonological cues to aid categorization as well.

However, the overall level of category learning was not as high as was found in the Monaghan et al. (2005) study. The differences between the training sessions were suggestive in that there may not have been sufficient exposure to the training sentences in the auditory modality, especially in the nonsignificant preceding cue experiment. The following Experiments replicate the above studies but with twice the amount of training. It was hypothesized that succeeding distributional cues will still remain a more powerful categorization cue but that both types of cues will facilitate grammatical categorization.

2.1. *Experiments 1 and 2*

Experiments 1 and 2, although run as separate Experiments, are presented together. Each Experiment will be analysed separately, and then combined to determine whether the difference in categorical learning from preceding and succeeding cues was significant. This combined analysis was justified as the same population of participants was used for both Experiments and they were run within a reasonable time frame (within a few months).

2.1.1. Participants

In Experiment 1, twenty-four[2] University of York undergraduates participated in the study. Participants had not taken part in any similar experiments. The average age was 19.88 years; range was 18 to 39. There were 18 female and six male participants in the sample.

In Experiment 2, twenty-six University of York undergraduate and postgraduate students participated in the study. Two participants were excluded as they produced unusable card sorting results (unequal groups)[3]. Participants had not taken part in any similar experiments. The average age of the included sample was 19.96 years; range was 18 to 24. There were 22 female and 2 male participants in the included sample. All participants were native English speakers and were paid £2 or given course credit.

2.1.2. Materials

The artificial language used was an adapted version of Valian and Coulson's (1988) written language. There were 12 category words in the language divided between two categories, A and B. There were two marker words in the language with one marker word always co-occurring with category A words and the other marker word co-occurring with category B

---

[2] The N number for Experiments 1 to 7 was determined based on the number of subjects in the previous experiments reported in St. Clair and Monaghan (2005) and Monaghan, Chater and Christiansen (2005).

[3] All statistical analysis relating to the similarity task was conducted with and without these excluded participants. There was only one difference, in which one statistically significant result became marginally significant. This result will be noted in the results section. The general reason for excluding participants who did not complete the card sorting task correctly was the possibility they disregarded the instructions throughout the entire Experiment. It was strongly emphasized by the experimenter before experimental session that *equal* groups were needed; this was emphasized again at the beginning of experimental session in the written instructions and then again while the participant completed the card sorting task. Failure to comply with this strongly emphasized instruction implies the possibility that these participants did not comply with any of the experimental instructions.

words. In Experiment 1, the marker word preceded the category words (*aAbB*) whereas the marker word succeeded the category words in Experiment 2 (*AaBb*). The language from Experiment 1 will be described in detail; the corresponding language in Experiment 2 only differed in the placement of the marker word.

The two marker words, *alt* and *ong,* were counterbalanced across participants to create two dialects. In dialect 1 *alt* preceded Category A words and in dialect 2 *alt* preceded Category B words. Half of the category words were of high frequency and occurred twice as often (8 times in each training cycle) as the low frequency category words (4 times per training cycle).

There were two conditions to test the role of phonological cues in categorisation. In the phonologically coherent condition (or coherent hereafter), words within the same category had shared phonological properties. Category A words had consonant clusters at the onset and offset, unrounded high vowels, and contained nasals and stops. Category B words had no consonant clusters, contained rounded low vowels, and fricatives. These phonological properties were found to differentiate lexical categories in the Monaghan et al. (2005) corpus analysis. For the phonologically incoherent condition (or incoherent hereafter) the low frequency words from the Category B were exchanged with the high frequency words from Category A, creating categories with no common phonological cues. Table 2.1 lists the words for the coherent condition. During training, 18 sentences were presented in the form *aAbB*, where *a* and *b* were the marker words and *A* and *B* were category words. *A* phrases (*aA*) and *B* phrases (*bB*) appeared equally often in first and second position. See Appendix A for full details of all training sentences.

Table 2.1 Category words by frequency status in the Phonologically Coherent condition

Frequency	*Category A*	*Category B*
High	Tweand	Foth
	Dreng	Vawse
	Klimp	Suwch
Low	Gwemb	Zodge
	Prienk	Thorsh
	Blint	Shufe

There were two test sessions and the order in which the participants received the test sessions was counterbalanced throughout the experiment. Each test session consisted of 24 sentences. Twelve of the test sentences had not occurred during training, but conformed to the artificial language's regularities (compatible). Half of these sentences were composed of high frequency category words and the remaining were composed of low frequency category words. The remaining 12 sentences violated the artificial language's regularities; six had the same marker word preceding both category words (e.g., *aAaB*, syntax-incompatible), and six had both marker words preceding the wrong category word (e.g., *aBbA*, incompatible). The *aAaB* test sentences are termed syntax-incompatible as they violate the general pattern of the language's syntax, of having both marker words in each sentence. For these sentences the violation occurred in half the cases for high frequency and half for low frequency category words. Table 2.2 summarises the three sentence types. See Appendix B for details of all test sentences for both test sessions.

Table 2.2 Examples of the test sentence types

Compatible	*Syntax-incompatible*	*Incompatible*
aAbB	*aAaB*	*bAaB*
bBaA	*bBbA*	*aBbA*

All stimuli were synthesized within the Festival Speech Synthesizer program (Black, Clark, Richmond, King, & Zen, 2004). The kal diphones voice at 22050 Hz. was used for all speech synthesized throughout this thesis. There was a standard pause inserted between each word in the sentences; this pause varied depending on the surrounding phonemes and was automatically inserted by the Festival program. The average pause was 216 ms. taken from the pauses within four random sentences.

There were also 12 cards with the category words printed on them which were used in a card sorting task. The task required the participants to sort the 12 category word cards into two equal piles; due to the nature of this task the lowest possible score is three Category A words and three Category B words. If two Category A words are sorted together, then the remaining four words are Category B words. Chance level for this task is calculated using the proportion of choosing one Category A word, and then choosing a second Category A word, and a third, fourth, fifth and sixth expected by chance. Chance level was calculated to be 3.91 cards, which is the chance level used for all experiments that include 12 category words.

2.1.3. Procedure

Participants were instructed to pay attention to the patterns within the made-up language. They first heard all the nonwords in the language individually. After this familiarization phase, the first training session began during which the participants listened to the 18 training sentences presented in a random order. Each sentence was presented four times making a total of 72 training sentence trials. A trial consisted of two presentations of the training sentence with approximately a one second interval after the first presentation. Participants were instructed to repeat the sentence aloud after both presentations; they were given approximately five seconds for this repetition to occur.

During the test phase, participants were instructed that half the sentences were similar to the training language and half dissimilar. They then had to judge whether each of the 24 test sentences was similar or dissimilar to the training language. The participants pressed "Y" if they thought the sentence was similar and "N" if they thought the sentence was dissimilar. Participants were then given four more cycles of the 18 training sentences, and then testing was repeated using a different set of 24 test sentences. After the second test session, the participants were given instructions to sort the 12 category cards into two groups according to which words they thought went together.

2.1.4. Results and Discussion

All *t* tests and pairwise comparisons were Bonferonni corrected. As this correction is very conservative, the tests that were significant before correction have been marked to give a more complete picture of what the data is showing. Therefore, $p = .06^{\dagger}$ indicates that the uncorrected *p* value was significant at the .05 level, $p = .06^{\dagger\dagger}$ indicates significance at the .01 level, and $p = .06^{\dagger\dagger\dagger}$ indicates significance at the .001 level. When the Levene's test for homogeneity of variance assumption was not met the test statistic for unequal variance is reported.

2.1.4.1. aAbB Experiment 1 Results

Overall accuracy in the similarity task was 72%. A mixed ANOVA was conducted with time (1st and 2nd test session) and frequency (high and low) as the within subjects factors and phonological coherency (coherent and incoherent) as the between subjects factor. In addition, two dummy variables (dialect and test counterbalance) were submitted into the

ANOVA (Pollatsek & Well, 1995). The dependent variable was the amount of correctly accepted compatible test sentences (as similar to the artificial language) and correctly rejected syntax-incompatible and incompatible sentences (as dissimilar to the artificial language).

Time was found to be highly significant, $F(1,16) = 8.78, p < .01, \eta_p^2 = .35$, with participants more accurately differentiating the test sentence types in the second test session, M = 9.02 for second test session; M = 8.25 for the first test session; both out of 12. There was no significant difference between the coherent and incoherent groups, $F(1,16) = 2.25, p = .15, \eta_p^2 = .12$; M = 8.92 and M = 8.35 out of 12, respectively, nor was there a difference between high and low frequency sentences or any interaction between frequency and phonological coherence, $F < 1$; M = 8.73 and M = 8.54 out of 12, respectively. There was no other significant effects or interactions; time by coherency: $F(1,16) = 1.44, p = .25$; all others $F < 1$.

Although there was no difference between the two phonological coherency groups learning was above chance level with both groups combined, $t(23) = 13.05, p < .001$; M = 34.54 with a chance level of 24. However, this may not be due to any sort of category learning but from knowledge of the syntax. The participants very easily learned that both marker words occurred in each sentence. Therefore, correct rejection of syntax-incompatible sentences (*aAaB*) was near ceiling level for all participants, as can be seen in Figure 2.1. When looking at the incompatible sentences a trend appeared in that the participants who received the coherent language had a higher ability to correctly reject these sentences than the participants who did not have any phonological cues to category membership, but this difference was not significant, $t(22) = 1.57, p = .13$; M = 3.83 for coherent and M = 2.17 for incoherent; of a possible 12, respectively.

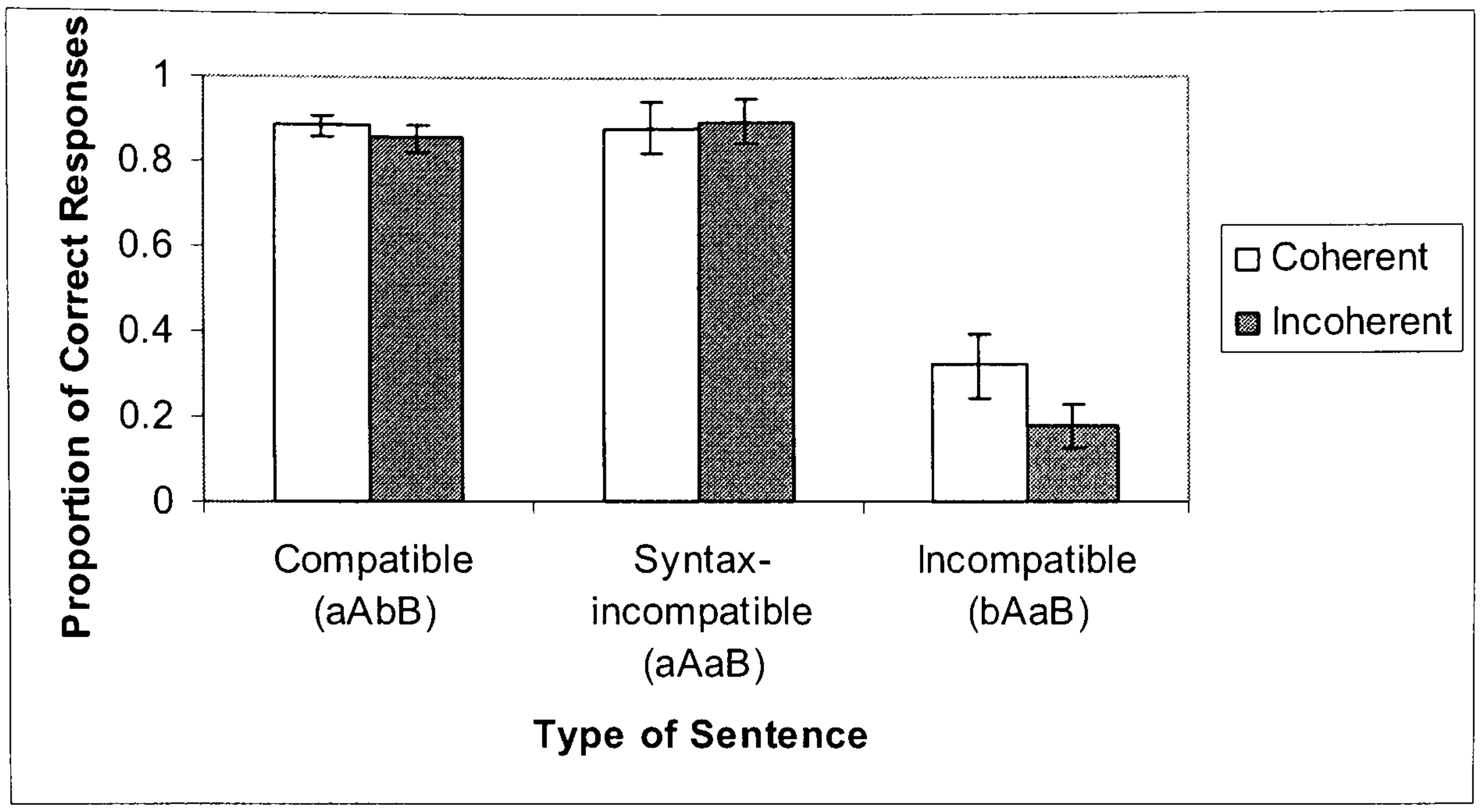

Figure 2.1 Proportion of correctly accepted (compatible) and rejected (syntax-incompatible and incompatible) test items in the phonologically coherent and incoherent conditions in Experiment 1.

As the correct rejection of syntax-incompatible sentences showed only syntactic knowledge, not categorical knowledge, a second ANOVA was conducted with only compatible and incompatible sentences considered. Differentiation of compatible and incompatible sentence would indicate that categorical knowledge had been learnt. This ANOVA contained the same variables as the first ANOVA. There was a marginally significant effect of time, $F(1,16) = 3.89, p = .07, \eta_p^2 = .20$; M = 5.73 first test session and M = 6.21 second test session; both out of a possible nine. The main effect of phonological coherency was significant, with higher accuracy with the coherent than the incoherent group, $F(1,16) = 5.07, p < .05, \eta_p^2 = .24$; M = 6.27 and M = 5.67 out of 9, respectively. As in the first analysis there were no differences between high and low frequency sentences nor an interaction with phonological coherence, $F < 1$; M = 6.04 and M = 5.90 out of nine, respectively. There were no other significant main effects or interactions; time by coherence: $F(1,16) = 2.65, p = .12$; all others $F < 1$.

Although there was a significant difference between the groups, neither the coherent nor incoherent condition was significantly different from a predefined learning level, $t(11) = 1.25, p = .47$, M = 25.08 for coherent; $t(11) = -2.24, p = .09^{\dagger}$, M = 22.67 for incoherent conditions. The predefined learning level in this case was 24, as if participants used only the strategy of looking for the repeated marker word to exclude a test sentence as dissimilar the level of correct responses we would expect is 24. This is calculated as there were 24 compatible test items across both test sessions, which, following the above strategy, would have been correctly accepted but the 12 incompatible test items would have been incorrectly accepted.

The card sorting task was analysed by looking at both the *distributionally* and *phonologically* defined categories. The distributionally defined categories were the category words demarcated by the marker words; these differed between the coherent and incoherent conditions. The phonologically defined categories were defined by the common phonological cues and did not differ between the coherent and incoherent conditions.

When analysing the distributionally defined categories, the coherent group sorted significantly more cards correctly together than the incoherent group, $t(22) = 2.36, p < .05$; M = 4.50 and M = 3.58 out of six, respectively. However, the incoherent group did not differ from the chance level of 3.91, $t(11) = -1.43, p = .35$, and the coherent group was also not significantly different from chance level, $t(11) = 1.88, p = .17$. When looking at the phonological categories there was no difference between the two conditions, $t(22) = .84, p = .41$, M = 4.50 for coherent and M = 4.17 for incoherent conditions.

The overall scores from the compatible and incompatible sentences were correlated with the distributionally defined card sorting results, $r = .49, p < .05$, indicating a positive relationship between the two dependent measures. For the coherent condition the correlation remained moderately high, $r = .49, p = .11$, but the significance of the correlation was lost. The correlation in the incoherent condition was not near significant, $r = .13, p = .69$. See Figure 2.2 for the scatterplot of the card sorting results and the results of the compatible and incompatible sentences.

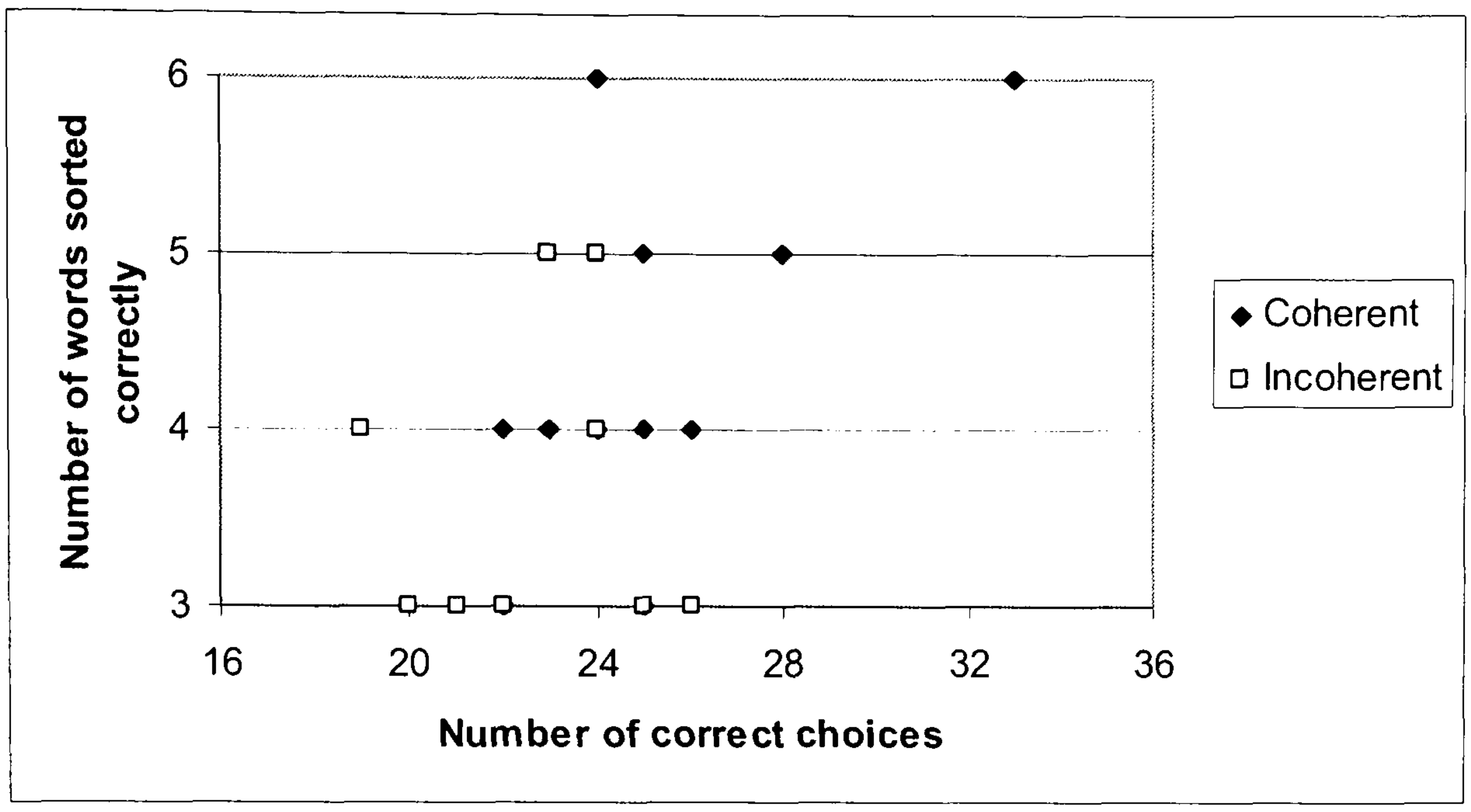

Figure 2.2 Scatterplot of the number of correctly sorted words (distributionally defined categories) and number of correct responses to compatible and incompatible test items in Experiment 1.

2.1.4.2. aAbB Discussion

The significant main effect of time when all of test sentence items are considered indicated that the participants were learning across the course of the experiment, as they were more accurate during the second test session than the first. However, this learning appears to be mainly due to an increase in syntactic knowledge, as this main effect is only marginally significant when the syntax-incompatible test items were removed from the analysis. The significant difference between the phonological coherency conditions when only compatible and incompatible test items were considered was promising in that it indicated that there was more learning when phonological and distributional cues combine to predict category membership. However, neither condition was above chance levels of responding so no firm conclusions can be made based on this analysis.

The card sorting results for the coherent condition indicated that the participants were able to sort the category words together better than would be expected by chance. However, the comparison between the two conditions with the distributionally defined categories was confounded by the incoherent condition. When faced with the 12 nonwords these participants were more likely to sort them according to their phonological properties than based on the associations that should have formed between the marker element and the category words. This will be discussed further in the general discussion. The results of the correlation analysis indicated that there was only a positive relationship between the ability to correctly differentiate the test sentences and to explicitly sort to the two categories of words in the card sorting task in the coherent condition, which was not surprising given the performance of the incoherent condition on the card sorting task.

This Experiment replicated St. Clair and Monaghan (2005) in finding no concrete evidence of categorical learning with only distributional cues. There was suggestive evidence that some categorical knowledge was learnt when both distributional and phonological cues were available; however, also replicating St. Clair and Monaghan (2005) no firm conclusions can be made. The most suggestive evidence came from the incompatible test sentences; the coherent participants were more able to correctly reject these sentences than the incoherent participants, although this comparison was not significant. One possible hypothesis for the lack of learning was that the syntax-incompatible test sentences drew too much attention during the test phases, and thus confounded any attempt to differentiate the compatible and incompatible test sentences. This issue will be discussed further in the general discussion.

Providing twice the amount of training did not increase participants' ability to learn from preceding cues, even when combined with phonological information. Experiment 2 tested whether increased training allowed further categorical knowledge to be learned from a succeeding distributional cue language than what was found in St. Clair and Monaghan (2005).

2.1.4.3. *AaBb Experiment 2 Results*

Overall accuracy for the similarity task was 74%. As in Experiment 1, a mixed ANOVA was conducted with time (1st and 2nd test session) and frequency (high and low) as the within subjects factors and phonological coherence (coherent and incoherent) as the between subjects factor. As in Experiment 1, two dummy variables (dialect and test counterbalance) were added to the ANOVA and the dependent variable was the number of correctly accepted or rejected test sentences.

The only significant difference was between the coherent and the incoherent group, with the participants who received both phonological and distributional cues performing more accurately at discriminating the test sentences, $F(1,16) = 15.53, p = .001, \eta_p^2 = .493$; M = 9.63 and M = 8.17 out of 12, respectively. There were no other significant main effects or interactions, frequency: $F(1,16) = 1.26, p = .278$; frequency by coherence: $F(1,16) = 1.97, p = .18$; test by frequency $F(1,16) = 2.63, p = .12$; all others $F < 1$.

Learning was again well above chance level, $t(23) = 12.86, p < .001$; M = 35.58 with a chance level of 24, however as mentioned in Experiment 1 these results are skewed by the fact that the syntax-incompatible sentences could be rejected on syntactic knowledge alone. As can be seen in Figure 2.3 correct rejection of syntax-incompatible sentence was at ceiling level in both the coherent and incoherent conditions, but the correct acceptance of compatible sentences was not ceiling level, although it was approaching ceiling level in both the coherent and incoherent conditions. The critical incompatible sentences show overall higher levels of correct rejections than in Experiment 1. Additionally, there was significantly more correct rejections of incompatible sentences in the coherent group than in the incoherent group, $t(22) = 2.11, p < .05$[4]; M = 5.67 and M = 3.75 (out of a possible 12), respectively.

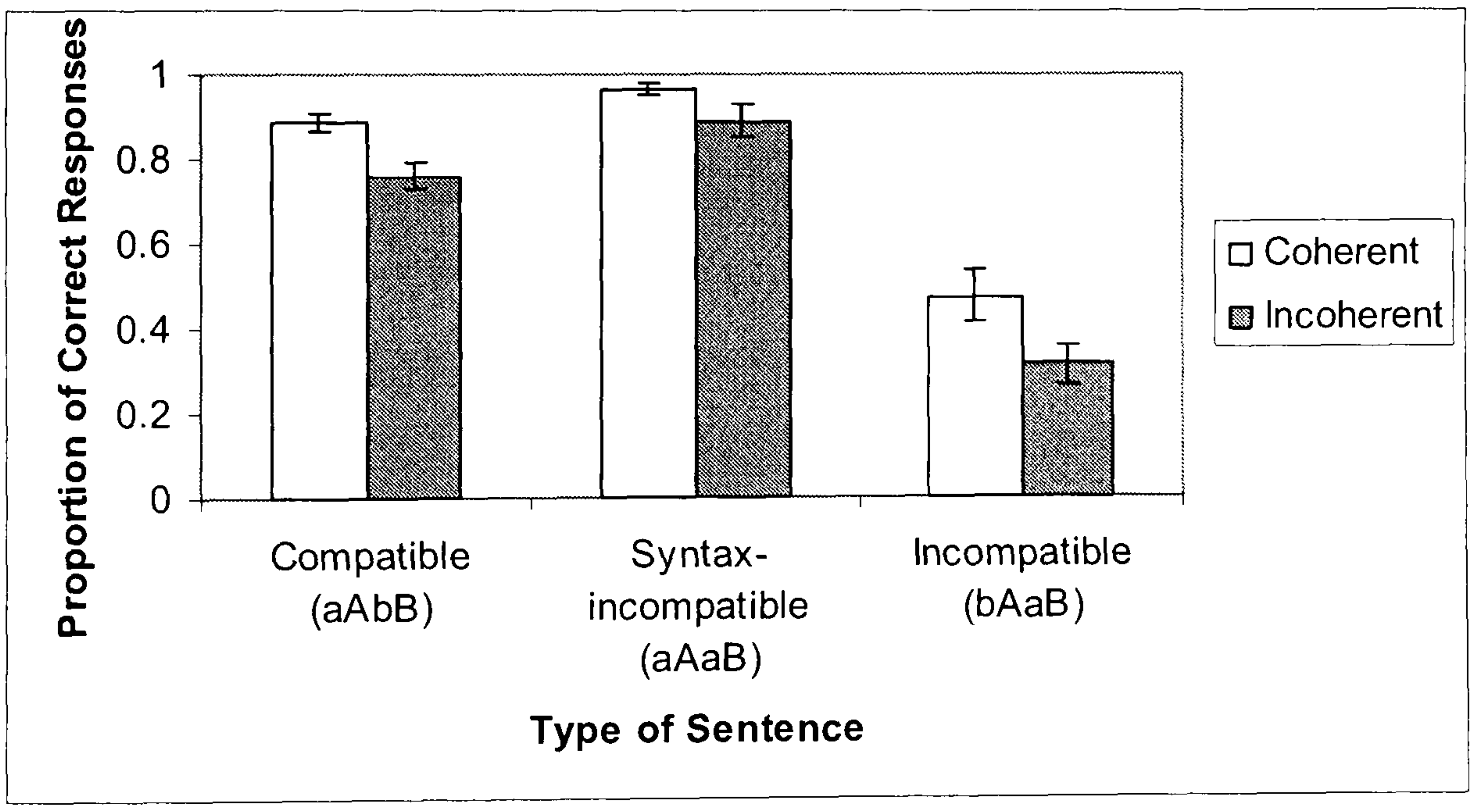

Figure 2.3 Proportion of correctly accepted (compatible) and rejected (syntax-incompatible and compatible) test items in the coherent and incoherent conditions in Experiment 2.

The results were investigated further with a second mixed ANOVA with the same variables but only compatible and incompatible sentence data. As in the first ANOVA the only significant difference was between the coherent and the incoherent groups, with the coherent group more accurately discriminating the compatible and incompatible test sentences, $F(1,16) = 14.69, p = .001, \eta_p^2 = .48$; M = 6.73 and M = 5.50 out of 9,

---

[4] With the excluded participants included, this result falls just short of significance: $t(22) = 1.93, p = .07$; M = 5.58 for coherent and M = 3.83 for incoherent.

respectively. There were no other significant main effects or interactions, frequency by coherency: $F(1,16) = 1.83, p = .20$; time by frequency: $F(1,16) = 3.09, p = .10$; all others $F < 1$.

Contrary to Experiment 1 when looking at only compatible and incompatible data the coherent condition was significantly above chance level of 24, $t(11) = 3.37, p < .05^{\dagger\dagger}$, M = 26.92 of a possible 36, as well as significantly different from the incoherent group, as the main effect reported above indicated. Furthermore, the incoherent group was marginally significantly below chance level, $t(11) = -2.48, p = .06^{\dagger}$, M = 22.0 of a possible 36.

For the card sorting task, the coherent participants sorted significantly more category words together than the incoherent participants, $t(16.872) = 2.66, p < .05$, when looking at the distributionally defined categories. The coherent group was not significantly different from chance level of 3.91, $t(11) = 1.64, p = .26$, M = 4.5 of 6, whereas the incoherent participants were marginally significantly below chance level, $t(11) = -2.56, p = .054^{\dagger}$, M = 3.42 of 6. When comparing the two coherency conditions based on the phonologically grouping there was no difference, $t(16.62) = .41, p = .69$.

The results of the compatible and incompatible test items were correlated with the distributionally defined card sorting results and a significant positive relationship was found, $r = .75, p < .001$. When broken down into the two conditions a strong correlation between the two measures was found in the coherent group, $r = .84, p = .001$, but not in the incoherent group, $r = .34, p = .279$. See Figure 2.4 for a scatterplot of the two dependent measures.

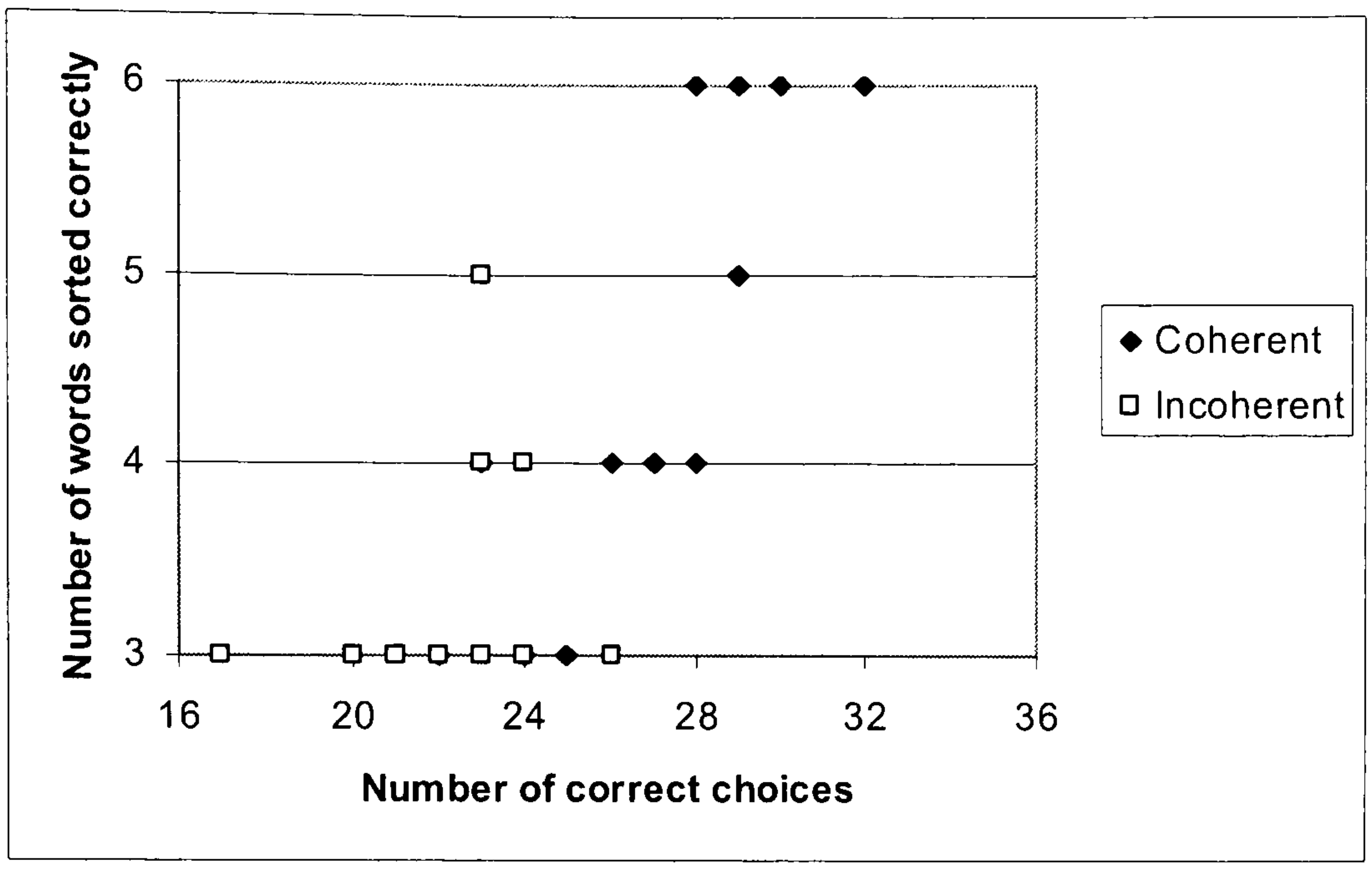

Figure 2.4 Scatterplot of the number of correctly sorted words (distributionally defined categories) and number of correct responses to compatible and incompatible test items in Experiment 2.

2.1.4.4. AaBb Discussion

The main result of this experiment, which is lacking in Experiment 1, was the increased categorical knowledge when both phonological and distributional cues were available to aid the categorization process. The significant difference between the coherency groups was found regardless of whether the syntax-incompatible test sentences were included in the analysis. Similarly, learning for the coherent group was significantly above chance with and without the syntax-incompatible sentences, which indicated that only when succeeding distributional cues were combined with phonological cues could word categorization occur.

The card sorting task results showed a significant difference between the two phonological coherency groups but did not show that the coherent group differed from chance level. The incoherent group showed the same phonology interference effect as in Experiment 1, in which they were more adept at sorting the nonwords according to their phonological characteristics and not according to their categories as defined by the pairings with the marker words. This was not surprising as there was no evidence that the incoherent condition obtained any knowledge of the distributionally defined categories.

The highly significant positive correlation between the similarity task and the card sorting results indicated that the two dependent measures were related and that both measured categorical knowledge. This correlation was intriguing it was only significant when the distributional cues occurred in the succeeding position relative to the category words, and not when the distributional cues precede the category words

This Experiment replicated St. Clair and Monaghan (2005) in that participants' could learn the two categories of words only with the combined succeeding distributional and phonological cues. No learning occurred with only succeeding distributional cues, or indeed with preceding cues. However, this difference did not indicate that succeeding cues were statistically better than preceding cues. The difference between preceding and succeeding cues was tested directly in the below combined analyses.

2.1.4.5. aAbB and AaBb combined results

A mixed ANOVA was conducted with time (1st and 2nd test session) and frequency (high and low) as the within subjects factors and phonological coherency (coherent and incoherent) and language structure (preceding or succeeding distributional cues; *aAbB* or *AaBb*) as the between subjects factor. In addition, two dummy variables (dialect and test counterbalance) were submitted into the ANOVA. The dependent variable was the number of correctly accepted compatible sentences (as similar to the artificial language) and correctly rejected syntax-incompatible and compatible sentences (as dissimilar to the artificial language).

The phonological coherency main effect was highly significant, with the coherent group performing better than the incoherent group, $F(1,32) = 14.73, p = .001, \eta_p^2 = .32$; M = 9.27 and M = 8.26, respectively. The time main effect was also significant, with the accuracy of discriminating between the compatible and incompatible test items increasing in the second test session, $F(1,32) = 5.08, p < .05, \eta_p^2 = .14$; M = 8.53 for the first test session and M = 9.00 for the second test session. The main effect of language structure was not significant, $F < 1$, as with all other main effects, $F < 1$. The phonological coherency by language structure interaction was marginally significant, $F(1,32) = 2.89, p = .099$. Although no post hoc analysis was done on this marginally significant interaction, an inspection of the means revealed there was an increase in the coherent condition when the distributional cues succeed the category words whereas there was no difference in the incoherent condition. See Figure 2.5 for a graph of this interaction. No other main effects or interactions were significant, frequency: $F(1,32) = 2.01, p = .17$; frequency by phonological coherence: $F(1,32) = 1.42, p = .24$; time by language structure: $F(1,32) = 2.11, p = .16$; frequency by time: $F(1,32) = 1.42, p = .24$; frequency by time by language structure: $F(1,32) = 1.13, p = .30$; all others $F < 1$.

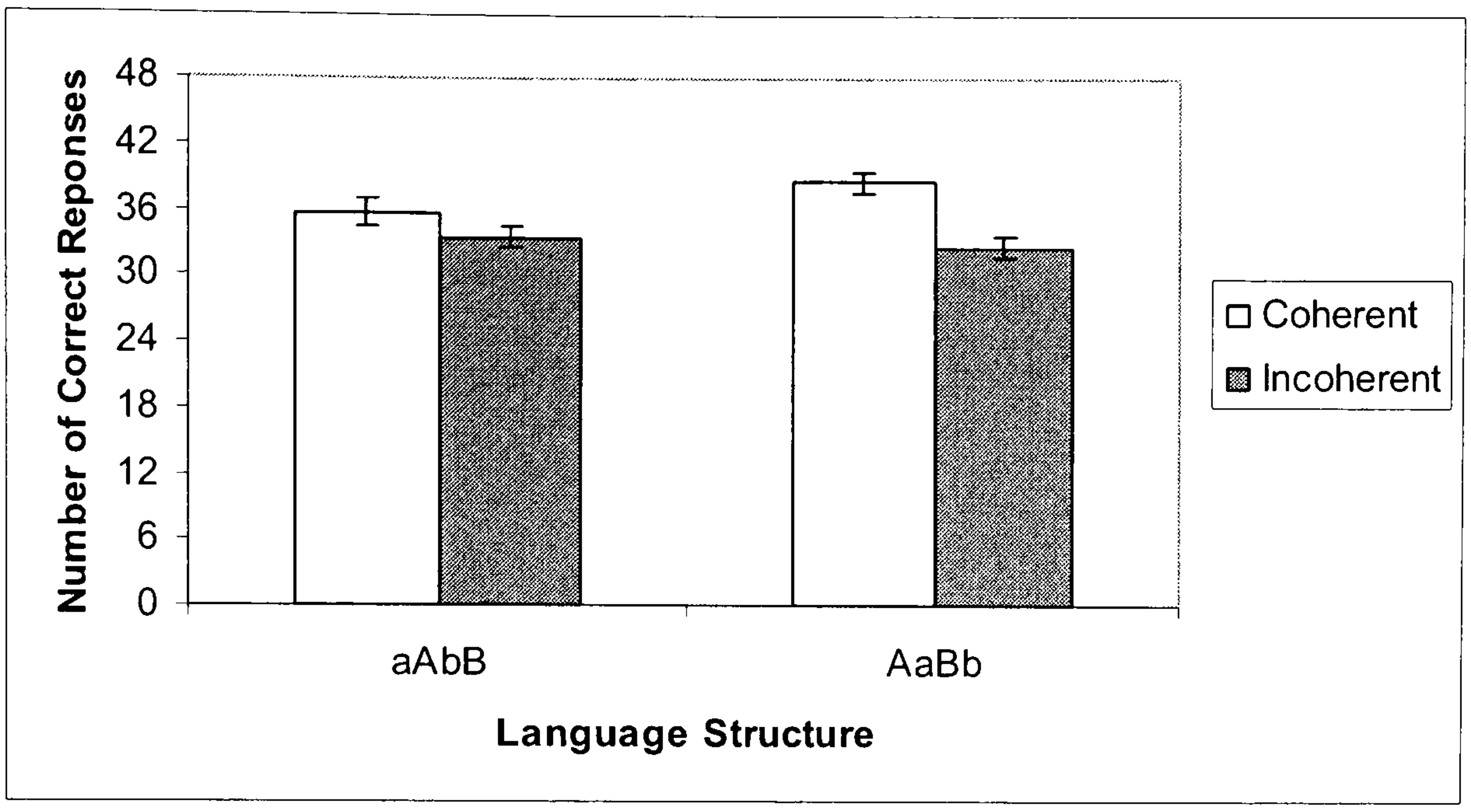

Figure 2.5 Number of correctly rejected and accepted test sentences (with standard error bars) for coherent and incoherent conditions in the *aAbB* and *AaBb* language structure conditions (Experiments 1 and 2).

As in both Experiments 1 and 2 the analysis with all of the data was confounded by the ceiling effects found with responses to syntax-incompatible sentences. Therefore a second mixed ANOVA was conducted with only compatible and incompatible sentences and included the same variables as the first ANOVA. The results did not change substantially, and thus was not reported in full.

The previous results indicated that the only difference between preceding and succeeding distributional cues occurred within the coherent condition. The coherent conditions were therefore directly compared (with compatible and incompatible test sentence data). Although the difference between preceding and succeeding distributional cues was not significant, $t(22) = -1.498, p = .15$, the group who received succeeding cues was significantly above chance whereas the group that received preceding cues was not different from chance level, $t(11) = 3.37, p < .05^{\dagger\dagger}$ and $t(11) = 1.25, p = .48$, respectively

2.1.4.6. Combined Discussion

The results of the combined analysis, although suggestive, were not conclusive. The above chance performance of the coherent condition when they received succeeding distributional cues was the best indicator for the idea that increased learning occurred when distributional cues occurred after the category words. It was interesting to note that any difference that was observed between succeeding and preceding distributional cues was due to the coherent condition; that is, any differences between the two language structures was carried by this condition. This may simply be due to the syntax-incompatible sentences, which may have led many of the participants in the incoherent condition to focus on whether the test item had one or two distinct marker words.

2.2. General Discussion

Across the two Experiments, there was no evidence of any categorical knowledge in the incoherent condition. However, the results for this condition in both Experiments 1 and 2 could have been unduly influenced by the presence of the syntax-incompatible test sentences; performance was at ceiling level for all participants, indicating they readily learnt the syntax of the sentences. Some participants did report that they were basing their judgements on whether the test sentences had one or two marker words. Therefore it is not surprising that the incoherent participants showed no evidence of any categorical knowledge and, in fact, may have randomly rejecting some of the *aAbB* and *aBbA* test sentences. The instructions explicitly state that half of test sentence are "dissimilar"; rejecting only the *aAaB* test sentences would only account for one fourth of the total. In order to reject approximately half of the total test items, it appeared that the incoherent participants randomly rejected some *aAbB* and *aAaB* test sentences.

The pattern was different when both phonological and distributional cues to category membership were inherent in the language; these participants were both able to correctly reject the *aAaB* test items and also showed limited evidence of an ability to discriminate the less obvious difference between the *aAbB* (compatible) and *aBbA* (incompatible) test items with preceding distributional cues. The ability to differentiate the compatible and incompatible test items in addition to correctly rejecting the syntax-incompatible sentence was firmly established in the coherent condition when there were succeeding distributional cues, but was not significant with preceding distributional cues.

An alternate possibility for the lack of learning found within the incoherent conditions was that the presence of the same nonwords that form phonologically defined categories in the coherent condition may have interfered with learning the categories from the distributional defined information in the incoherent condition. Within this experimental design it was impossible to create matched nonwords without phonological cues, as the phonological cues were embedded with the category words themselves, instead of syllable level "phonological markers" as previous literature has used (e.g., Brooks, Braine, Catalano, Brody, & Sudhalter, 1993; Frigo & McDonald, 1998). However, this previous research has found no differences between conditions that contained phonologically cued words randomly spread across two categories and conditions that contained matched nonwords without phonological cues (Frigo & McDonald, 1998). This lack of difference with more noticeable syllable level phonological cues indicated that the use of these phonologically similar category words spread over two categories to create the incoherent condition was unlikely to have been the cause of the lack of learning within this condition.

The conclusion from the incoherent conditions in Experiment 1 and 2 was that the data may be confounded due to the presence and the apparent reliance on the *aAaB* test items. Although further interpretation and conclusions from this data is impossible, it is worth noting that even in the *AaBb* language, which overall allowed higher learning, the participants were not able to learn about the categories from the distributional cues alone; the addition of phonological cues was necessary in order for grammatical categorization to occur.

Due to the above analysis it was difficult to draw any firm comparisons between the distributional cues only and phonological and distributional cues conditions. For this reason, discussion of the lack of the expected frequency by phonological coherency interaction is impossible as the data from the phonologically incoherent condition is uninterpretable. The phonology interference effect with the card sorting task is similarly difficult to interpret. The fact that the incoherent participants were more likely to sort the category words according to the phonological characteristics was likely due to the lack of knowledge from the distributional cues. When faced with all category words, the incoherent participants sorted the words according to both their phonological and orthographic similarities. Whether this lack of distributional knowledge was the result of the presence of the syntax-incompatible sentences or from a general inability to learn from preceding distributional cues cannot be determined.

Therefore, across both Experiments 1 and 2 the only condition to show any sign of categorical knowledge was the coherent condition, which had both phonological and distributional cues to category membership. There was no firm evidence of categorical knowledge in the coherent condition of Experiment 1. However, the results of Experiment 2 tell a different story as the participants in this condition were able to differentiate the critical *AaBb* (compatible) and *BaAb* (incompatible) test items at above chance levels.

Thus, the main finding from St. Clair and Monaghan (2005) that only succeeding cues induce categorical knowledge has been replicated. However, the combined analysis of Experiment 1 and 2 did not show a statistical difference between the *aAbB* and *AaBb* language structures. The lack of statistical difference in these comparisons may have been due to the syntax-incompatible test sentences which could have masked categorical knowledge in the coherent as well as the incoherent group, although to a lesser extent in the coherent group. Although the comparison was not significant the general pattern of results has been replicated; therefore two explanations as to why succeeding distributional cues allow more learning are discussed below. The first is based on the Rescorla-Wagner (1972) model of associative learning and the second is based on the notion that the marker words may be treated as inflections.

The R-W model of associative learning was developed in animal learning research, but the principles inherent in this model apply generally to all associative learning situations. The current application focuses on the idea that is it easy to form associations between two stimuli when the first stimulus always predicts the second stimulus. For the *AaBb* language, each particular category word, say A_i, always predicts that a is the next word. This one to one correspondence is very easy to learn as $A_i \rightarrow a$. Once each individual A_i has been associated with a all of the A_i's would be associated together due to their shared association with a, which would form Category A. The salience of learning the association with a will be high as the association is so prevalent, which leads to a high asymptote level of maximum associative strength. In addition, the salience of the initial word in each pair was predicted to be higher in the coherent condition due to the common phonological cues, but this will have to be tested further in future Experiments.

In the *aAbB* language the link is still one to one between each A word and a, but now in the opposite direction. When one hears a there are six possible A_i's that could be the next word. R-W model still predicts learning in this case, however it will be at a lower overall level as the associative strength between *a* and the six A_is will be lower than in the *AaBb* language. This is due to a reduction in the salience of the second word; as discussed in Chapter 1, reduction in the salience of the second stimuli reduces the overall asymptote level. The saliency in this case is reduced as there is much greater variability in the number of words that can occur in the second position. Thus, the correspondence between the marker and category words (and later categorization) is harder to learn within the *aAbB* language structure. See Figure 2.6 for a hypothetical graphical representation of the differing associative strength increases with preceding and succeeding distributional cues. As can be clearly seen, the associative strength with the *AaBb* language reaches a much higher asymptote than in the *aAbB* language, which has smaller associative strength increases and eventually reaches a lower asymptote.

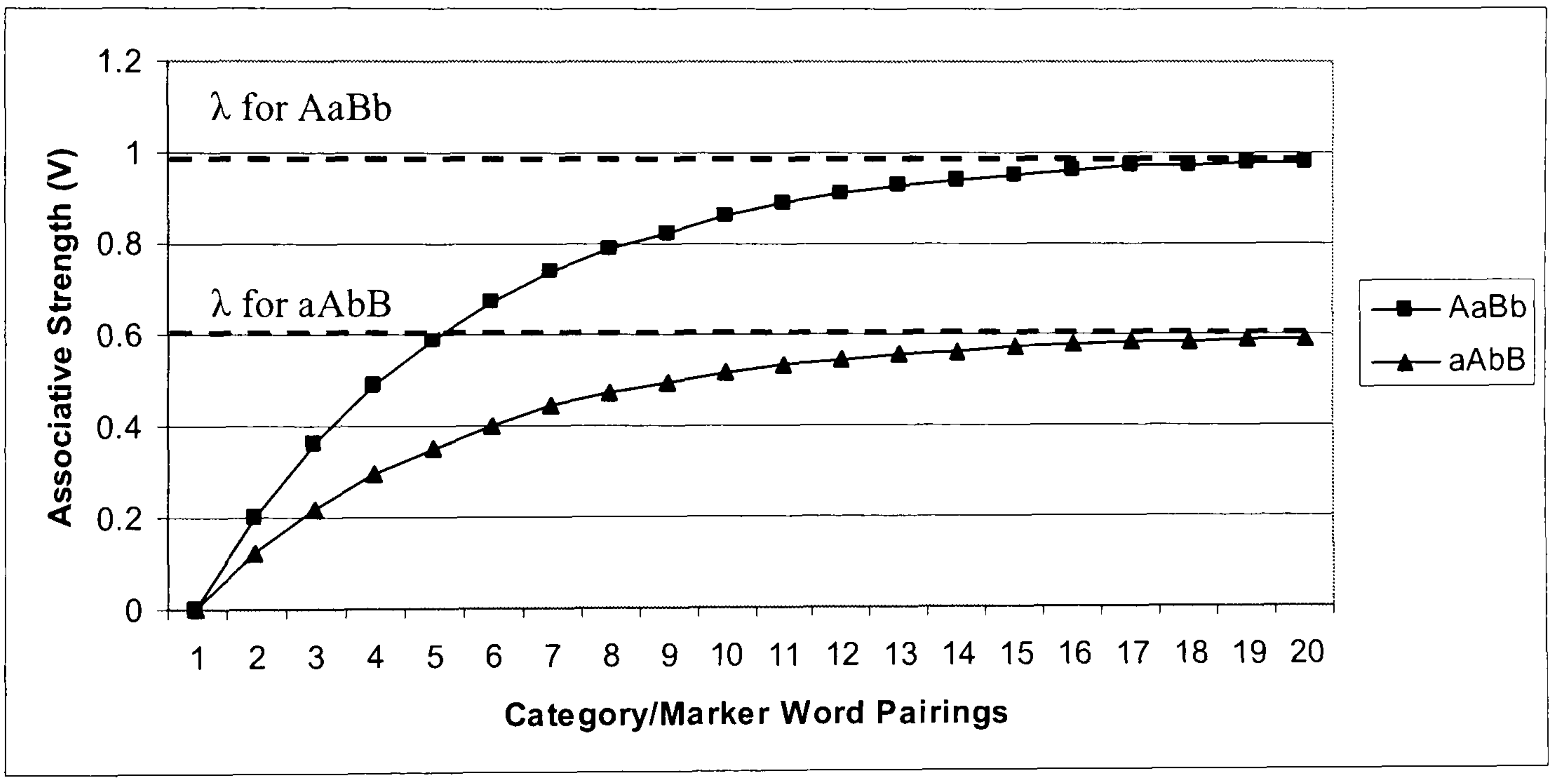

Figure 2.6 Theoretical associative strength increases with preceding (aAbB) and succeeding (AaBb) distributional cues.

The possible role of the phonological cues within this framework is that the phonological information may be supplementary cues that also have associative links with the relevant distributional cue and may aid in grouping the category words together. It cannot be determined at the present time whether the phonological cues must be present in order for categorization of all the *A* words to occur due to the confound within the incoherent condition. However, one hypothesis that was tested in further experiments was that the phonological cues to category membership are necessary to allow the participants to make full use of the distributional cues. If the phonological information provide cues that all the A_is are similar, this may increase and enhance the co-occurring distributional information.

The second hypothesis that might explain the overall findings is the idea that the particular phonological form of the marker words may have been more salient in the succeeding location. With the *alt* and *ong* marker words it seemed plausible that the initial vowel may attach to the end of category words (*dreng-ong*) more naturally than a marker word affixing to the beginning of the category word (*ong-dreng*). This may also relate to the prototype theory of associative learning categorization. If the marker words shared many of the same characteristics of typical succeeding cues in natural language, the marker words in the succeeding location may have higher salience than corresponding cues in the preceding location. This would lead to learning progressing faster with succeeding cues, as the associative strength asymptote may increase even further with the increased salience of the succeeding word. Conversely, the salience of these marker words in the preceding location may have been decreased due to a potential dissimilarity with naturally occurring preceding cues, which would lead to slower learning in the preceding cue Experiment. However, if marker words shared many characteristics of typical natural language preceding cues, salience would increase for preceding word cues and learning would be predicted to be progress at a faster rate.

In order to test this hypothesis further, the marker words were changed to a *cvc* phonological structure in Experiments 3 and 4, as it was thought that this structure would equate the salience of the preceding and succeeding word cues. Thus, *dreng-feg* and *feg-dreng* would be approximately equal in how well they could be thought of as prefixes or suffixes. Chapter 3 reports Experiment 3 and 4, which were replications of Experiment 1 and 2 with the marker word change. In addition to this change the syntax-incompatible test sentences were removed from these Experiments to avoid the problems associated with these test sentences.

3. Marker word phonology within bigram distributional contexts

Experiments 3 and 4 tested the above hypothesis by changing the phonological form of the marker words to a consonant-vowel-consonant (*cvc*) form, which was hypothesised to be less salient and less likely to be treated as a suffix in the *AaBb* experiment. Syntax-incompatible test sentences were excluded in Experiment 3 and 4 as these test sentences did not test categorical knowledge and may have, as stated above, masked any categorical knowledge in the incoherent condition. Experiment 3 was a replication of Experiment 1 and Experiment 4 was a replication of Experiment 2; both of these experiments used *cvc* marker words. It was hypothesised that succeeding distributional cues would still induce more categorical knowledge but that there would be conclusive evidence of categorical knowledge when the distributional cues preceded the category words as well. Furthermore, the incoherent conditions would show limited evidence of categorical knowledge, particularly with the high frequency category words.

3.1. Experiment 3 and 4

Experiments 3 and 4, although run as separate Experiments, are presented together. Each Experiment was analysed separately, and then combined to determine whether the difference in preceding and succeeding cues was significant. This combined analysis is justified as the same population of participants was used for both Experiments and the Experiments were run within a reasonable time frame (within a few months).

3.1.1. Participants

Twenty - six University of York undergraduate and graduate students participated in the *aAbB* Experiment. Participants had not taken part in any similar experiments. Two participants were excluded; one due to technical faults and the other due to external interference (fire alarm). The mean age for the participating sample was 19.63 years; range was 18 to 22. There were 17 females and seven males in the sample. All participants were native English speakers and were paid either £4 or received course credit.

Twenty - four University of York staff members, undergraduate and graduate students and participated in the *AaBb* Experiment. Participants had not taken part in any similar experiments. The mean age for the sample was 21.58 years; range was 18 to 44.

There were 14 females and 10 males in the sample. All participants were native English speakers and were paid either £3 or given course credit.

3.1.2. Materials

The training stimuli were almost identical to the stimuli used in Experiment 1, in that there was a phonologically coherent and incoherent condition and the frequency of the category words was manipulated so that half of the category words occurred twice as often as the remaining half. However, the marker words differed from Experiment 1; each participant received two of four possible nonwords, *feg, nep, vot,* and *zid.* Two of the marker words were randomly chosen for each participant. This was done to ensure that the results were not influenced unduly by any individual words.

In addition to the marker word change the test sentences were changed as well. It was apparent from four previous experiments (Experiments 1 and 2 and the two experiments in St. Clair & Monaghan, 2005) that the test items with the structure *aAaB* (or *AaBa*; syntax-incompatible sentences) did not allow any inference about categorical learning. These test items only indicated whether the participants had learned the overall syntax of the language and rejection of these test items was at ceiling level in all Experiments. Focusing on the syntax did not allow these participants to differentiate the compatible and incompatible test items (*aAbB* and *aBbA*), thus masking any categorical knowledge.

The *aAaB* test items were therefore excluded from this experiment and six additional *bAaB* (incompatible) test items per test session were created. Half of these items were made up of high frequency words and the other half low frequency words. This created two test sessions that contained 12 compatible test items (*aAbB*) and 12 incompatible test items (*bAaB*). The same test items were used in both the *aAbB* and *AaBb* conditions, with the only change being the location of the marker word. See Table C in Appendix C for details on the new test items in Experiment 3.

3.1.3. Procedure

The procedure was the same as in Experiment 1, in that there was two training and two testing sessions, followed by the card sorting task at the end of the Experiments.

3.1.4. Results and Discussion

All *t* tests and pairwise comparisons were Bonferonni corrected. The tests that were significant before correction have been marked to give a more complete picture of the data. Therefore, † indicates uncorrected significance at the .05 level, †† indicates significance at the .01 level, and ††† indicates significance at the .001 level.

3.1.4.1. *aAbB Experiment 3 results*

Overall accuracy was below the average of Experiment 1 and 2 at 55%. This was explained by the fact that this Experiment excluded syntax-incompatible sentences. The chance level of 67% with the syntax-incompatible sentences taken into account therefore dropped to 50% in this Experiment with these sentences excluded. A mixed ANOVA was conducted with frequency (high or low) and time (test session 1 or 2) as the within subjects factors and phonological coherency (coherent and incoherent) as the between subjects factor. In addition, one dummy variable was added as a between subjects factor (order of test sessions). The dependent variable was the number of correctly accepted and rejected test items.

There was a significant main effect of frequency, $F(1,20) = 25.79, p < .001, \eta_p^2 = .56$, with higher accuracy when the sentences contained high frequency rather than low frequency category words, M = 7.21; M = 5.88 of a possible 12, respectively. There was also a main effect of phonological coherency, with the coherent condition performing better than the condition with no phonological cues to category membership, $F(1,20) = 9.36, p < .01, \eta_p^2 = .319$; M = 7.02 and M = 6.06 out of a possible 12, respectively. The phonological coherency by frequency interaction was also significant, $F(1,20) = 7.28, p < .05, \eta_p^2 = .27$. There were no other significant main effects or interactions, time: $F(1,20) = 1.87, p = .19$; frequency by time by phonological coherency: $F(1,20) = 2.08, p = .17$; all others $F < 1$.

Pairwise comparisons were conducted on the frequency by coherency interaction and it was found that there was no significant difference between high and low frequency sentences in the coherent group ($p = .11$), nor between the coherent and incoherent high frequency sentences ($p = .53$). The significant difference was between the high and low frequency sentences in the incoherent condition ($p < .001^{\dagger\dagger\dagger}$) and between the low frequency sentences across the coherency conditions ($p = .001^{\dagger\dagger}$). Therefore, this interaction was due to the relative deficiency with low frequency sentences in the incoherent condition. When the category words were of high frequency the incoherent participants performed at equivalent levels to the coherent condition. This indicated that when words are of lower frequency, distributional and phonological cues are necessary to induce categorization. This finding is in line with the Monaghan et al. (2005) corpus analysis and artificial language learning findings. See Figure 3.1 for a graph of this interaction.

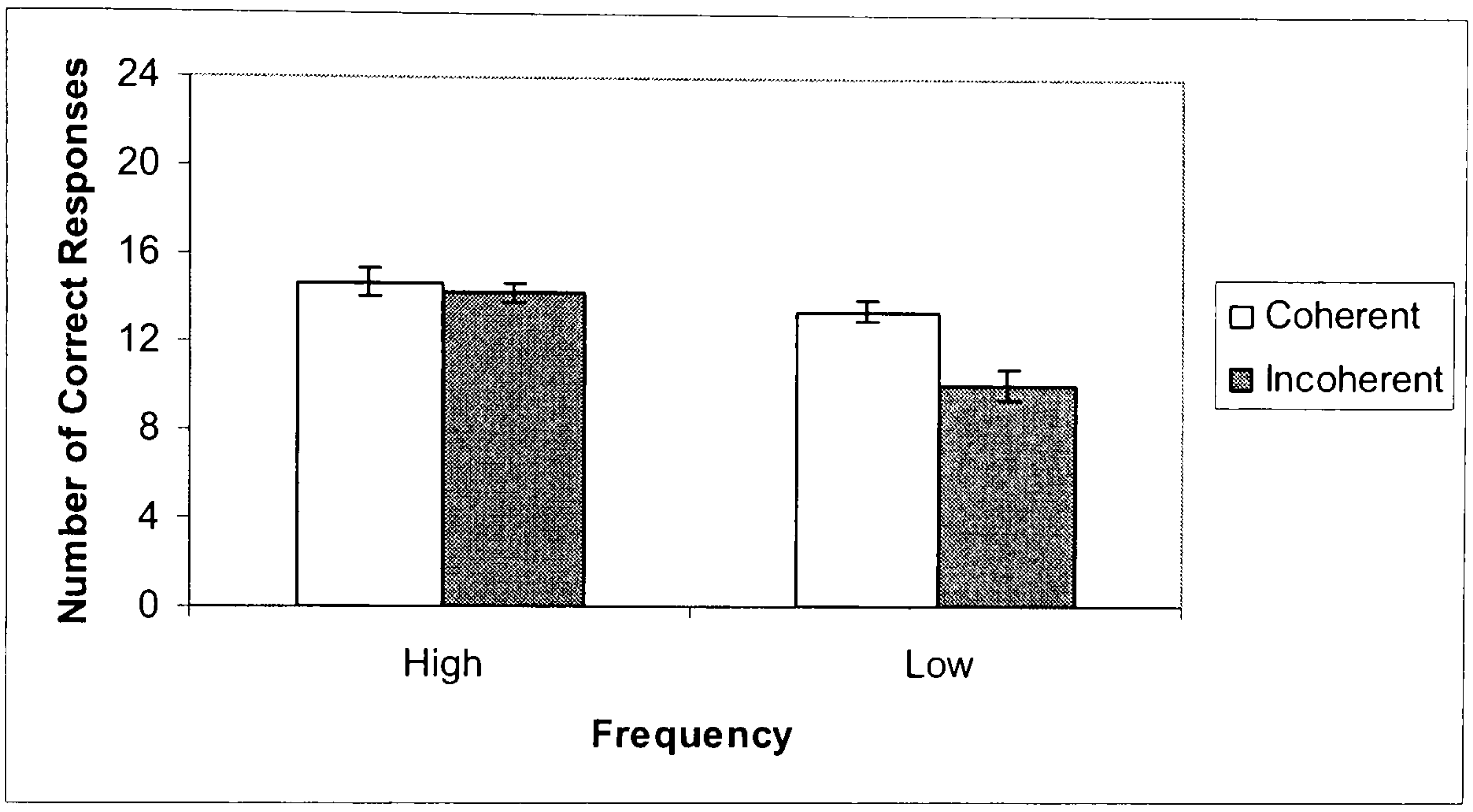

Figure 3.1 Number of correctly rejected and accepted test sentences (with standard error bars) for high and low frequency sentences in the coherent and incoherent conditions in Experiment 3.

In order to show a clear effect of categorical learning the two conditions were compared to chance level; in this case 24 as there were 48 total test items. The coherent condition was significantly above the chance level of responding, $t(11) = 4.87$, $p < .005^{\dagger\dagger\dagger}$, whereas the incoherent condition was not significantly different than chance level, $t(11) = .293, p = 1.0$, M = 28.08 and M = 24.25, respectively. However, this may not be an accurate story for the incoherent condition. As was discussed previously, the high frequency test items were similar across both coherency conditions. Therefore, the performance on the high frequency sentences from the incoherent condition were singled out to test against chance level. It was found performance on the high frequency sentences was indeed higher than chance level of 12, $t(11) = 4.91, p < .001$; M = 14.17.

The results of the card sorting task indicated a slightly less consistent picture. When looking at the distributionally defined categories there was no difference between the coherent and incoherent conditions, $t(22) = 0, p = 1.0$, M = 3.75 and M = 3.75, of a possible 6. Neither condition was above chance level, $t(11) = -.57, p = 1.0$ for the coherent condition and $t(11) = -.89, p = .78$ for the incoherent condition. When comparing the coherent and incoherent condition on categorization based on the phonological groupings, there was no difference between the groups, $t(22) = -1.53, p = .14$, M = 3.75 and M = 4.42 out of a possible 6, respectively.

A correlational analysis was conducted with the overall number of correctly rejected and accepted test items and the distributionally defined card sorting results. Overall, there was not a significant correlation between the similarity judgement task and the card sorting, $r = -.36, p = .08$, which was slightly surprising given the significant correlations in Experiment 1 and 2.

3.1.4.2. aAbB Discussion

This Experiment differs from Experiment 1 and St. Clair and Monaghan (2005) in that there was evidence of categorical learning within the coherency condition. In addition, the high frequency words were apparently categorized more readily than the low frequency words, as shown by a difference between the high and low frequency test sentences. The combination of these two main effects produced the previously illusive frequency by coherency interaction, such that the high frequency sentences in the incoherent condition were similar to the coherent condition. Correct responding to the low frequency sentences was depressed only when phonological cues were unavailable to aid categorization.

In sum, this indicated that preceding distributional cues were sufficient for grammatical categorization given enough exposure of the category-marker words pairings. However, when there were insufficient pairings, as with low frequency sentences, additional phonological cues were necessary in order for categorization to occur. This result contrasts with Experiment 1 where there was no difference between the high and low frequency items in the incoherent condition. The most plausible explanation for this result is the lack of syntax-incompatible sentences (*aAaB*) during the test phase. As previously discussed it was possible that categorical knowledge in the incoherent condition was masked by the incoherent participants' reliance upon looking for the repeated marker word. When these test items were removed the expected increase in correct responses to the high frequency test items in the incoherent condition appeared.

When looking at the card sorting task, the phonology interference effect was replicated in the incoherent group. This group was more likely to sort the words according to their phonological cues than by how they were distributionally defined within the language. This was not exactly surprising given the previous findings and indicated that the phonology interference effect was not due to the syntax-incompatible sentences. The coherent participants' results were surprising as they were unable to show any evidence of categorical knowledge within this task. Even though the coherent participants in Experiment 1 were much better at this task than the participants in Experiment 3, neither group performed at above chance levels which prevented any firm conclusions as to why only the incoherent condition were able to utilize the phonological cues in the present Experiment.

As a whole, these results provide convincing evidence that learning from preceding bigram distributional cues can occur either when accompanied with additional phonological cues or when the frequency of the distributional cue and category word is of sufficient frequency. This change in learning pattern from Experiment 1 to Experiment 3 can be attributed to two possible causes: the change in the marker words from *alt* and *ong* to a *cvc* phonology or the exclusion of the syntax-incompatible test sentences. These possibilities will be fully addressed in the general discussion. The most immediate question is whether, with *cvc* marker words, succeeding distributional cues still have an advantage over preceding distributional cues, as was found in Experiments 1 and 2. Experiment 4 will, therefore, replicate Experiment 3 with the only change being the succeeding location of the marker words (*aAbB* to *AaBb*).

3.1.4.3. *AaBb Experiment 4 results*

Overall accuracy in this experiment was 60.8%, which was an increase from 55% accuracy in Experiment 3. A mixed ANOVA was conducted with frequency (high and low frequency) and time (test session 1 and 2) as the within subjects factors and phonological coherency (coherent and incoherent) as the between subjects factor. In addition, a dummy variable was included as a between subjects factor (order of test sessions). The dependent variable was the number of correctly accepted and rejected test items.

There was a significant main effect of frequency, $F(1,20) = 9.66, p < .01, \eta_p^2 = .326$, with higher accuracy with high frequency sentences than with low frequency sentences, M = 7.75; M = 6.86 of a possible 12, respectively. Also replicating Experiments 2 and 3, there was a main effect of phonological coherency, with the coherent condition more accurately discriminating the test items than the incoherent condition, $F(1,20) = 12.50, p < .005, \eta_p^2 = .39$; M = 8.23 and M = 6.38 of a possible 12, respectively. There was also a significant main effect of time, $F(1, 20) = 10.93, p < .005, \eta_p^2 = .35$. Participants were better at discriminating the test items during the second test session; M = 6.75 for test session 1; M = 7.75 for test session 2, of a possible 12. Unlike the previous Experiment the frequency and phonological coherency interaction was not significant, nor were any other interactions; frequency by test session by phonological coherence: F(1, 20) = 1.23, p = .28; all others F < 1.

As in previous Experiments, it was necessary to show above chance differentiation of the test items. The coherent group was well above the chance level of 24, $t(11) = 4.77$, $p < .005^{\dagger\dagger\dagger}$, M = 32.92. However, the incoherent group did not differ from chance level, $t(11) = 1.23$, p = .735, M = 25.50. As the high frequency test items within the incoherent group were predicted to be above chance level, these items were tested against the chance level of 12, $t(11) = 2.18, p = .052$, M = 13.83; there was a marginally significant above chance level differentiation.

When looking at the distributional defined categories there was a trend towards a significant difference between the coherent and incoherent conditions, $t(22) = 2.31, p = .09^{\dagger}$; M = 4.58; M = 3.75, respectively. When compared against chance level of 3.91, the participants in the coherent condition sorted more cards than the expected by chance, $t(11) = 2.94, p < .05^{\dagger}$. The incoherent condition did not differ significantly from chance level, $t(11) = -.574, p = 1.0$. When comparing the coherent and incoherent groups on the phonological groupings, there was no difference, $t(22) = .98, p = 1.0$, M = 4.58 for coherent and M = 4.25 for incoherent.

The overall number of correctly accepted and rejected test items was correlated with the results of the distributional defined card sorting task. As in Experiment 3, the correlation was not significant, r = .30, $p = .15$.

3.1.4.4. AaBb Discussion

This Experiment again finds that categorical learning was most prevalent with both distributional and phonological information. Specifically, learning was only above chance in the coherent condition, with only a marginally significant above chance differentiation with the high frequency test sentences in the incoherent condition. Additionally, high frequency category words were categorized more easily than lower frequency category words. However, with succeeding distributional cues the high frequency advantage occurred for both the coherent and incoherent condition; thus, the hypothesized interaction was not present. As was found in Experiment 1, participants learned over the course of the experiment; they were more accurate at differentiating the compatible and incompatible test sentences during the second test session. This indicated that not only did increased pairings between the marker word and category words manifest itself into higher accuracy with high frequency test sentences, but higher accuracy in the second test session as well.

The card sorting task provided more consistent results with the similarity results than was found in Experiment 3. The incoherent group demonstrated the phonology interference effect, indicating that they did not learn anything substantial from the distributional cues and thus relied upon the phonological and orthographic regularities within the category words themselves. When looking at the distributional groupings, the coherent condition sorted more words than the incoherent condition and also sorted more category words together than expected by chance; incoherent condition did not differ from chance level. This supports the results of the similarity judgement task in that the combination of the phonological and distributional cues aided the segregation of the category words into two separate categories that is shown in the card sorting task. Succeeding distributional cues alone did not appear to aid this process significantly.

A more general comparison with Experiment 3 indicated that Experiment 4, with succeeding distributional cues, may have induced more categorical knowledge. If we only look at the overall accuracy, the *AaBb* language increased the accuracy from 55% (the *aAbB* Experiment) to almost 61%. It appears that the differences in the incoherent condition across the two Experiments were slight and that the increase in accuracy originated in the coherent condition. Furthermore, the card sorting results also support the idea of increased learning with the *AaBb* language structure, particularly with the coherent condition. These differences were directly tested to determine the legitimacy of these apparent differences.

3.1.4.5. Combined analysis

A mixed ANOVA was conducted with frequency of the category words (high and low) and test session (test session 1 or 2) as the within subjects variables and phonological coherency (coherent and incoherent) and language structure (*aAbB* or *AaBb*) as the between subjects factors. In addition, one dummy variable (test counterbalance) was submitted into the ANOVA as a between subjects factor.

Frequency of the category words was significant with participants responding correctly more often to the high frequency test items, $F(1, 40) = 32.69, p < .001, \eta_p^2 = .45$; M = 7.48 for high frequency and M = 6.37 for low frequency, of a possible 12. The manipulation of phonological coherency was also significant with better discrimination of the test items when both types of cues were consistent predictors of the categories, $F(1, 40) = 21.20, p < .001, \eta_p^2 = .35$; M = 7.63 for the coherent condition and M = 6.22 for the incoherent condition, of a possible 12. In addition, there was a significant effect of language structure, with higher accuracy when the marker words succeeded the category words (*AaBb*), $F(1, 40) = 6.20, p < .05, \eta_p^2 = .13$; M = 7.30 for *AaBb*, M = 6.54 for *aAbB*, of a possible 12. The frequency by phonological coherency interaction was significant, $F(1, 40) = 5.28, p < .05, \eta_p^2 = .12$, as was the time by language structure interaction, $F(1, 40) = 11.31, p < .005, \eta_p^2 = .22$. All other main effects and interaction were not significant; time: $F(1, 40) = 2.31, p = .14$; frequency by language structure: $F(1, 40) = 1.26, p = .27$; frequency by phonological coherency by language structure: $F(1, 40) = 1.79, p = .19$; frequency by time by phonological coherency: $F(1, 40) = 3.29, p = .08$; all others $F < 1$.

Post-hoc analyses using pairwise comparisons were conducted on the phonological coherency by frequency and time by language structure interactions. See Figure 3.2 for a graph of the phonological coherency by frequency interaction. There was a significant difference between the high and low frequency test items in both the coherent and incoherent conditions ($p < .05^{\dagger}$ and $p < .001^{\dagger\dagger\dagger}$, respectively). There was also a significant difference in the high frequency sentences between the coherent and the incoherent condition ($p < .05^{\dagger}$). Similarly, there was a highly significant difference in the low frequency sentences between the coherent and the incoherent condition ($p < .001^{\dagger\dagger\dagger}$). The interaction was, in part, caused by a relatively small difference between the high frequency sentences across the coherency conditions and a relatively large difference between the low frequency sentences across the coherency conditions. In addition, there was a larger difference across the two frequency levels in the incoherent condition when compared to the coherent condition.

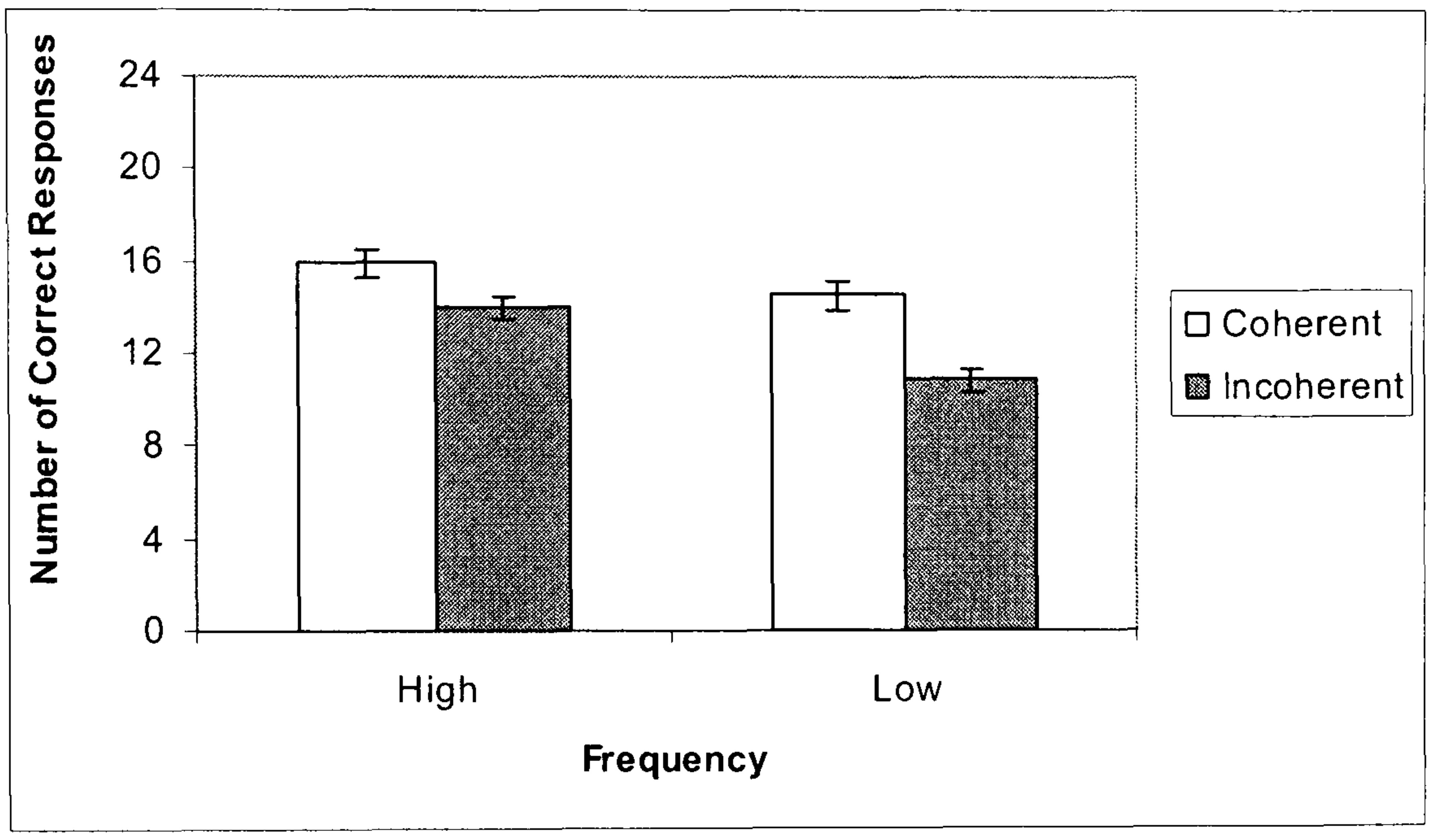

Figure 3.2 Number of correctly rejected and accepted test sentences (with standard error bars) for high and low frequency sentences in the coherent and incoherent conditions in the *aAbB* and *AaBb* Experiments combined (Experiment 3 and 4).

For the time by language structure interaction, no difference was found between scores at the first test session in the *aAbB* and *AaBb* levels of language structure (p = 1.0) but there was a significant difference between performance during second test session between the two Experiments, with higher discrimination of the compatible and incompatible test items when the distributional cues were in a succeeding position ($p = .001^{\dagger\dagger\dagger}$). There was no difference between the two test session in the *aAbB* experiment, but there was a significant difference in the *AaBb* experiment, with higher learning in the second test session (p = .20 and $p = .001^{\dagger\dagger\dagger}$, respectively). The interaction was due to the increased accuracy in the second test session in the *AaBb* language structure condition, which differed significantly from all other test session and language structure combinations. See Figure 3.3 for a graph of this interaction.

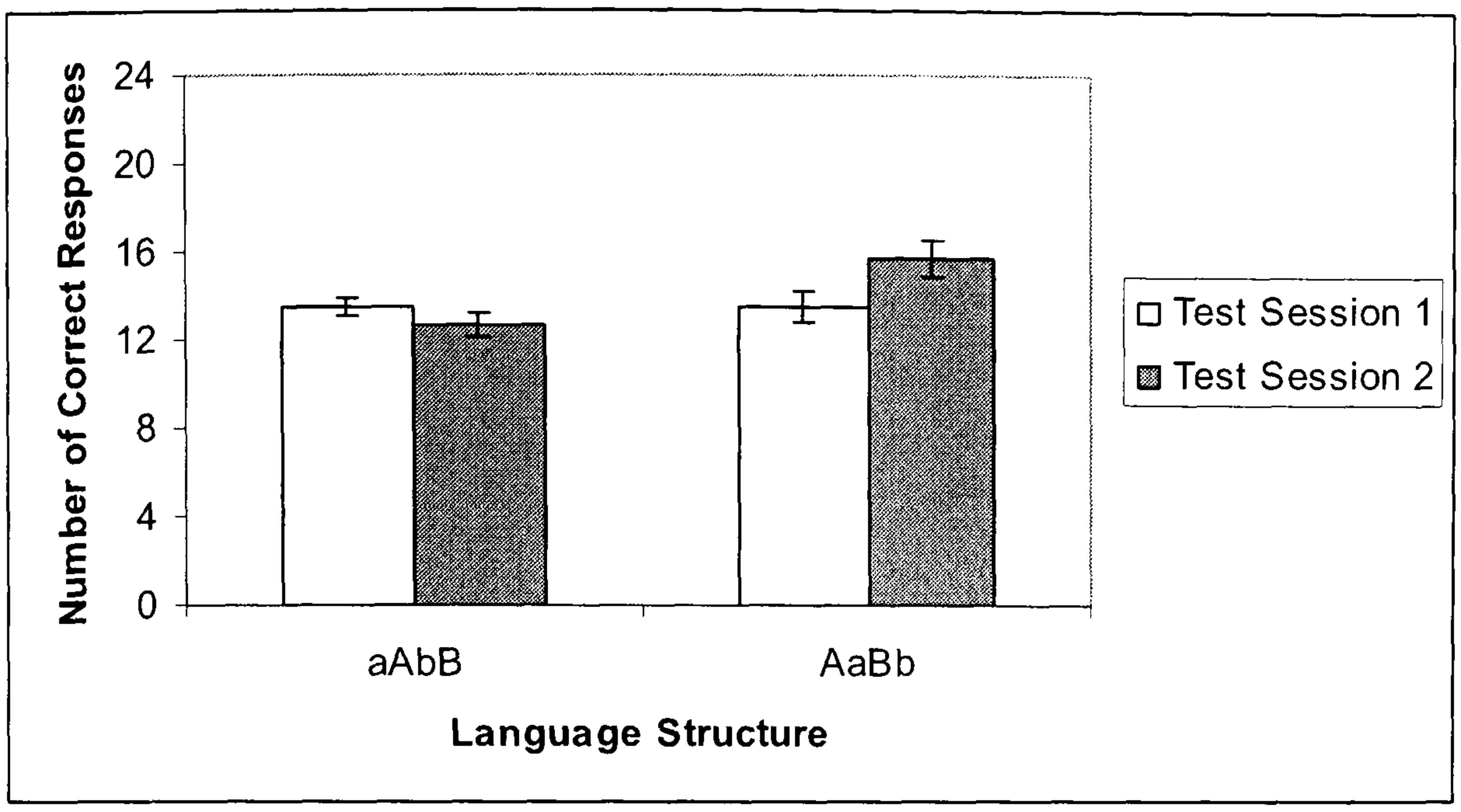

Figure 3.3 Number of correctly rejected and accepted test sentences (with standard error bars) for the *aAbB* and *AaBb* language structures in test sessions 1 and 2.

The two language structure conditions were tested against chance level to determine whether categorical knowledge has been learnt in each condition. For the preceding language structure condition, performance was significantly above the chance level of 24 (there were 48 total test items; $t(23) = 3.06, p < .05$; M = 26.17). As expected, the language structure with succeeding distributional cues was also well above chance level, $t(23) = 3.89, p < .005$; M = 29.21.

3.1.4.6. Discussion from combined analysis

The combined analysis of Experiments 3 and 4 provided evidence of the increased ability of succeeding distributional cues to aid grammatical categorization; this convincing evidence was missing from the combined analysis of Experiments 1 and 2, but was strong in current analysis. As expected there was a significant difference between the coherent and incoherent conditions across both of the Experiments; phonological and distributional cues allowed more learning than distributional cues alone. The main effect of category word frequency was also significant, which indicated that the high frequency words were more readily learnt due to the increased pairings with the distributional cue.

The frequency by coherency interaction, which was found in Experiment 3, was also present in the combined analysis. The interpretation discussed in Chapter 3 holds for this analysis as well. When only distributional cues to category membership were available, categorization occurred only if there were sufficient category-marker word pairings, as in the case of the high frequency category words. However, when phonological cues to categorical membership were also available both high and low frequency test items were categorized, as the phonological cues compensated for the lack of distributional information.

Additionally the time by language structure interaction was of importance in the combined analysis. This was explained by the increase in accuracy in the second test session of the succeeding cue Experiment while there was no difference between the test sessions in the preceding cue Experiment. This result was intriguing as it was the preceding cue Experiment in Chapter 2 that showed marginally significant increase in learning across the test sessions (with only compatible and incompatible test sentences). This was explained by the general slower growth of the associations between the marker and category words that occurs with preceding, but not succeeding cues. The question that must be answered is why there is increased learning with more exposure in the succeeding cue Experiment, but not the preceding cue Experiment, while the opposite was true in Chapter 2. This issue will be further discussed in the General Discussion.

3.2. General Discussion

The results of the combined analysis confirmed the previous trend that preceding and succeeding distributional cues differentially influence categorical learning. Succeeding distributional cues induced more categorical knowledge than preceding distributional cues. However, this only occurred when both phonological and distributional cues were present to aid the categorization process. As discussed previously, the Rescorla-Wagner model of associative learning can account for the advantage of succeeding distributional cues. The association between a particular category word (A_i) and the corresponding marker word (a) will be easier to form when the marker word (or the predictor cue to category membership) occurs after the category word. If the predictor cue occurs in the initial position any of the six possible A_is could occur as the next word, therefore strong associations between individual A_is and a will take longer to form.

However, from the data it appears that supplementary phonological cues that characterize the category words, A_is, must be present in order for the categorization of all the A_i words to occur. When these cues were missing, as in the incoherent condition, the *AaBb* language did not induce increased learning over *aAbB* language. This can be explained in terms of the feature hypothesis of categorical associative learning. The main feature in each category was the consistent co-occurrence with the marker word. As described above, this association will be stronger with succeeding cues than with preceding cues. However, this one feature was not sufficient for grammatical categorization to occur; additional phonological cues within the words also needed to be present. These multiple, correlated cues may also form associations with the marker word: it is the combined associative strength of all the shared phonological features across the category words along with the individual associations of each category word that allows the creation of the category. These ideas are illustrated within Figure 3.4.

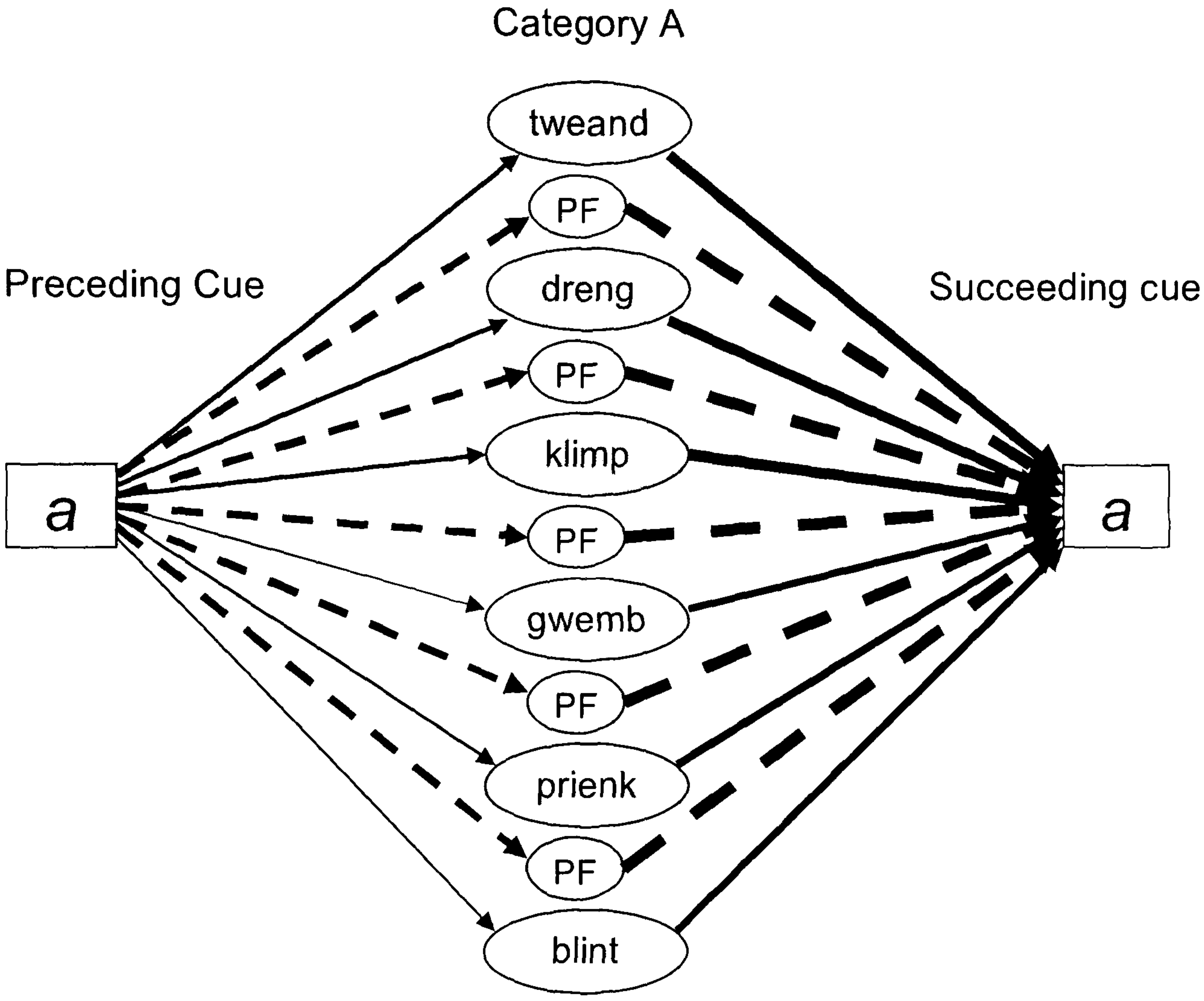

Figure 3.4 Diagram illustrating the relative associative strengths between the preceding and succeeding cues and the individual category words (high frequency category words are in grey) and the phonological features (PF) within the coherent category words[5].

---

[5] The thickness of the arrows from the preceding cue to the category words indicated increased asymptote level, or maximum associative strength, as the different features of the second stimulus differed in salience. The phonological features were most salient, therefore had higher asymptote levels. The thickness of the arrows from the category words to the succeeding cues indicated rate of associative strength increases, with thicker lines representing faster learning, as it was the first stimuli that had saliency differences. In this case, the phonological features with higher saliency would result in quicker learning.

The associations displayed in Figure 3.4 all behave in the manner predicted by the R-W model; in every case, the associations with succeeding distributional information should form faster and easier. As illustrated above, the lines (representing associative strength) are stronger with succeeding cues than with preceding cues; in addition, the associative strength of the phonological features defining the category is predicted to be higher than the category words themselves. This was due to the fact that the abstract phonological features of the category were present during every category-marker word pairing, whereas the pairing of a specific category words and the marker word was relatively infrequent in comparison. The above diagram also illustrates the difference between high and low frequency category words; the associative strength of low frequency category words is lower than the high frequency category words due to fewer pairings with the marker word.

Thus, this explanation accounts for the higher learning with succeeding cues when there are phonological and distributional cues. The increased learning with the high frequency category words in the incoherent condition is also explained by this account. The incoherent condition was significantly above chance with high frequency sentences in the preceding cue Experiment. This comparison was nearly significant with the succeeding cue Experiment, with a *p* value of .052. This overall finding is explained by the increased associative strength between the high frequency words with both the preceding and succeeding cues. More generally, this account predicts categorical learning with only distributional cues only if there are sufficient category-marker word pairings. Valian and Coulson (1988) found learning from a written artificial language with preceding bigram distribution cues with four total training sessions. The written form of the language as well as the numerous training sessions would have allowed the preceding cues sufficient pairings with the category words to allow categorical learning.

As mentioned in Chapter 2 there was a slight worry that the presence of two groups of six phonologically similar words may interfere with the ability of the incoherent condition to learn effectively from the distributional cues. However, the results from the incoherent group in both Experiment 3 and 4 indicated that the high frequency category words were grouped together but the lower frequency words were not. This indicated that words with common phonological characteristics across the categories did not interfere with the categorization of these high frequency category words.

The expected frequency of the category words by phonological coherency interaction was found in Experiment 3 and the combined analysis, but not Experiment 4. This interaction is of importance as it confirms the results found from both corpus and experimental investigations by Monaghan et al. (2005). Distributional cues were found to be both useful and usable in categorizing high frequency words together, but phonological information within the category words was found to be of particular importance with lower frequency words. Experiment 3 provided the first evidence that within the auditory language learning domain that categorization of high frequency words can progress with only distributional cues but that additional phonological cues were necessary for lower frequency words. The lack of this interaction in Experiment 4 must be thought of in terms of the increased effectiveness of the succeeding distributional cues. The lack of the predicted interaction was caused by increased accuracy with high frequency test items in both the coherent and incoherent conditions. In other words, the effectiveness of the succeeding distributional cues increased with the number of category word-marker words pairings; there was no effect within the coherent condition with preceding distributional cues, perhaps because the associative strength asymptote had been reached by the combined influence of the category word and phonological cue associations. As discussed above, the asymptote level for succeeding distributional cues was higher than for preceding cues perhaps allowing further learning with increased pairings for only succeeding cues.

There was a large increase in learning from the succeeding cues from the first test session to the second test session. This increase in learning was not mirrored with preceding cues; there was no increase in learning with the second training session. The comparison with Experiments 1 and 2 not only found no hint of a difference in the succeeding cue Experiment with the *alt* and *ong* marker words but a marginally significant increase in learning from the first to the second test session within the preceding cue Experiment. These differences can only be attributed to two possible causes: the exclusion of the syntax-incompatible test sentences or the change in the marker word phonology from *vc*/*vcc* to *cvc*. It is premature to discuss these two alternate causes as subsequent Chapters will provide further evidence.

The results of the card sorting task of Experiment 3 and 4 supported the similarity judgement results. The lack of above chance levels of category word groupings with preceding distributional cues, but above chance performance in the case of succeeding distributional cues indicated increased categorical knowledge with succeeding cues. However, the similarity tasks did show above chance levels of categorical knowledge with the preceding distributional cues as well. One possibility for the lack of congruence between the two measures of categorical knowledge is that the *aAbB* language induced categorical learning that was mainly implicit, and therefore was unlikely to be expressed with a more explicit task. The succeeding distributional cues produced more robust learning that may have translated into some form available for explicit use and this could have led to higher card sorting results. Research has shown that implicitly learned knowledge may become explicitly available due a process of feeling-based inferences; participants can become aware that the are basing their judgements on implicitly learned knowledge without explicit access to this knowledge (Dienes & Perner, 2002). Feeling based inferences about which category words should go together may be easier with succeeding distributional cues due to higher overall learning than with preceding distributional cues.

Another possibility for the card sorting results is that the knowledge gained in the *aAbB* language may have been insufficient for the card sorting task, while perfectly sufficient in the similarity judgement task. In the *AaBb* language, each individual category word was predictive of the succeeding marker word, but in the *aAbB* language it was the marker word that predicted the individual category words. With the similarity task, both marker word and category word were presented together, so both types of dependencies could be used to guide responses. However, only the category words were given in the card sorting task. The fact that the category words in the *AaBb* language were directly associated with the marker word may have aided in grouping the correct words together in the card sorting task. Conversely, in *aAbB* language the marker word to category word direct associations would not have aided performance in this task. However, it must be pointed out that even in the *aAbB* language there were partially predictive succeeding cues; whenever an individual A_i category word was the second word of the phrase it predicted that the marker word from the opposite category would be the succeeding word. These partially predictive cues may have aided the *aAbB* language participants in the card sorting task, but this would likely have been lower than the level from the AaBb language, in which the succeeding word was always predicted by the category word. This is the pattern found within these Experiments.

The last topic that needs consideration in regards to these Experiments is the prototype hypothesis. With *vc/vcc* marker words no concrete evidence of categorical learning was found with preceding distributional cues, but there was significant learning with succeeding distributional cues. This was hypothesized to be influenced by the *vc/vcc* marker words potential similarity to natural language succeeding cues (or suffix inflections), and therefore increased saliency of the succeeding marker words. This hypothesis is supported by previous work by Brooks et al. (1993) who found that suffix cues with the phonological form of simply a vowel or *vc* induced grammatical categorization. However, with *cvc* marker words significant learning was found with both types of distributional cues, but significantly more learning was found with the succeeding cues. While the syntax-incompatible test items created problems within the *vc/vcc* Experiments the evidence did not point to this being a major confound for the coherent condition. The coherent condition of the *aAbB* Experiment with *vc/vcc* marker words did not show any categorical learning, while the coherent condition with *cvc* marker words showed significant categorical knowledge. Previous research does not provide any answers in this case; Frigo and McDonald (Frigo & McDonald, 1998) used *cvc/ccvc* prefix (as well as suffix) cues but did not find categorical learning unless both prefix and suffix cues were combined. However, this may have been due to many differences in experimental design, specifically the presence of additional phonological cues within the current Experiments that were missing from the Frigo and McDonald studies.

The main question now becomes why *cvc* marker words induced categorical knowledge while the *vc/vcc* marker words did not. One possibility is that the *vc/vcc* marker words were simply incongruent in the preceding location, and therefore led to a reduced saliency with those particular marker words. This may account for the lack of learning with *vc/vcc* marker words, but this therefore implies increased saliency of the preceding cues with *cvc* marker words. Increased saliency with *cvc* preceding marker would lead to an increase rate of learning, allowing the associations between the marker words and category words to reach the asymptote level faster. This increased saliency could either be due to neutral saliency between preceding and succeeding *cvc* cues or the possibility that *cvc* marker words might be more phonologically similar to existing prefixes in the English languages and the *vc/vcc* marker words were similar to suffixes. In other words, *cvc* marker words may be more prototypical of prefixes while *vc/vcc* markers may be consistent with a suffix prototype. This would lead to increased saliency of the *cvc* marker words in the preceding location and increased saliency of the *vc/vcc* marker words in the succeeding location.

In order to test this explanation a comprehensive phonological analysis of English prefixes and suffixes was conducted. Chapter 4 presented this phonological analysis; it was predicted that this analysis would provide a more complete explanation as to the difference between *vc/vcc* and *cvc* marker words as well as provide specific predictions about the phonological form of the marker words for future Experiments.

4. Analysis of suffix and prefix phonology

This Chapter investigated the hypothesis that the phonology of the marker words may influence how well participants can acquire categorical knowledge. Specifically, the salience of the marker words may increase or decrease depending on the phonological similarity between preceding and succeeding distributional cues in natural language. Prefixes and suffixes were the chosen preceding and succeeding distributional cues under investigation, although it must be acknowledged that there are many highly useful preceding and succeeding distributional cues at the word level in natural language. In particular, high frequency function words are potentially instrumental in the language acquisition process (McDonald, 1997).

The phonological differences between function and content words (or closed and open class words) have been extensively investigated in previous research; the findings of Shi et al. (1998) and Monaghan et al. (2005) were described in detail in Chapter 1. However, the general findings are that function words are unstressed words which are generally phonological minimal in comparison to content or open class words (Cutler, 1993; Shi, Morgan, & Allopenna, 1998). Interestingly, there also is some evidence that points to function words and affixes having functionally the same grammatical role. Patients with language disorders display similar patterns of addition, deletion and substitution errors with affixes and functions words; additionally, the affix errors were mostly suffix errors (although this may simply be due the fact that there are a greater number of suffixes; Patterson, 1980). This finding indicates that perhaps function words and affixes perform the same grammatical role; in particular, both may be useful in the acquisition of grammatical categories.

The previous research gives a good picture of the phonological form of function words. However, the general phonological form of affixes has not been investigated in detail previously. This analysis focused on prefixes and suffixes mainly due to the observation that the *vc/vcc* marker words may have been more natural as a suffix (*dreng-ong*) than as a prefix (*ong-dreng*). This chapter will investigate whether the pattern of learning found in Experiments 1 to 4 was influenced by marker words similarity (and thus increased saliency) with the common phonological characteristics of affixes.

To do this, the general and specific phonological characteristics of prefixes and suffixes within the English language were investigated. It was predicted that the *vc/vcc* phonological form would be more typical of suffixes and the *cvc* phonological form would be common in prefixes. This phonological similarity would lead to increased salience of *vc/vcc* marker words in the succeeding location and *cvc* marker words in preceding location; increased salience leads directly to higher learning ability. The amount of *cv* prefixes and suffixes was also investigated as it was predicted the *cv* phonological form would be more common in prefixes than suffixes leading to increased salience of *cv* marker words in the preceding location. This prediction was mainly due to the location of the vowel; if *dreng-ong* allowed the marker word to act similarly to suffixes perhaps marker words ending in a vowel would similarly attach to the beginning of words.

The general phonological form of the beginning and ending phonemes of affixes was also investigated. More specifically, the number of affixes that both began and ended with consonants (C_C), began with a vowel and ended with a consonant (V_C), or began with a consonant and ended with a vowel (C_V) were investigated. It was hypothesized that the overall phonological form of beginning and ending phonemes would be relevant as to whether the marker words were more salient as a preceding or succeeding cue.

To further determine the general characteristics of beginning and ending phonemes of prefixes and suffixes these phonemes were analysed in terms of numerous phonological measures. The consonants were analysed in terms of voicing and manner and place of articulation and the vowels in terms of height, position and roundedness. There were two reasons for focusing on the beginning and ending phonemes; these phonemes are key in attaching prefixes and suffixes to the root word, thus fully understanding the phonological characteristics of these phonemes is important when investigating the prototype hypothesis. Secondly, previous research has found that the beginning and ends of words provide useful information to the prediction of the grammatical category of the word (Onnis & Christiansen, 2005) and had also been shown to be useful in categorizing words in artificial language paradigms (Frigo & McDonald, 1998).

The results of these analyses were tested in Experiments 5, 6 and 7 which altered the marker words to conform to the general phonological characteristics of prefixes and suffixes. The saliency, and thus ease of learning, of preceding and succeeding cues was predicted based on the phonological similarity between the general form of prefixes and suffixes and the marker words.

4.1. Prefix and Suffix Phonological Analysis

4.1.1. Compilation, transcription and description of analysis

A comprehensive list of prefixes and suffixes were taken from Fudge (1984). The prefixes and suffixes that were labelled as compound forming elements or were described as "not really a suffix [prefix]" (for example, -elle, Fudge, 1984, p. 67) were excluded. There were 57 prefixes and 217 suffixes, but the prefixes occurred more frequency than the suffixes, $U = 2638, p < .001, r = -.39$. The prefix/suffix difference reflects the suffixing preference in English morphology, which is a potential language universal preference (Cutler, Hawkins, & Gilligan, 1985).

Each affix was transcribed according to the International Phonetic Alphabet (IPA). The transcription was done by hand by a native English speaker with the aid of an online dictionary with IPA transcriptions. For instance, the *pre-* prefix was transcribed as /pɹi/ and described as [plosive (bilabial, unvoiced)-approximant (alveolar, voiced)-close/front (unrounded)]. The general characteristics of the affixes as well as the specific phonological features of the beginning and ending phoneme were then coded in detail. Table 4.1 shows the general features and specific features coded and analysed for both consonants and vowels. The number of *cvc, vc* and *cv* affixes directly related to the prototype hypothesis. The beginning and ending phonemes pairs (e.g., C_C) measurement refers to the general phonological form of prefixes and suffixes, which extended the previous analysis to determine the prototypical general form of prefixes and suffixes. Phoneme number was analysed to determine whether there were any general length differences in prefixes and suffixes. Finally, the more specific phonological characteristics of beginning and ending phonemes were analysed in an attempt to determine what characteristics allow affixes to attach to the root word.

Table 4.1 Phonological features of prefixes and suffixes

	Phonological features
	cvc, vc and *cv* affix numbers
	Beginning and ending phoneme pairs
	Phoneme number
Consonants	Manner of articulation
	Place of articulation
	Voicing
Vowel	Height
	Position
	Roundedness

The general aim of the analysis was to determine which phonological features differentiated the two types of affixes. Distinctive phonological features may influence participants' ability to treat marker words as prefixes or suffixes. If suffixes were found to be more likely to have an initial vowel than prefixes, then marker words with initial vowels may have increased salience in the succeeding location. Conversely, if prefixes are found to be more likely to begin with a consonant and end with a vowel, then perhaps marker words of that phonological form would increase learning in the *aAbB* language but not in the *AaBb* language.

Both type and token analyses were conducted. The type analysis counted each affix only once but the token analysis weighed each affix based on the frequency of the affix occurring either at the beginning or end of a word. The English CELEX database was used to generate the frequency data (Baayen, Pipenbrock, & Gulikers, 1995), which was taken as frequency per million words. An affix that was also a word (e.g., -ability) was only counted when the affix occurred within a larger word (e.g. learnability). In addition, the affixes that consisted of only one letter (e.g. a-, o-, -t) were not included in the token analyses, as the frequency counts for these affixes would include all words that either began or ended with that particular letter.

4.1.2. Results

All multiple Chi-Square and Mann-Whitney U tests were Bonferonni corrected. As this correction is very conservative, the tests that were significant before correction have been marked to give a more complete picture of the data. Therefore, † indicates uncorrected significance at the .05 level, †† indicates significance at the .01 level, and ††† indicates significance at the .001 level.

All measures except phoneme number and affix frequency were categorical variables. The dependent categorical variables were the number of prefixes and suffixes that fell into each category. For instances, in the phonemes pairs C_C category, the number of prefixes and suffixes that both began and ended in a consonant was the dependent variable. Chi-square tests were conducted on all categorical measures directly comparing prefixes and suffixes in order to determine whether the two affix types differed on each measure. However, some chi-square comparisons had expected values too low for the Chi-square test to be appropriate. In these cases, the Fisher's Exact test replaced the Chi-square test (Everitt, 1992). Cramer's V was used as a measure of association between the two variables to estimate the effect size. This measure was especially useful with the token analysis, as the Chi-Square tests were unduly influenced by the high N numbers. Cramer's V was not influenced by the number of data points in the analysis, and therefore provided an impression of the effect size for whether a particular phonological feature was important in differentiating prefixes and suffixes.

The continuous variables (frequency and phoneme number) had non-normal distributions (skewness and kurtosis values were significant) which could not be corrected by log transformation, square root transformation or inverse transformation. Therefore, non-parametric Mann-Whitney U tests replaced standard *t*-tests.

Logistic regression was used to evaluate the effectiveness of all phonological measures of the beginning and ending phonemes combined. Three sets of logistic regressions were conducted (type and token analyses) for the C_C, C_V, and V_C phoneme pairs (the V_V pair only occurred in suffixes, so comparisons between affix groups could not be conducted). The phoneme pair data was analysed as beginning consonant features could not be analysed with beginning vowel features, as an affix with a beginning consonant would have missing values for beginning vowels; thus, one logistic regression model with all data was impossible.

The *cvc*, *vc* and *cv*, phoneme pairs and phoneme number analyses were presented first followed by the combined logistic regression analyses of the individual consonant and vowel phonological measures. The distribution of beginning and ending consonants and vowels across prefixes and suffixes was also investigated. Finally, each individual phonological measure was tested to determine what cues individually differentiate prefixes and suffixes. These individual analyses are summarized in this Chapter and are presented in detail in Appendix I.

4.1.2.1. cvc, vc and cv prefixes and suffixes

The prevalence of *cvc*, *vc* and *cv* affixes were compared across the prefix and suffix group to determine whether each phonological form type was more typical of prefixes or suffixes. For the type analysis, the distribution of *cvc*, *vc* and *cv* affixes indicated that *cvc* affixes were more common as prefixes (56.7%), *vc* affixes were more common as suffixes (61.3%) and *cv* affixes were more common as prefixes (54.5%). However, the distribution did not differ significantly, $\chi^2(2) = 3.02, p = 1.0, V = .17$.

Table 4.2 Prevalence of *prefix and suffix* affixes within the *cvc, vc* and *cvc* categories in the type and token analyses.

		Affix Type	
		Prefix	*Suffix*
Type Analysis	*cvc* *(N = 30)*	56.7%	43.3%
	vc *(N = 62)*	38.7%	61.3%
	cv *(N = 11)*	54.5%	45.5%
Token Analysis	*cvc* *(N = 36218)*	86.7%	13.3%
	vc *(N = 269485)*	34.9%	65.1%
	cv *(N = 91148)*	68.1%	31.9%

The same pattern of results was found in the token analysis, but the results were much stronger as 86.7% of *cvc* affixes were prefixes, 65.1% of *vc* affixes were suffixes, and 68.1% of *cv* affixes were prefixes. In this case, prefixes were significantly different than suffixes, $\chi^2(2) = 55108.45, p < .05^{\dagger\dagger\dagger}, V = .37$, with a medium effect size. These results were consistent with the increased learning from preceding *cvc* marker words that was found in Chapter 3. The fact that *cvc* prefixes were more common than *cvc* suffixes supported the idea that *cvc* marker words were more salient in the preceding location. Similarly, the results from the *vc* affixes indicated that this phonological form was more common with suffixes, which would support the hypothesis of increased saliency for the *vc*/*vcc* succeeding marker word cues. This would increase the learning found with succeeding but not preceding cues. Similarly, the *cv* phonological form was more common in prefixes than in suffixes, which would indicate that *cv* marker words may be more salient as preceding cues.

Table 4.3 Prevalence of *cvc, vc,* and *cv* affixes within the prefix and suffix categories for the type and token analyses

		cvc, vc, and cv		
		cvc	*vc*	*cv*
Type Analysis	*Prefixes (N = 47)*	36.2%	51.1%	12.8%
	Suffixes (N = 56)	23.2%	67.9%	8.9%
Token Analysis	*Prefixes (N = 187381)*	16.8%	50.1%	33.1%
	Suffixes (N = 209470)	2.3%	83.8%	13.9%

However, the results within the prefix and suffix groups provide more insight into possible effects on marker word salience. The *vc* phonological form was common in both prefixes and suffixes, but was more predominant in the suffix group. The *cvc* and *cv* phonological forms accounted for 16.8% and 33.1% of all prefix tokens, while the *cvc* form was virtually not represented in the suffix group, accounting for only 2.3% of the tokens, while the *cv* phonological form accounted for 13.9% of all tokens. The results of this analysis suggested that the *cvc* and *cv* marker words could lead to increased saliency in preceding cues. However, there may be no saliency difference in preceding and succeeding *vc* marker words; this phonological form was the most common in both suffixes and prefixes.

4.1.2.2. Beginning and ending phoneme pairs

The previous analysis was extended beyond the particular *cvc*, *vc* and *cv* affixes to the more general combined phonological form of beginning and ending phonemes. Specifically, this analysis looks at the differences across prefix and suffix groups in the amount of affixes that begin and end with a consonant (C_C), begin with a consonant but end with a vowel (C_V), and begin with a vowel and end with a consonant (V_C). It was predicted that the C_C and C_V phonological patterns would be more common in prefixes, but that the V_C pattern would be more common in suffixes. The V_V phonological form only occurred in the suffix case, and was therefore excluded from this analysis.

There was no difference in the general distribution of prefixes and suffixes that have C_C, C_V, and V_C phonological patterns within the type analysis, $\chi^2(2) = .60, p = 1.0$, V = .16. See Table 4.4 for a summary of the data.

Table 4.4 Initial and ending phoneme pairs for prefixes and suffixes with type and token analyses

		Initial and Final Phonemes		
		C_C	*C_V*	*V_C*
Type Analysis	*Prefixes (N = 55)*	40.0%	14.5%	45.5%
	Suffixes (N = 206)	31.1%	12.6%	45.6%
Token Analysis	*Prefixes (N = 196349)*	16.6%	35.0%	48.4%
	Suffixes (N = 265458)	4.2%	11.2%	80.8%

The token analysis was significant, $\chi^2(2) = 65611.42, p < .05^{\dagger\dagger\dagger}$, V = .40, with a medium effect size. There were significant amounts of C_C, C_V, and V_C prefixes, with about half of all prefixes of the V_C phonological form. However, the vast majority of suffixes were also of the V_C phonological form, while 11.2% of suffixes begin with a consonant and end with a vowel. Although the V_C phonological form characterized the majority of affixes, the differences between prefixes and suffixes within each phonological form strengthened the result from the previous section. While keeping in mind that there were more total suffixes than prefixes, 69% of V_C affixes were suffixes and only 31% were prefixes. The opposite pattern was found with the C_V affixes, with 69.8% prefixes and only 30.2% suffixes. Additionally, 75% of C_C affixes were prefixes, which supported the previous hypothesis of increased saliency with preceding *cvc* marker words due to similarity with prefixes. The V_C and C_V patterns also support the hypothesis of a possible increased saliency for *vc*/*vcc* succeeding marker words and *cv* preceding marker words.

4.1.2.3. Phoneme number

Phoneme number was analysed to determine whether prefixes and suffixes differed in length, which would have implications for the marker words used in the artificial language learning studies. For the type analysis there were significantly more phonemes in suffixes than in prefixes, $U = 3387, p < .005^{\dagger\dagger\dagger}, r = -.33$. There was also a significant difference between the affix conditions in the token analysis as well, $U = 2586365730, p < .005^{\dagger\dagger\dagger}, r = -.26$. However, as can be seen in Table 4.5, the pattern reversed when frequency of the affixes was accounted for. There was no difference between prefixes and suffixes when looking at median phoneme lengths, but the means indicated that prefixes contained more phonemes than suffixes, which was the opposite pattern found in the type analysis.

Table 4.5 Descriptive statistics for phonemes number of prefixes and suffixes for the type and token analyses

		Phoneme Number					
		Median	*Range*	*Min.*	*Max.*	*Mean*	*Std.Dev.*
Type Analysis	*Prefix (N=57)*	2.0	4	1	5	2.54	.91
	Suffix (N=219)	4.0	8	1	9	3.90	1.78
Token Analysis	*Prefix (N=196023)*	2.0	3	2	5	2.24	.488
	Suffix (N=357001)	2.0	8	1	9	1.97	.814

The reversal in length preference between the type and token analysis was explained by the fact that the suffix group included many longer, but infrequent suffixes, while the prefix group contains shorter and more frequent prefixes. Thus, when frequency of the affixes is taken into account, the pattern reversed with the large number of highly frequent short suffixes overwhelming the smaller amount of longer, lower frequency suffixes. This result is consistent with the general property of language that frequently used morphemes tend to become shorter over time (Bloomfield, 1933).

4.1.2.4. Combined beginning and ending consonants

The next three analyses investigated how well the combined phonological features of beginning and ending phonemes predict affix status. The first analysis investigated the combined influence of beginning and ending phonemes in C_C affixes. Backwards logistic regression analyzed the beginning and ending consonant manner and place of articulation and voicing, which were entered in one block. There were four steps to the logistic regression model with the type data. The variables were analyzed during each step in terms of the change in log likelihood ratio, which indicated if the model changed significantly when that variable was excluded. During each step the variable that altered the overall model the least (highest significance level) was removed in the next step, until only variables that impacted the model significantly were left in the final step.

Table 4.6 Variables included in the beginning and ending consonant (C_C) logistic regression model (type data) and change in log likelihood in the model when the variable was removed from the model.

	Variables in equation	*Change in model Log Likelihood Ratio*	*Significance*
Step 4	Beginning consonant manner of articulation	37.44	< .001
	Beginning consonant place of articulation	14.31	.014
	Ending consonant manner of articulation	10.38	.035

Table 4.6 shows that only beginning consonant manner and place of articulation and ending consonant manner of articulation aided the C_C logistic regression model in differentiating prefixes and suffixes. Beginning consonant voicing, ending consonant place of articulation and voicing were not found to be significant predictors of affix status. The model did well at differentiating suffixes from prefixes, with an overall accuracy of 83.9% with the Step 4 variables. The model chi square was significant, $\chi^2(18) = 42.36, p < .001, R^2 = .57$ (Nagelkerke). The Beta values are reported in Table D in Appendix D.

Table 4.7 Variables included in the beginning and ending consonant (C_C) logistic regression model (token data) and change in log likelihood in the model when the variable was removed from the model.

	Variables in equation	*Change in model Log Likelihood Ratio*	*Significance*
Step1	Beginning consonant manner of articulation	6258.77	< .001
	Beginning consonant place of articulation	10892.19	< .001
	Beginning consonant voicing	317.55	< .001
	Ending consonant manner of articulation	18907.35	< .001
	Ending consonant place of articulation	46.88	< .001
	Ending consonant voicing	26.19	< .001

The logistic regression model with token C_C data revealed a different pattern of results as all six variables contributed significantly to the model. However, it was clear from the change in the model log likelihood ratio that the significant values in the type analysis were the most influential factors in the token analysis. The token analysis model was 97.2% accurate at categorizing the prefixes and suffixes. The model chi square was significant, $\chi^2(18) = 41580.94, p < .001, R^2 = .91$ (Nagelkerke). The Beta values are reported in Table E in Appendix E.

4.1.2.5. *Combined beginning consonants and ending vowels*

Backwards logistic regression analyzed beginning consonant manner and place of articulation and voicing with ending vowel height, position of articulation and roundedness (C_V). These variables were entered into the equation in one block. As there were no differences between the type and token logistic regression analyses, only the token analysis will be reported. All variables except ending vowel roundedness (which the logistic regression procedure suppressed) were entered into the Step 1 logistic regression model. Ending vowel position of articulation was the only variable dropped from the final model, as can be seen in Table 4.8.

Table 4.8 Variables included in the beginning consonant and ending vowel (C_V) logistic regression model (token data) and change in log likelihood in the model when each variable was removed from the model.

	Variables in equation	*Change in model Log Likelihood Ratio*	*Significance*
Step 2	Beginning consonant manner of articulation	51129.05	< .001
	Beginning consonant place of articulation	3458.74	< .001
	Beginning consonant voicing	97718.79	< .001
	Ending vowel height	27729.89	< .001

Interestingly, this logistic regression model kept beginning consonant voicing, which the previous C_C type data model eliminated. It appeared that beginning consonant voicing became more important when the ending phoneme was a vowel, which indicated that some variables importance in predicting affix status differed depending on the overall phonological form of the affix. The Step 2 model differentiated suffixes from prefixes with an overall accuracy of 92.9%. The model chi square was significant, $\chi^2(13)$

$= 95710.79, p < .001, R^2 = .88$ (Nagelkerke). The Beta values are reported in Table F in Appendix F.

4.1.2.6. Combined beginning vowels and ending consonants

Backwards logistic regression was used to analyze the beginning vowel height, position of articulation and roundedness with ending consonant manner and place of articulation and voicing. These variables were entered as one block. With the type data, beginning vowel height, ending consonant manner of articulation and voicing were dropped from the final model. There were four steps and the final logistic regression model only included the beginning vowel position of articulation, roundedness and ending consonant place of articulation.

Table 4.9 Variables included in the beginning vowel and ending consonant (V_C) logistic regression model (type data) and change in log likelihood in the model when each variable was removed from the model.

	Variables in equation	*Change in model Log Likelihood Ratio*	*Significance*
Step 4	Beginning vowel position of articulation	11.43	.043
	Beginning vowel roundedness	19.53	< .001
	Ending consonant place of articulation	13.32	.010

This result continued a trend of the beginning phoneme phonological features contributing more towards the differentiation of prefixes and suffixes. The final model was significant, $\chi^2(11) = 33.21, p < .001, R^2 = .36$ (Nagelkerke). Interestingly, the R^2 value was much lower in this analysis than in the prior analyses with beginning consonants. However, the overall accuracy of predicted prefixes and suffixes was still high at 85.9% in the final model. The Beta values are reported in Table G in Appendix G.

The token analysis did not show the same results as the type analysis, as all variables contributed to aid in predicting affix status. When frequency was taken into account beginning vowel height, and ending consonant manner of articulation and voicing influenced the model and there was a decrease in the effectiveness of the model when these variables were removed, as can be seen in Table 4.10. The model containing all variables was 87.0% accurate at discriminating between prefixes and suffixes. The model chi square was significant, $\chi^2(24) = 174609.20, p < .001, R^2 = .60$ (Nagelkerke). The Beta values are reported in Table H in Appendix H.

Table 4.10 Variables included in each step in the beginning vowel and ending consonant (V_C) logistic regression model (token data) and change in log likelihood in the model when each variable is removed from the model.

	Variables in equation	*Change in model Log Likelihood Ration*	*Significance*
Step1	Beginning vowel height	18215.82	< .001
	Beginning vowel position of articulation	12235.26	< .001
	Beginning vowel roundedness	1660.99	< .001
	Ending consonant manner of articulation	50139.59	< .001
	Ending consonant place of articulation	63279.38	< .001
	Ending consonant voicing	2643.84	< .001

These combined analyses have demonstrated that the fine grain phonological features of the beginning and ending phonemes provide accurate information that can serve to differentiate prefixes from suffixes. The combined analysis and the individual results (presented in detail in Appendix I) will aid in determining the most distinctive cues of prefixes and suffixes that may relate to the salience of the preceding and succeeding marker words.

4.1.2.7. Beginning phoneme type

There was no difference in the distributional of beginning consonants and vowels between prefixes and suffixes in the type analysis, $\chi^2(1) = .99, p = 1.0$, V = .06, but there was a significant difference in the token analysis, $\chi^2(1) = 13406.34, p < .05^{\dagger\dagger\dagger}$, V = .16. Both types of analysis showed virtually no difference between consonants and vowels for prefixes but the suffixes were more likely to have a vowel as the beginning phoneme than a consonant, as can be seen in Table 4.11. The difference was more pronounced within the token analysis. This result supported the earlier evidence that *vc* marker words may be more similar to suffixes than prefixes, and thus increase the saliency of succeeding distributional cues more than preceding distributional cues.

Table 4.11 Beginning phoneme type for prefixes and suffixes with type and token analyses

		Beginning Phoneme Type	
		Consonant	*Vowel*
Type Analysis	*Prefix (N = 59)*	50.8%	49.2%
	Suffix (N = 227)	43.6%	56.4%
Token Analysis	*Prefix (N = 196349)*	51.6%	48.4%
	Suffix (N = 361752)	35.6%	64.4%

4.1.2.8. *Ending phoneme type*

As Table 4.12 illustrates the ending phoneme was more likely to be a consonant across both affix types in both type and token analyses. The type analysis did not reveal any differences between prefixes and suffixes, $\chi^2(1) = 1.58$, $p = 1.0$, V = .08, but there were significant differences between prefixes and suffixes in the token analysis, $\chi^2(1) = 26035.06$, $p < .05^{\dagger\dagger\dagger}$, V = .24. The pattern for ending consonants was stronger in suffixes than in prefixes. However, when an affix does end in a vowel, it is likely to be a prefix (63.2%; token analysis). This supports the hypothesis that marker words which end in vowels are more likely to aid learning of preceding distributional information rather than succeeding information.

Table 4.12 Ending phoneme type for prefixes and suffixes with type and token analyses

		Ending Phoneme Type	
		Consonant	*Vowel*
Type Analysis	*Prefix (N = 55)*	85.5%	14.5%
	Suffix (N = 216)	77.8%	22.2%
Token Analysis	*Prefix (N = 196349)*	65.0%	35%
	Suffix (N = 270983)	85.2%	14.8%

4.1.2.9. Beginning and ending phonological measures

Individual Chi-Square tests for beginning and ending consonant voicing, manner and place of articulation and beginning and ending vowel height, position and roundedness were conducted. The results for the type analysis are summarized in Table 4.13; individual explanation of each of the analyses was presented in Appendix I.

Overall, the results for the type analysis indicated beginning consonant manner of articulation and voicing were the most influential beginning phoneme phonological measures, while ending consonant place of articulation was the only phonological measure for the ending phonemes that differentiated prefixes and suffixes individually. Interestingly, these variables were not always present in the logistic regression analyses for type data presented earlier in this Chapter. Ending consonant place of articulation was only significant when combined with beginning vowels. Similarly, beginning consonant voicing was only significant when combined with ending vowels. However, beginning manner of articulation was present in both the C_C and C_V analyses.

Table 4.13 Chi-square (or Fisher's exact test) and Cramer's V results for the type analysis with significance levels for 11 categorical phonological feature measures.

Type of analysis		*Chi-Square*	*Significance*	*Cramer's V*
Beginning Phoneme	*Consonant manner of articulation*	19.79	< .05[††]	.39
	Consonant voicing	6.85	.14[††]	.23
	Consonant place of articulation	5.84[±]	1.0	.23
	Vowel height	6.01[±]	1.0	.21
	Vowel position	4.78[±]	1.0	.18
	Vowel roundedness	2.17[±]	1.0	.13
Ending Phoneme	*Consonant manner of articulation*	1.13	1.0	.07
	Consonant voicing	.44	1.0	.05
	Consonant place of articulation	18.82	< .05[††]	.30
	Vowel height	8.95[±]	1.0	.48
	Vowel position	9.67[±]	1.0	.48
	Vowel Roundedness	6.11	1.0	.33

[±] Fisher's Exact Test

As can be seen in Table 4.14 the results for the token analyses differed from the type analyses. Details of these results can be found in Appendix I. Due to the frequency weighting of the affixes, all of the consonant and vowel phonological features significantly differentiated prefixes from suffixes individually. As the Chi-Square tests can be unduly influenced by large N numbers the effect size for each test may be a more reliable indicator of how well each measure differentiated the affix groups.

Table 4.14 Chi-square and Cramer's V results for the token analysis with significance levels for 11 categorical phonological feature measures.

Type of analysis		*Chi-Square*	*Significance*	*Cramer's V*
Beginning Phoneme	*Consonant manner of articulation*	51079.70	<.05[†††]	.47
	Consonant voicing	17060.99	<.05[†††]	.27
	Consonant place of articulation	82667.76	<.05[†††]	.60
	Vowel height	72442.19	<.05[†††]	.47
	Vowel position	44389.50	<.05[†††]	.37
	Vowel Roundedness	7499.74	<.05[†††]	.15
Ending Phoneme	*Consonant manner of articulation*	41161.62	<.05[†††]	.34
	Consonant voicing	1706.42	<.05[†††]	.07
	Consonant place of articulation	37904.40	<.05[†††]	.33
	Vowel height	16424.55	<.05[†††]	.40
	Vowel position	16572.30	<.05[†††]	.39
	Vowel Roundedness	11596.29	<.05[†††]	.33

The beginning and ending consonant results were not surprising, as all these variables were all included in the relevant combined logistic regression models. However, the Cramer's V value for ending consonant voicing was particularly low, indicating that this measure may be more useful in differentiating the affix groups when combined with additional phonological feature information. Specific results found that beginning consonant phonemes in prefixes were often plosives and fricatives, whereas the beginning consonants of suffixes were more widely distributed between plosive, fricative, nasal, approximant and lateral approximant phonemes. This finding indicated that perhaps the beginning plosives and fricatives in prefixes allowed higher salience, which may be beneficial for preceding grammatical category cues. The suffixes, being succeeding cues, do not necessarily need this additional salience. Conversely, the majority of ending consonants in suffixes were found to be nasals, which have been found to aid in attaching suffixes to the root word (Bauer, 2003b).

The beginning vowel position and roundedness features were consistent with the results of the V_C logistic regression model, but beginning vowel height was also highly significant. The ending vowel position was highly significant at differentiating the affix groups individually, but did not aid when combined with beginning consonant and the other two ending vowel phonological features. Overall, the vowel measures indicated that there was more variety in vowel usage for ending prefixes and beginning suffix locations; in other words when an affix is attaching to the root word. There was less variety in vowels at the beginning and ending of words.

4.1.3. Discussion

This chapter investigated the phonological characteristics of prefixes and suffixes in order to determine whether the changes in learning found with phonologically different marker word cues in Experiments 1 to 4 can be explained due to phonological similarity to affixes. The general conclusions from these analyses were:

1. There are fewer prefixes than suffixes, but prefixes occur more frequently.
2. The *cvc* and *cv* phonological form occurred most often in prefixes while the *vc* form occurred most often in suffixes.
3. V_C phonological pattern was more common in suffixes and affixes with ending vowels were likely to be prefixes.
4. Beginning consonant phonemes promote higher salience in prefixes.
5. Ending consonant phonemes allow prefixes to attach to the following root word and promote higher salience in suffixes.
6. Ending vowel phonemes consist of two vowels for suffixes, but a wider range of vowels for prefixes.

One important description of affixes was that there are far more suffixes than prefixes; in other words, succeeding affix cues are more common in English than preceding affix cues. This finding is in line with the hypothesis that succeeding cues allow faster learning whereas learning will proceed slower with preceding cues. The fact that suffixes tend to be more prevalent across most languages (Cutler, Hawkins, & Gilligan, 1985) is an indication that language may be designed around what we are able to learn from; Experiments 1 to 4 in this thesis provide convincing evidence that succeeding cues induce both higher and faster learning than preceding cues. The preference for suffix information across languages seems consistent with this general learning preference.

The significant difference within the frequency of prefixes and suffixes supports this point as well. There are fewer total prefixes but they occurred more frequently than suffixes. This is precisely what is needed from preceding cues according to the associative learning ideas discussed in terms of Experiments 1 to 4. Preceding cues, due to the lower overall associative strength with following words, need to occur more frequently than succeeding cues to allow language learning. The lower number of total prefixes is consistent with the fact that these cues are less helpful in language learning, and the higher frequency of these prefixes is a necessary requirement for any learning to occur at all from preceding distributional cues. This finding is supported by evidence that the English language has generally shifted away from prefixation over the centuries, and that prefixes have also become more "wordlike" in that they are now more able to function away from the root word than in the past (Bauer, 2003a).

The remainder of the conclusions relate to the general phonological characteristics of prefixes and suffixes. It was hypothesized that the saliency of preceding and succeeding marker words might vary depending on the phonological similarity to naturally occurring preceding and succeeding distributional cues. In the *vc/vcc* Experiments marker words did not induce categorization in the preceding location but did in the succeeding location. However, with *cvc* marker words categorical information was learnt in both locations. The results of this Chapter indicated that the *cvc* structure was more consistent with prefixes than suffixes and the *vc* structure occurred more frequently in suffixes than prefixes, although was common to both types of affix. This provides direct support that the *cvc* structure is similar to preceding language cues and the *vc* structure is similar to succeeding cues; these similarities may have increased the saliency of the *cvc* preceding marker words and the *vc/vcc* succeeding marker words.

Additionally, it was found that *cv* affixes were more likely to be prefixes; this leads to the prediction that *cv* marker words may be more salient as preceding cues.

The combination of beginning and ending phoneme pairs confirmed the findings from the *cvc*, *vc* and *cv* affixes by looking at the general beginning and ending form of the affixes. The majority of suffixes began with a vowel and ended with a consonant, while prefixes were more flexible in beginning and ending phoneme combinations. This result supported the idea that *vc/vcc* marker words were similar in phonological form to suffixes; thus, the saliency of the *vc/vcc* marker words would be increased in succeeding location, but not in the preceding location. The pattern that ending vowel phonemes are usually from prefixes supports previous results with *cv* affixes and the hypothesis that *cv* marker words may be more salient as preceding cues.

Generally, the finding of higher prevalence of *cvc* prefixes seems to correspond with Bauer (2003a) finding that English prefixes are becoming more wordlike and able to function separately from the root word. This finding may seem to indicate that preceding cues may generally be more "wordlike" than succeeding cues. Also the findings that preceding word cues are better indicators for grouping words according to grammatical category than succeeding word cues should also be taken into account as well when investigating marker word phonology (Redington, Chater, & Finch, 1998). It may be that typical preceding cues in English are generally word or "wordlike" cues and succeeding cues are generally suffix or bound morpheme cues. This will be investigated in further chapters.

The remaining analysis investigated how the phonological details of the first and last phoneme differ between prefixes and suffixes. This detailed analysis allowed a closer look at the specific characteristics of prefixes and suffixes; specifically what phonological details make prefixes distinct from suffixes. These phonological details may increase or decrease marker word salience depending on similarity (or dissimilarity) to prefixes and suffixes. The first main point from this analysis was the findings that prefix beginning consonant phonemes were frequently plosive and fricative consonants; this was not the case for suffixes. The predominance of these two phonemes may aid in increasing the salience of the prefixes as these are distinctive phonemes. As discussed above, learning from preceding information is not as easy as learning from succeeding information; beginning consonant manner of articulation may serve to increase the saliency of naturally occurring preceding language cues, which would increase learning from these cues.

The beginning vowel characteristics also reliably differed between prefixes and suffixes. Beginning vowel position, roundedness and height combined to differentiate prefixes and suffixes in the logistic regression models. Overall, there was a greater variety in beginning suffix vowels than beginning prefix vowels. This was likely to be due to the difference between word initial beginning vowels in prefixes and word internal beginning suffix vowels. The higher consistency in beginning prefix vowels may also serve to increase the saliency of prefixes by creating similarity across the prefix category. Alternatively, this result may simply be due to the phonotactic constraints of word initial and word internal vowels.

The majority of ending consonants in prefixes were nasals whereas there was a greater variety in ending consonants in suffixes. Previous research indicates that ending nasal consonant may aid the prefixes in attaching to the root word (Bauer, 2003b). Conversely, suffixes usually end with a plosive; this is interesting as it relates to the result with the beginning consonant of prefixes. From this result, it appears that plosives are common as both the word initial and final phoneme. However, this does not mean that the word beginning prefix and ending suffix consonants may not increase the saliency of the prefixes and suffixes. It is interesting to note that the prevalence of beginning plosive consonants in prefixes is much higher than ending plosive consonants in suffixes; one explanation for this is that the prefixes, as preceding cues, need higher salience than succeeding information, as discussed above.

There were only two ending vowels of suffixes but there was a larger variety of ending prefix vowels. This leads to the general finding that word initial or ending vowels are limited in number, but there is a wider variety of possible vowels in word internal positions. This is likely due to general phonotactic constraints but it still may be that the decreased amount of vowels that can occur at the end of suffixes may act to increase the saliency of these succeeding cues.

The results of this chapter support the prototype hypothesis that the similarity to naturally occurring preceding and succeeding cues (in this case, prefixes and suffixes) may increase the saliency of the artificial marker words used as distributional cues. The increase in learning with *cvc* preceding marker word cues when compared to *vc*/*vcc* preceding cues may be due to the greater occurrence of *cvc* prefixes and *vc* suffixes. Interestingly, Frigo and McDonald (1998), who integrated beginning and ending prefix and suffix cues into their language, had similarly *cvc* or *ccvc* prefix and suffix cues. However, categorical learning did not occur in their design unless both prefix and suffix cues were simultaneously present. One main difference between their design and the current Experiments is that the prefix and suffix cues were being used as "phonological markers" and there was an additional succeeding word distributional cue. Therefore, there were no phonological cues within the category words themselves, as are found in natural language and in the current Experiments.

Due to these differences in design it was decided to further test the hypothesis described above. However, due to the fact that Frigo and McDonald did not find learning with preceding *cvc*/*ccvc* prefix cues, it was decided to include additional phonological cues to the category words (i.e., only the coherent condition). Increased learning was expected when the marker words were phonologically similar to naturally occurring affixes with three Experiments that directly compared preceding and succeeding conditions. Experiment 5 investigated *cvc* marker words again, to provide a thoroughly valid comparison between preceding and succeeding cues. The *vc* marker words were tested in Experiment 6; succeeding cues were hypothesized to have more of an advantage than preceding cues due to the *vc* phonological structure being more prevalent with suffixes. Finally, Experiment 7 will investigate *cv* marker words; preceding cues were predicted to have higher salience than succeeding cues, which may lead to higher learning with preceding cues.

5. Suffixing and prefixing preference: Role of phonological similarity

Experiments 5, 6 and 7 directly compared preceding and succeeding distributional cues when there were also phonological consistencies within the category words. In other words, the coherent conditions of the *aAbB* and *AaBb* Experiments were directly compared. Experiment 5 used the same *cvc* marker words that were used in Experiments 3 and 4; this Experiment was necessary to have a fully valid *aAbB* and *AaBb* comparison. Experiment 6 used *vc* marker words and Experiment 7 used *cv* marker words. As stated in the previous Chapter Experiment 5 should lead to high levels of learning from preceding cues but an overall slight advantage for succeeding cues was predicted; Experiment 6 should improve suffix learning and Experiment 7 should improve prefix learning (possibly by equating learning from preceding and succeeding information).

5.1. Experiment 5

It was hypothesized that both language structure conditions would show evidence of categorical knowledge. The results of Experiment 3 and 4 indicated that the *AaBb* condition should produce more categorical knowledge because of the higher predictiveness of succeeding distributional cues. However, the affix analysis indicated that *cvc* marker words are more natural as prefix information, so the categorical knowledge in the *aAbB* condition was predicted to be substantial.

5.1.1. Participants

Twenty-four University of York undergraduate and masters students participated in this experiment. Participants had not previously been included in any similar experiments. The average age was 20.21 years; range was 18 to 37. There were 19 female and five male participants in the sample. All participants were native English speakers and were paid £3 or given course credit.

5.1.2. Materials

All stimuli for the preceding language structure condition (*aAbB*) were the same as the coherent condition of Experiment 3 while the stimuli for the succeeding language structure condition (*AaBb*) were the same as the coherent condition of Experiment 4. In other words, both phonological and distributional cues were available to aid learning; the position of the marker words varied based on the language structure condition. All other materials were the same as in Experiment 3 and 4, in that there were two frequency conditions within the category words. The high frequency words occurred twice as often as the low frequency.

As mentioned in the previous chapter 84.1% of three phoneme C_C affixes were prefixes. The marker words used in this experiment (*feg, nep, vot,* and *zid* or /fɛg/, /nɛp/, /vɔt/, or /zId/) were overall more phonologically similar to prefixes than suffixes. Out of the three phoneme C_C affixes, fricatives were more common as a beginning phoneme for prefixes than for suffixes. However, the one beginning nasal phoneme was more common as a beginning phoneme of suffixes. The ending plosive phonemes were relatively rare for both affix types. The majority of these prefixes ended with the /ɹ/ phoneme. However, it was decided not to alter the marker words in order to keep the results directly comparable to Chapter 3. In addition, as the ending plosive manner of articulation was neutral between the prefixes and suffixes it was not thought to detract from overall similarity to prefixes.

5.1.3. Procedure

The procedure was the same as in Experiments 3 and 4, in that there were two training and testing sessions and each training sentence was repeated aloud after auditory presentation.

5.1.4. Results

All multiple *t* tests were Bonferonni corrected. The tests that were significant before correction have been marked to give a more complete picture of the data; † indicates .05 level, †† indicates .01 level, and ††† indicates the .001 level.

The overall accuracy of differentiating the test items was 71.3%. The increase in accuracy from Experiments 3 and 4 was probably due to the lack of incoherent conditions, which had suppressed overall levels of categorical learning in previous Experiments. A mixed ANOVA was conducted with time (1st and 2nd test session) and frequency (high and low) as the within subjects factors and language structure (*aAbB* or preceding distributional cues and *AaBb* or succeeding distributional cues) the between subjects factor. As in Experiment 3 and 4, a dummy variable (test counterbalance) was included in the ANOVA as a between subjects variable. The dependent variable was the number of correctly accepted and rejected test items.

There were no significant main effects or interactions, but the results from the language structure main effect were striking in that the preceding distributional cue condition was more accurate at the similarity task, $F(1,20) = 2.69$, $p = .12$; M = 9.06 for the *aAbB* language and M = 8.04 for the *AaBb* language. The remaining main effects or interactions were not significant, frequency: $F(1,20) = 1.90$, $p = .18$; frequency by test by condition: $F(1,20) = 1.12$, $p = .30$; all others: $F < 1$.

In addition to the comparison between the two language structure conditions, both conditions were tested against a chance level of responding. With a chance level of 24, responding to the similarity judgement task was well above chance for both preceding and succeeding distributional cues, $t(11) = 5.89, p < .005^{\dagger\dagger\dagger}$, M = 36.25 for the *aAbB* condition and $t(11) = 6.39, p < .005^{\dagger\dagger\dagger}$, M = 32.17 for the *AaBb* condition. This result was not surprising due to the combination of both phonological and distributional cues, which has been shown in both this Experiment and Experiments 1-4 to induce substantial categorical knowledge.

The card sorting results indicated that there was no difference between preceding and succeeding cues in influencing how well the participants were able to explicitly differentiate the category words; $t(22) = 1.48, p = .15$, M = 4.67 for the *aAbB* condition and M = 4.0 for the *AaBb* condition. The *aAbB* condition was marginally significantly above the chance level of 3.91; $t(11) = 2.27, p = .09^{\dagger}$ for the *aAbB* language structure condition whereas the *AaBb* condition did not differ from chance level, $t(11) = .29, p = 1.0$.

As in Experiments 1-4 the card sorting results were correlated with the overall results of the similarity judgements. With both groups combined there was a significant positive relationship between the ability to explicitly group the category words together and the ability to complete the more implicit similarity judgements, $r = .62, p = .001$. The relationship was significant with the participants who received preceding distributional cues, $r = .69, p < .05$ and was not significant with those who received succeeding distributional cues, $r = .41, p = .18$. See Figure 5.1 for a scatterplot of the similarity scores and the card sorting results.

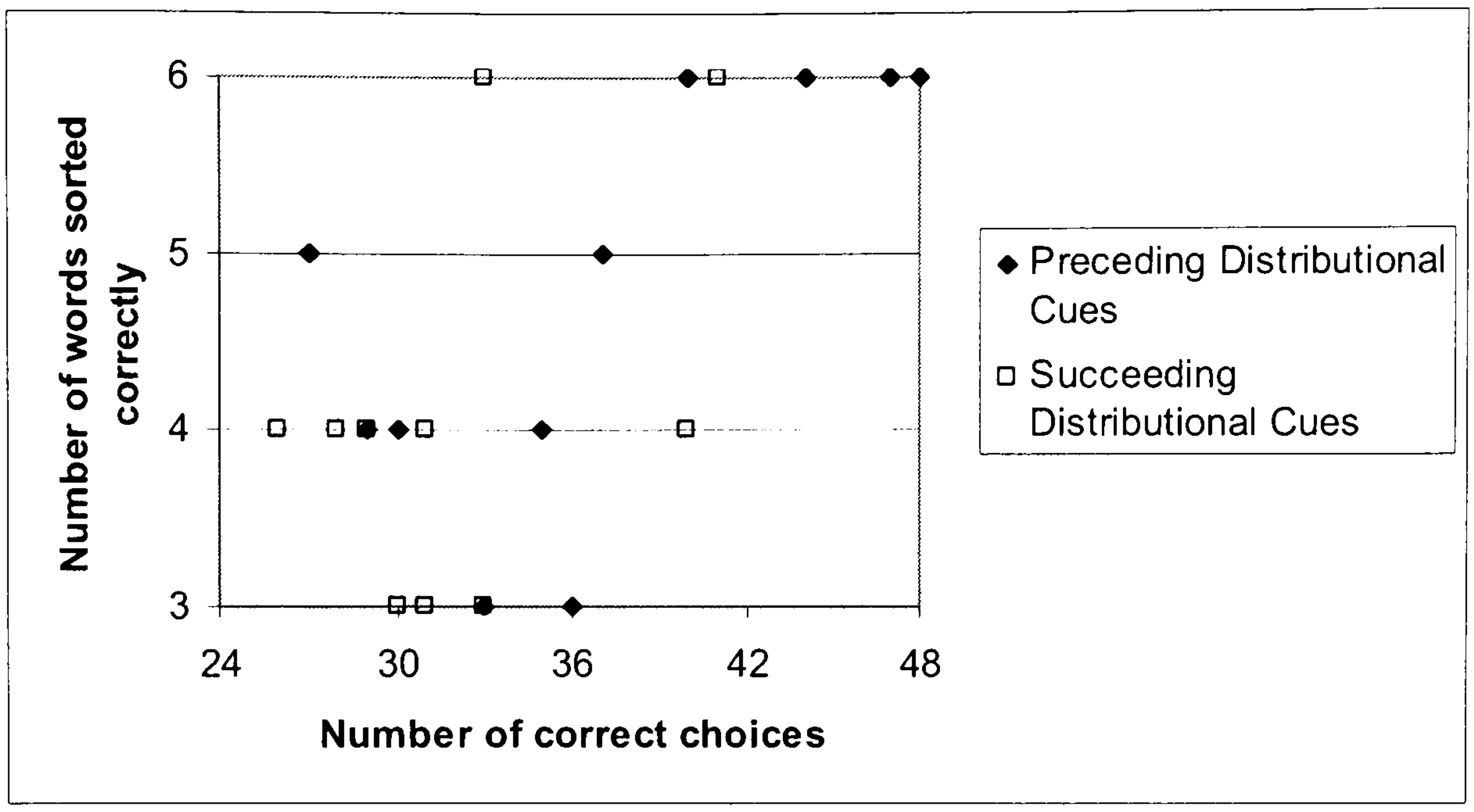

Figure 5.1 Scatterplot of the number of correctly sorted words and number of correct responses to the test items in Experiment 5.

5.1.5. Discussion

The results from the similarity task were not consistent with the combined results of Experiments 3 and 4, although they were consistent with the predictions from the prototype hypothesis. If marker word similarity to prefixes and suffixes increased the salience, and thus learnability, of the marker words increased learning would be expected in the *aAbB* preceding cue condition rather than in the *AaBb* succeeding cue condition as the *cvc* marker words were more phonologically similar to prefixes than to suffixes. There was not a significant difference in favour of the preceding cue condition, but a trend in that direction was found, which considering past results was unexpected. The prevalence of the succeeding cues in inducing more categorical knowledge has been found throughout three pairs of similar experiments (St. Clair & Monaghan, 2005; and Experiments 1-4).

The slight advantage for preceding distributional cues indicated that the salience of the preceding distributional cues may have increased due to similarity with naturally occurring preceding distributional cues; in other words, the participants were able to apply their expectations from their native language to the artificial language in the Experiment. The increased salience of preceding *cvc* marker words would have increased the learning rate, allowing the associations between the preceding cue and the category words to reach the asymptote level faster. The saliency of the succeeding *cvc* marker words should not have increased substantially. As discussed in Chapter 4, *cvc* suffixes were uncommon which might lead to a decrease in succeeding *cvc* marker word salience if any change occurred. The increased learning rate for preceding cues would have allowed the preceding cue participants to learn a substantial amount about the categories in a short amount of time. According to the results, this led to equivalent amounts of learning in both the preceding and succeeding marker word positions.

The lack of replication from the combined results of Experiment 3 and 4 may be due to several reasons. Both of these Experiments used the same *cvc* marker words as in the current Experiment. Experiment 3 contained a preceding cue language and had both coherent and incoherent conditions. Experiment 4 also contained both coherency conditions but was a succeeding distributional cue language. When these experiments were directly compared the succeeding cue experiment showed more categorical knowledge than the preceding cue experiment. Furthermore, this effect was carried by the coherent condition making the combined results and the current results comparable. A direct comparison indicated that the similarity judgement results in the *aAbB* preceding cue condition increased an average of eight correct sentences from Experiment 3 to Experiment 5; M = 28.08 in the coherent condition of Experiment 3 and M = 36.25 in the *aAbB* condition of Experiment 5. This difference between these specific conditions was significant when investigated within a mixed ANOVA with the same variables as described above, $F(1,22) = 13.28, p < .001$. There was no great difference between the coherent condition of Experiment 4 and the *AaBb* condition of Experiment 5; M = 32.92 and M = 32.17, respectively. From this comparison it was clear that the difference in results was principally due to the difference in the *aAbB* Experiment 3 and the corresponding condition in Experiment 5.

To investigate this further, the card sorting tasks of the two conditions were consulted. If the two groups of participants were comparable, the two card sorting results should also be comparable. However, the coherent condition of Experiment 3 was the only coherent condition from all Experiments to have below chance card sorting results; although these participants were able to perform above chance with the similarity judgements they were somehow impaired with the card sorting results (M = 3.75). However, in the *aAbB* condition of Experiment 5, participants were able to sort the category cards at a normal rate in comparison to previous experiments (M = 4.33). The exact reason for this difference is not apparent, but the overall picture suggests that the current Experiment is a more representative picture of learning from *cvc* preceding distributional cues.

Experiment 6 further investigated how the phonological form of marker words can influence grammatical category learning from preceding and succeeding cues. The marker words used in this experiment were *vc* and it was hypothesized that the succeeding cue condition would benefit as this phonological form occurs more frequently in suffixes.

5.2. Experiment 6

5.2.1. Participants

Twenty-six University of York undergraduate and PhD students participated in this experiment. One participant was excluded as they produced unusable card sorting results (unequal groups)[6]. A second participant was excluded due to a technical failure. Participants had not previously been included in any similar experiments. The average age of the included sample was 19.54 years; range was 18 to 33. There were 18 female and six male participants in the included sample. All participants were native English speakers and were paid £3 or given course credit.

5.2.2. Materials

All stimuli and materials were identical to Experiment 5, except in regards to the marker words. The phonological form of the marker words in Experiment 6 was *vc*. There were four possible *vc* marker words (*og, im, ev,* and *ud,* or /ɔg/, /Im/, /ɛv/, and /ʌd/) and two of the four possible marker words were randomly chosen for each participant. This was done to ensure that no individual marker words unduly influenced the results.

The overall phonological form of the marker words was more similar to suffixes than prefixes. The majority of suffixes began with a vowel. Additionally, suffixes were more likely than prefixes to end with a consonant. When looking at the ending consonant, the ending plosives and fricative in three of the marker words was more typical of suffixes as well. However, the ending nasal consonant was more typical of

---

[6] All results were conducted with this participant excluded and included. There were no differences between the two sets of results.

prefixes. All beginning vowels were unrounded, which follows the pattern found in beginning affix vowels. However, the vowels that were used did not match the height and position of articulation pattern found in prefixes or suffixes.

5.2.3. Procedure

The procedure was the same as in Experiments 5, in that there were two training and testing sessions and each training sentence was repeated aloud after auditory presentation.

5.2.4. Results

All multiple *t* tests and pairwise comparisons were Bonferonni corrected, with significance before correction indicated as for Experiment 5.

The overall accuracy of this Experiment was 70%, which was comparable to Experiment 5. A mixed ANOVA was conducted with time (1^{st} and 2^{nd} test session) and frequency (high and low) as the within subjects factors and language structure (*aAbB* and *AaBb*) as the between subjects factor. As in Experiment 3 and 4, a dummy variable (test counterbalance) was included in the ANOVA as a between subjects factor. The dependent measure was the number of correctly accepted or rejected test sentences.

The main effect of time was significant, with participants showing an increased ability to differentiate the test items during the second test session, $F(1,20) = 7.82, p < .05, \eta_p^2 = .28$; M = 7.65 of 12 for the first test session and M = 8.96 of 12 for the second test session. Additionally, the time by language structure interaction was significant, $F(1,20) = 10.49, p < .005, \eta_p^2 = .34$; see Figure 5.2. There were no other significant main effects and interactions; language structure: $F(1,20) = 2.29, p = .15$; frequency: $F(1,20) = 1.07, p = .31$; frequency by language structure: $F(1,20) = 1.83, p = .19$; all others: $F < 1$.

Pairwise comparisons of the time by language structure interaction revealed that there were no differences between the two test sessions with the *aAbB* language structure ($p = .76$) but that participants showed more categorical knowledge in the second test session of the *AaBb* language structure ($p < .005^{\dagger\dagger\dagger}$). In addition, there was no difference in accuracy between the language conditions during the first test session ($p = .34$), but the *AaBb* language structure condition was significantly more accurate than the *aAbB* condition during the second test session ($p < .01^{\dagger\dagger}$). Therefore, this interaction was due to the increase in accuracy over the course of the experiment with succeeding distributional cues, but not with preceding cues.

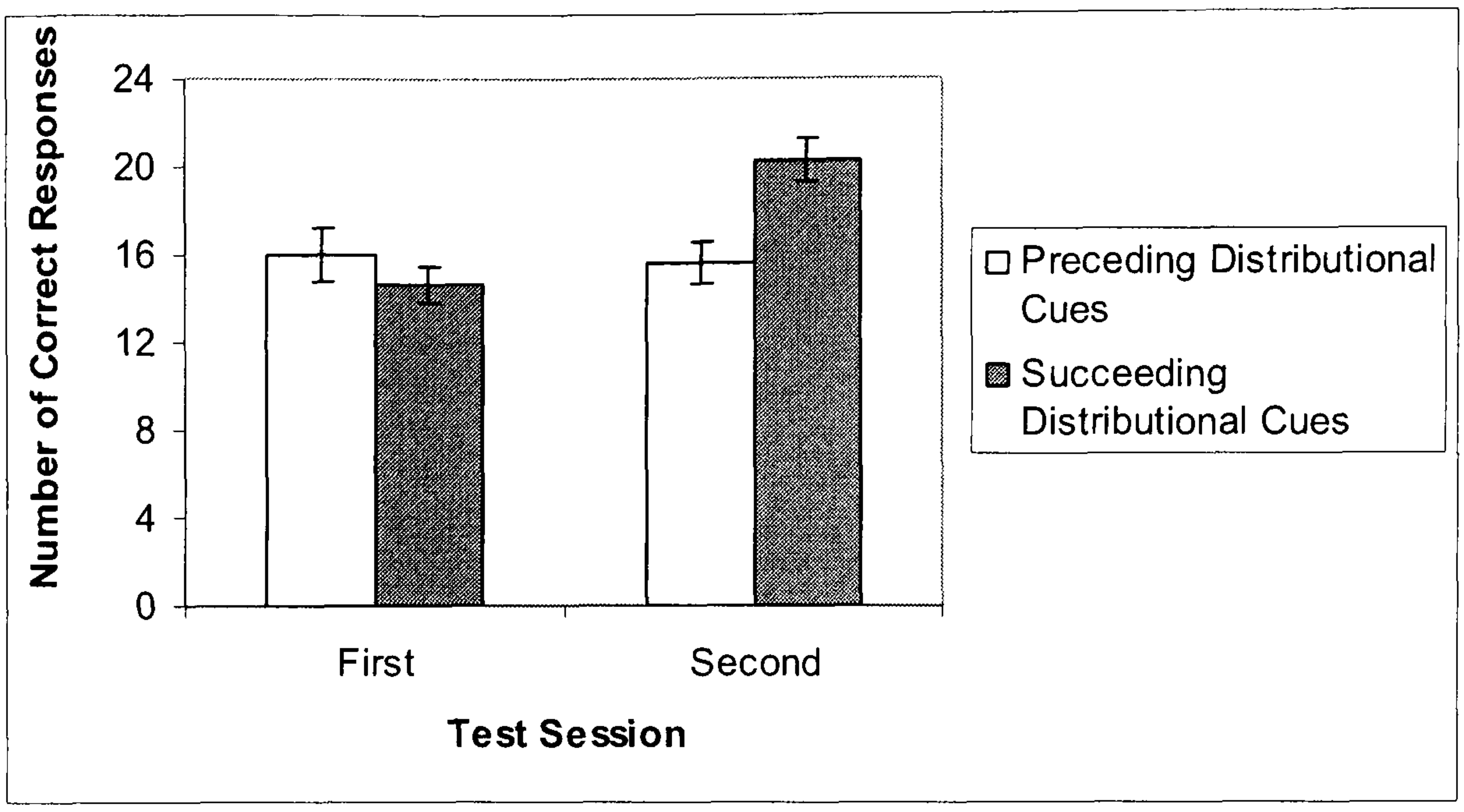

Figure 5.2 Number of correctly rejected and accepted test sentences (with standard error bars) for the first and second test session in the preceding and succeeding language structure conditions in Experiment 6.

Testing against the chance level of responding (24) revealed that both preceding and succeeding distributional cues led to above chance levels of learning, $t(11) = 5.64$, $p < .005^{\dagger\dagger\dagger}$, M = 31.58 for the *aAbB* condition and $t(11) = 6.98$, $p < .005^{\dagger\dagger\dagger}$, M = 34.83 for the *AaBb* condition. Although the difference between the two conditions was not found to be significant ($p = .15$) it is interesting to note that the pattern of the means has reversed from the pattern found in Experiment 5; more learning was found in the *aAbB* language structure condition with *cvc* marker words but, with *vc* marker words, more learning was found in the *AaBb* language structure condition.

The results of the card sorting test did not show any significant difference between the two language structure conditions, $t(22) = -.39$, $p = .70$; M = 4.42 for the *aAbB* language structure condition and M = 4.58 for the *AaBb* language structure condition. When tested against the chance level of 3.91 neither the *aAbB* language structure group nor the *AaBb* language structure was above chance level of 3.91, $t(11) = 1.95$, $p = .15$ and $t(11) = 2.00$, $p = .14$, respectively.

The results of the card sorting task were correlated with the overall score of correctly accepted or rejected test items in the similarity task. A positive significant relationship was found, $r = .53$, $p < .001$, in which the higher the participants scored on the similarity task the more likely they were to sort the correct category words together. When broken down by language condition there was a significant relationship in the *aAbB* condition, r = .67, $p < .05$, but not in the *AaBb* condition, r = .44, $p = .15$. See Figure 5.3 for a scatterplot of the two dependent measures.

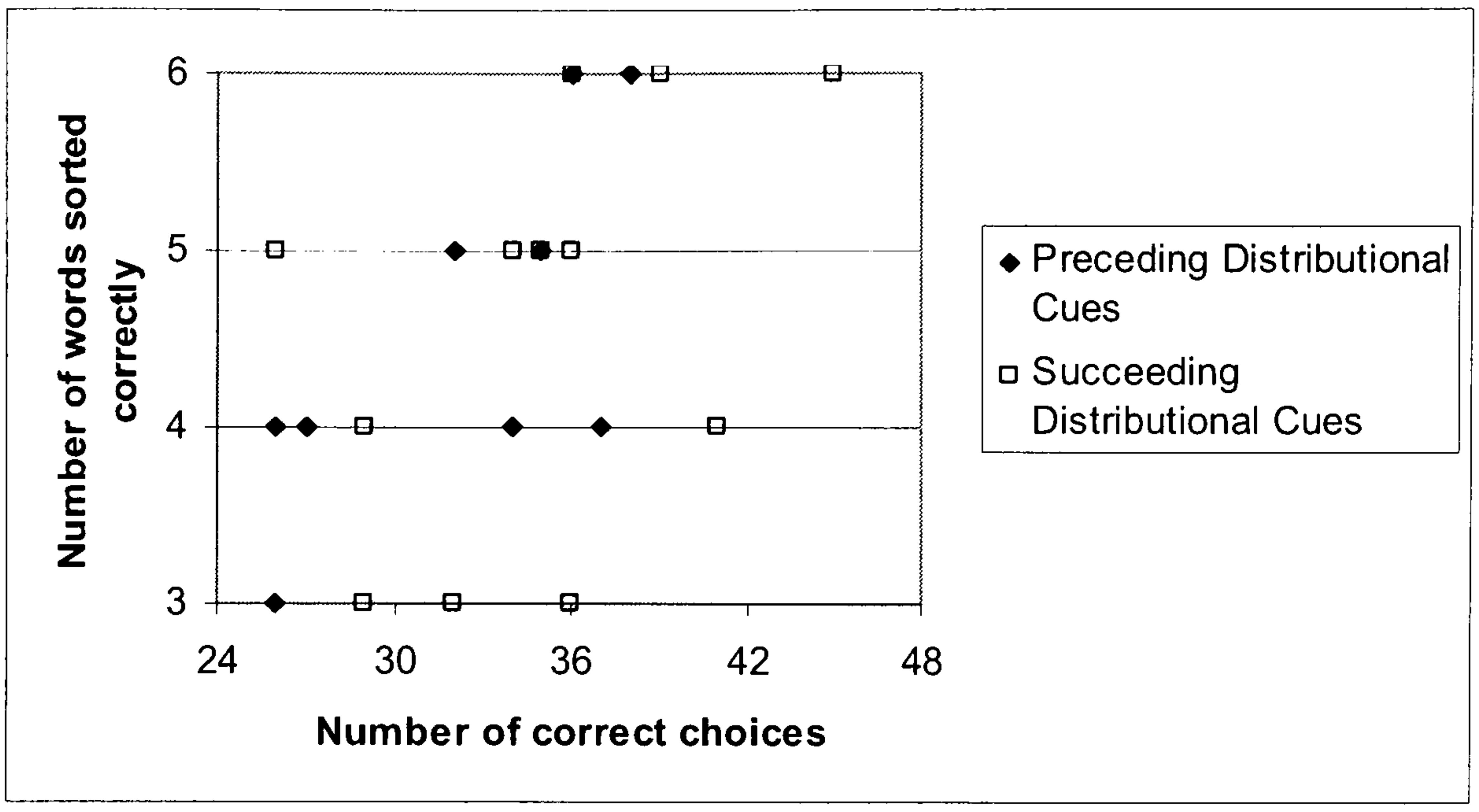

Figure 5.3 Scatterplot of the number of correctly sorted words and number of correct responses to the test items in Experiment 6.

5.2.5. Discussion

The pattern of results of the preceding and succeeding cue conditions was in the hypothetical direction; as predicted there was more evidence of categorical knowledge (in the similarity judgements) with succeeding cues when the marker words were phonologically more common in suffixes. However, the difference in the similarity task was non-significant, which does not allow any firm conclusions to be made. One possibility touched upon earlier relates to the fact that the *vc* phonological form is common with both prefixes and suffixes; thus, this form may have increased salience in both the preceding and succeeding locations.

Interestingly, a succeeding cue advantage emerged in the second test session. The participants with the succeeding cues were able to learn more about the categories over the course of the experiment, leading to higher categorical knowledge during the second test session. This can be explained by the R-W model; increased salience in the succeeding cue would lead to an overall increase in the asymptote level, whereas increased salience in the preceding cue would lead to an increase in the learning rate, which would allow the associative strength to reach the asymptote level faster. If both preceding and succeeding cues achieved increased saliency with *vc* marker words, the preceding cue language should reach asymptote (and highest possible learning) relatively quickly. However, the succeeding cue language should take longer to reach the maximum associative strength as the asymptote has increased due to the increased saliency of the *vc* succeeding marker words. The difference in the test sessions was reflecting the fact that the associative strength between the marker and category words took longer to reach asymptote with the succeeding cues; the increase in categorical knowledge was reflecting the noticeably higher associative strength that occurred within the second training session. There was no corresponding increase with the preceding distributional cues as the associative strength increase within the second training session was not noticeable as it was already approaching asymptote. See Figure 5.4 for a graphical demonstration of the theoretical changes in associative strength with both preceding and succeeding cue saliency increases.

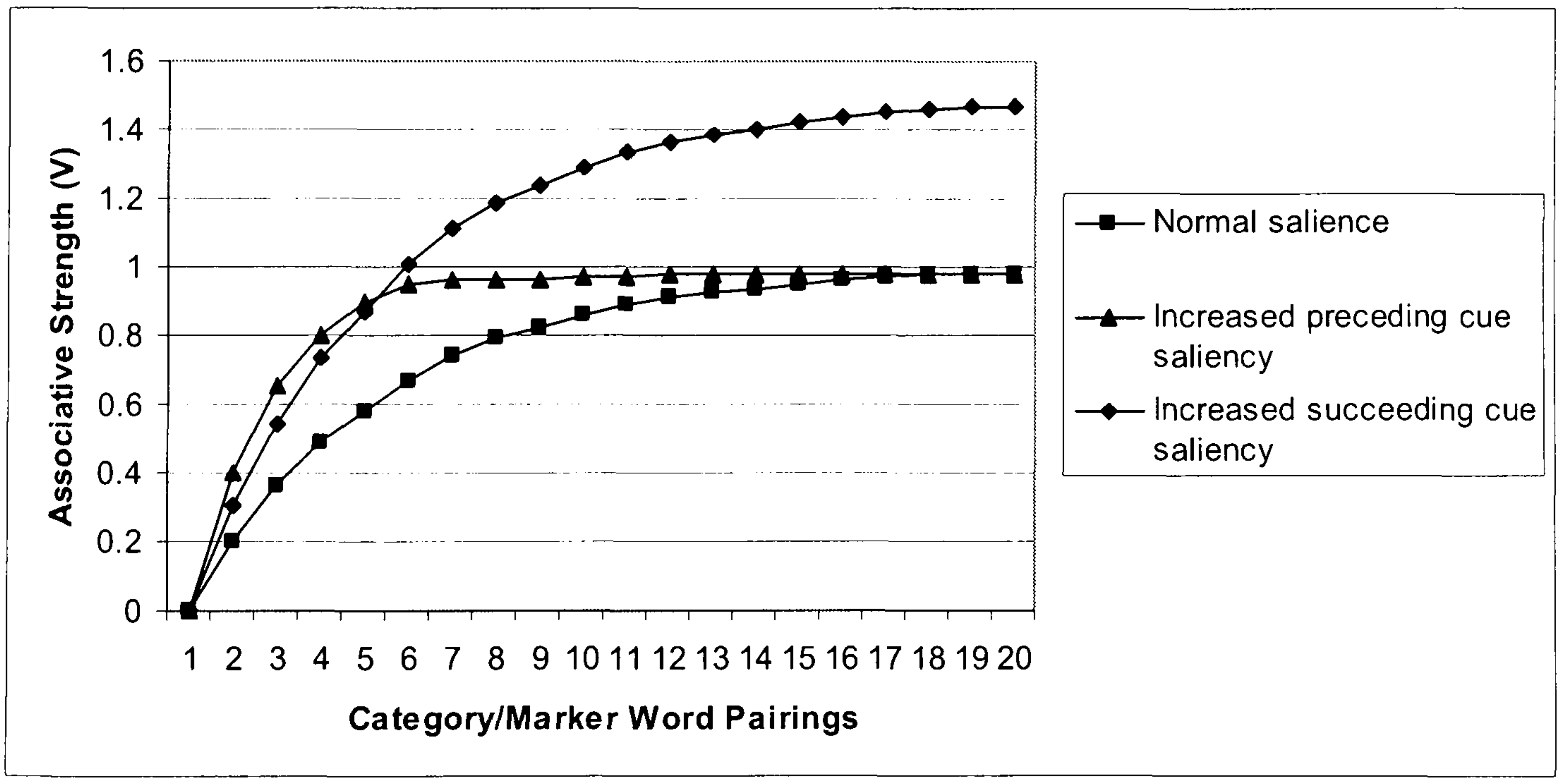

Figure 5.4 Hypothetical graph depicting the change in the associative strength with increased salience of preceding cues and of succeeding cues.

Compared with Experiment 5 it appeared that the *aAbB* condition may have been less able to learn categorical information from the *vc* marker words, M = 36.25 for *cvc* and M = 31.58 for *vc*. Consistent with the hypothesis, the *AaBb* condition seemed to learn slightly more from *vc* marker words than from *cvc* marker words, M = 32.17 for *cvc* and M = 34.83 for *vc*. These differences indicate that *cvc* preceding marker words may have been more salient than the *vc* marker words, even though *vc* prefixes are more common than *cvc* prefixes. As expected, the *vc* succeeding marker words appear to be slightly more salient than the *cvc* succeeding marker words, although this difference is not as large as with the preceding cues. This may be because the general higher learning from succeeding cues reduces the additional impact of higher salience marker words. These differences were fully investigated in the General Discussion.

The next experiment investigated the role of *cv* marker words. The overall prediction was that *cv* marker words would induce more learning in the preceding cue condition, as they were phonologically similar to prefixes but not suffixes.

5.3. Experiment 7

5.3.1. Participants

Twenty-six University of York undergraduate students participated in this Experiment. Two participants were excluded as they produced unusable card sorting results (unequal groups)[7]. Participants had not previously been included in any similar experiments. The average age of the included sample was 19.04 years; range was 18 to 22. There were 19 female and five male participants in the included sample. All participants were native English speakers and were paid £3 or given course credit.

5.3.2. Materials

All stimuli and materials were identical to Experiment 5 and 6, except in regards to the marker words which were of the phonological form *cv*. The four possible marker words contained the same phonemes as used in Experiment 6, but in reverse order. That is the marker words *og, im, ev,* and *ud* for Experiment 6 became *go, mi, ve,* and *du* (/gɔ/, /mI/, /vɛ/, and /dʌ/).

The beginning consonants in these marker words were consistent with prefixes, except the /m/ phoneme, which more commonly occurred as a beginning consonant for suffixes. Although beginning and ending consonants were more common in prefixes, affixes that ended with a vowel were more likely to be prefixes than suffixes. Additionally, the range of vowel height and position of articulation was more consistent

---

[7] All results were conducted with these participants excluded and included. There were no differences between the two sets of results.

with prefixes than suffixes, although the roundedness of the vowels was inconsistent. This was mainly due to the fact that the vowels used in Experiments 6 and 7 were kept constant. The marker words in this Experiment were not quite as consistent as in Experiments 5 and 6, but it was still thought the overall phonological similarity to prefixes would increase preceding marker word saliency.

5.3.3. Procedure

The procedure was the same as in Experiments 3 and 4, in that there were two training and testing sessions and each training sentence was repeated aloud after auditory presentation.

5.3.4. Results

All multiple *t* tests were Bonferonni corrected, with significance before correction indicated as for Experiment 5.

The overall accuracy in the similarity task was 74%, which was slightly higher than in Experiments 5 and 6. A mixed ANOVA was conducted with time (1st and 2nd test session) and frequency (high and low) as the within subjects factors and language structure (*aAbB* and *AaBb*) as the between subjects factor. As in Experiment 3 and 4, a dummy variable (test counterbalance) was included in the ANOVA as a between subjects factor. The dependent variable was the number of correctly accepted or rejected test sentences.

Language structure was the only significant main effect, with the succeeding cue condition performing more accurately than the preceding cue condition, $F(1,20) = 4.92, p < .05, \eta_p^2 = .20$; M = 9.60 and M = 8.08 of a possible 12, respectively. There were no other significant main effects or interactions; test session by condition: $F(1,20) = 2.04, p = .17$; frequency by test session by language structure: $F(1,20) = 1.28, p = .27$; all others $F < 1$.

As with the previous Experiments the two language structure conditions were compared to chance level. Both the preceding and succeeding condition were significantly above chance level of 24, $t(11) = 4.96, p < .005^{\dagger\dagger\dagger}$, M = 32.33 and $t(11) = 6.92, p < .005^{\dagger\dagger\dagger}$, M = 38.42, respectively.

In regards to the card sorting results, there was no significant difference between the preceding and succeeding language structure conditions, $t(22) = -.48, p = .63$. The preceding group did not differ from chance level of 3.91, $t(11) = 1.56, p = .30$, M = 4.5, whereas the succeeding language structure condition was marginally significantly above chance level of responding, $t(11) = 2.39, p = .07^{\dagger}$, M = 4.75.

The two dependent measures, similarity judgements and card sorting results, were correlated and it was found that the overall correlation was highly significant, $r = .56, p = .005$. For the preceding language structure condition, the correlation was marginally significant, $r = .57, p = .053$. However, the succeeding language structure condition correlation was significant, $r = .58, p < .05$. See Figure 5.5 for a scatterplot of the two dependent measures.

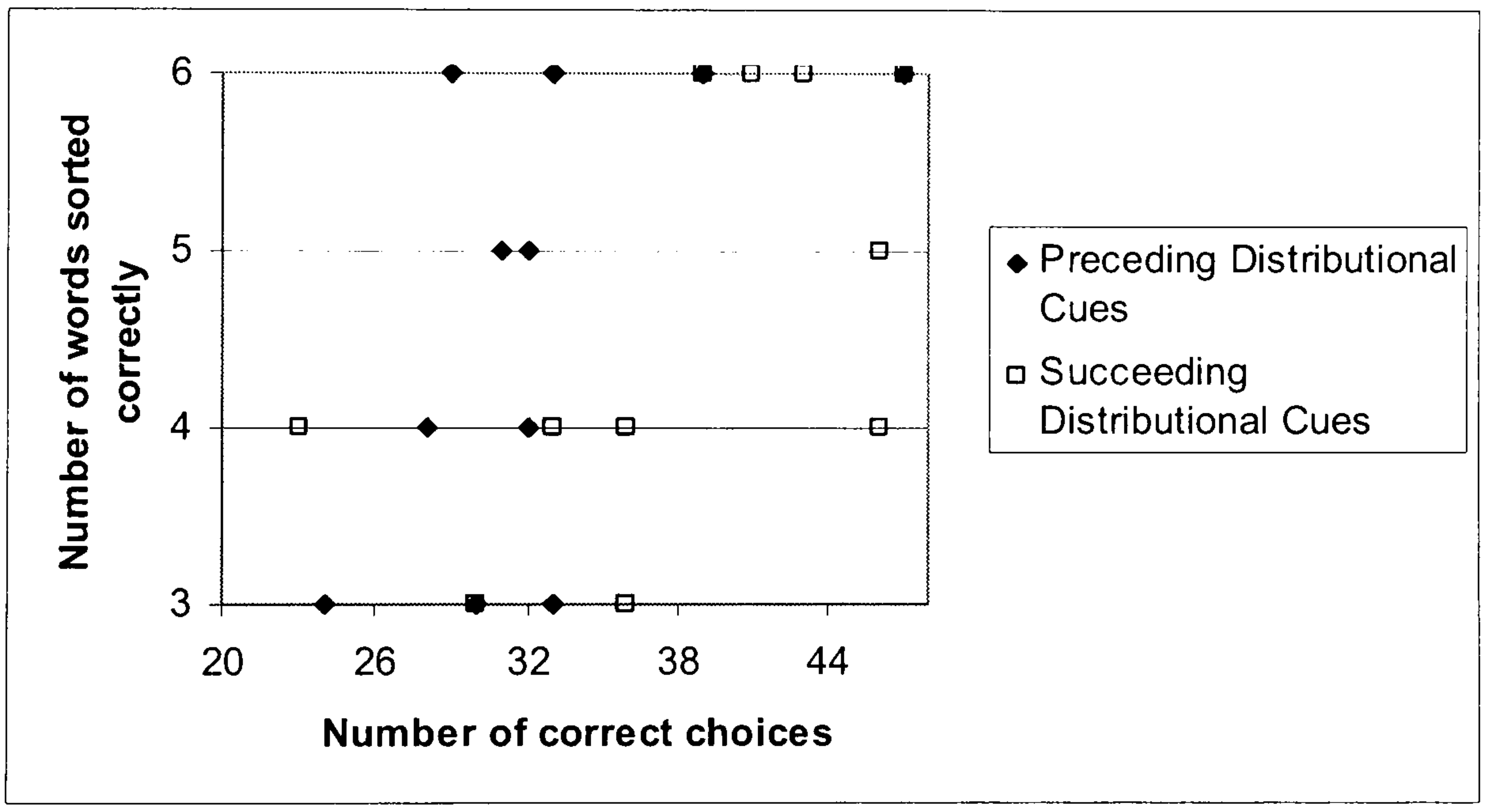

Figure 5.5 Scatterplot of the number of correctly sorted words and number of correct responses to the test items in Experiment 7.

5.3.5. Discussion

The results were not expected, as the succeeding cue participants learned significantly more about the categories than the preceding cue condition. The *cv* form of the marker words was predicted to increase the salience of preceding, but not succeeding cues which would lead to increased learning with preceding cues. This was predicted as the *cv* phonological form was more common in prefixes than suffixes; additionally, affixes with ending vowels were likely to be prefixes. However, it should be pointed out that *cv* affixes were relatively uncommon amongst both prefixes and suffixes. Therefore, the underlying premise of the *cv* marker word similarity to prefixes may not hold as the *cv* phonological form was not a general characteristic of prefixes. The advantage for succeeding marker word cues is not surprising when this is taken into account. As the preceding cue salience was not increased due to similarity to the general phonological characteristic of preceding natural language cues, the succeeding cue advantage emerged once again.

The card sorting results mirrored the findings in the similarity task in that the succeeding condition performed at above chance levels, while the preceding cue condition, although numerically higher than chance, did not differ from chance level. The significant (or marginally significant) correlations between the two dependent measures indicated that both the implicit and explicit measures of categorical knowledge used in these Experiments were in fact testing categorical knowledge.

When compared to the results from Experiment 5 and 6, the preceding cue condition was at about the same level as in the *vc* Experiment but was lower than in the *cvc* Experiment, M = 32.33 for *cv*, M = 31.58 for *vc*, M = 36.25 for *cvc* marker words. The succeeding cue condition had the highest learning found in any of the three Experiments, M = 38.42 for *cv*, M = 34.83 for *vc*, M = 32.17 for *cvc* marker words. The pattern of means was somewhat unexpected, as learning from succeeding cues was hypothesized to be highest with *vc* marker words; all three Experiments were compared with an ANOVA that is discussed in the General Discussion.

5.4. General Discussion

Overall, these Experiments aimed to provide evidence that phonological similarity to prefixes and suffixes can influence the saliency of marker words, which in turn influences their effectiveness at aiding grammatical categorization. However, as briefly mentioned in the preceding discussion section the pattern of means across the three Experiments was some what contradictory to the pattern expected by the prototype hypothesis.

The results from all three Experiments were investigated in a 3 by 2 ANOVA, with marker word phonology (*cvc, vc,* and *cv*) and language structure (*aAbB* and *AaBb*) as the between subjects factors. Neither of the main effects were significant but the interaction between them was significant, $F(2,66) = 4.76, p < .05$. The interaction was analysed with pairwise comparisons. It was found that there was an almost significant difference between the preceding and succeeding conditions in the *cvc* experiment (p =

.094), a lack of significant difference in the *vc* experiment ($p = .181$) and the significant difference in the *cv* experiment ($p < .05$), which replicated the individual findings. When comparing the results across the Experiments only the difference that was significant was between the succeeding cue conditions in the *cv* and *cvc* Experiments ($p < .05^{\dagger}$).

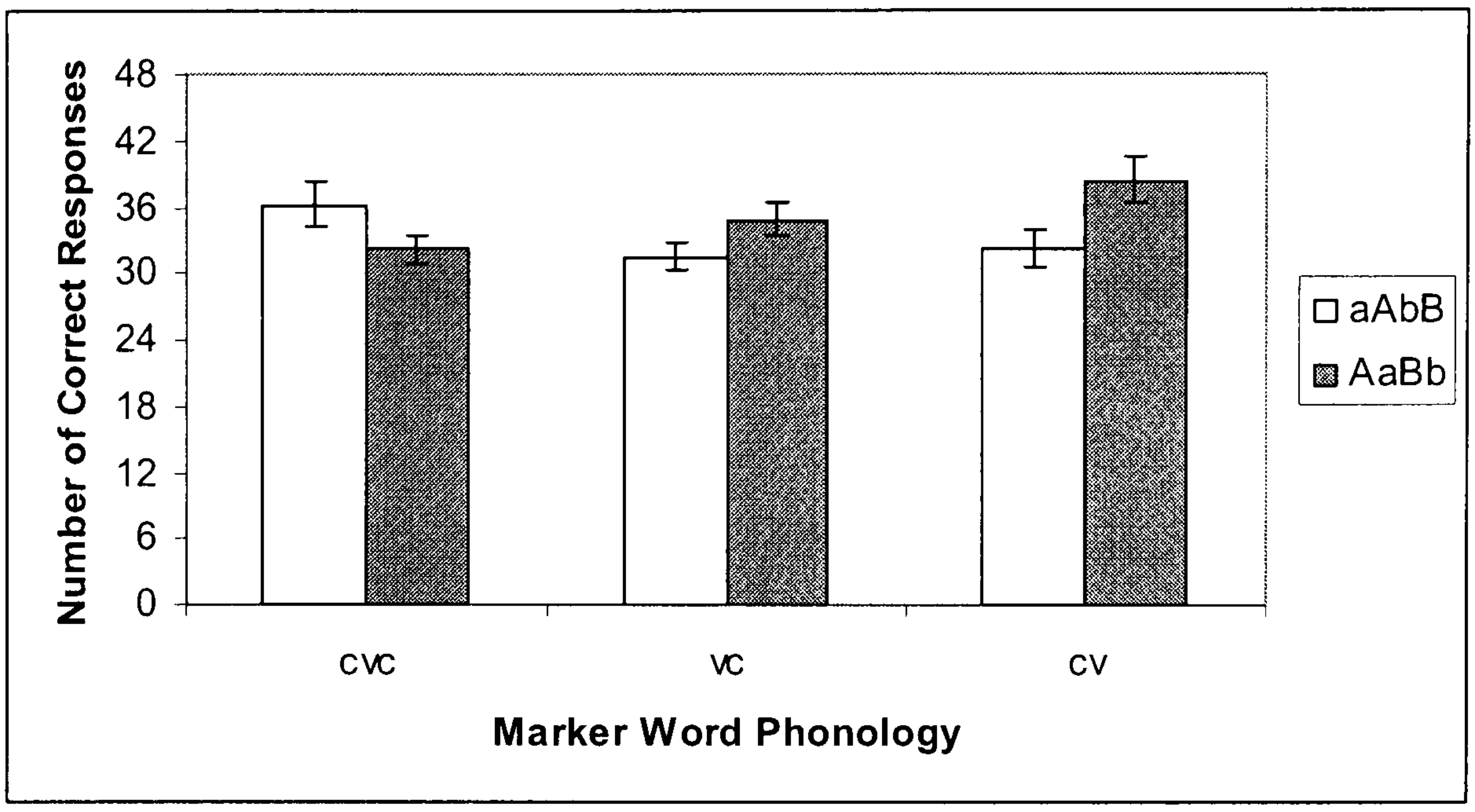

Figure 5.6 Number of correct responses in the preceding and succeeding cue conditions in Experiment 5 (*cvc*), Experiment 6 (*vc*) and Experiment 7 (*cv*).

There were no significant differences with preceding cues, which was slightly surprising as the interpretation of the individual analyses indicated that the *cv* condition may show lower levels of categorization than the *cvc* and *vc* Experiments. The salience of the *cv* preceding marker words was thought to be lower than the salience of the *cvc* and *vc* marker words; however, this result indicated that perhaps it was not reduced salience of the *cv* preceding marker word that caused the significant difference between the preceding and succeeding conditions in the *cv* Experiment, but an increase in the learning from succeeding cues with this particular marker word phonology. Alternatively, it could be that the learning found within the *cv* and *vc* marker words was not influenced by any saliency differences, and instead the *cvc* marker words were particular useful as preceding cues for a different reason. The difference between the *cvc* Experiment and the *vc* and *cv* Experiments did approach significance with the Bonferroni uncorrected *p* values ($p = .056$ and .11, respectively).

When looking at the succeeding cues, a significant difference was found between the *cvc* and *cv* marker words ($p < .05$), with the *cv* marker words allowing more categorical knowledge. There was no hint of a difference between the *vc* and *cvc* marker words (p = .814) nor between *vc* and *cv* marker words (p = .42). These results are intriguing and could lead to two possible conclusions; the first being that *cvc* marker words somehow suppressed learning from succeeding distributional cues or that *cv* marker words increased the induction of categorical knowledge based on succeeding cues. The lack of significant differences between the *cv* and *vc* succeeding marker word Experiments weakens the argument for the latter possibility. This leaves the idea that *cvc* marker words led to lower rates of categorical learning than the marker words of other phonological forms. The rates of learning categorical information from succeeding *cvc* maker words was approximately the same in Experiment 5 as in the coherent condition of Experiment 4, M = 32.17 in Experiment 5 and M = 32.92 in Experiment 4, which indicated the levels of learning from *cvc* succeeding words were constant across all Experiments.

This possibility is intriguing and somewhat puzzling. One explanation may be to look beyond the affix level we have focused on thus far and look at the word level. When looking at English many of the important cues for grammatical category are highly frequent function words. For example, *a, an, the, to, you, I, we, it* are very important in learning which grammatical category particular words belong to. Many of these words occur in the preceding location to the target words; it may be that word cues are the most prevalent preceding cues, while suffixes are the most consistent succeeding cues to grammatical categories. Previous research supports the idea that preceding word cues are highly reliable grammatical category cues (Redington, Chater, & Finch, 1998). Prefixes themselves, as mentioned earlier, have evolved into a more word like form than suffixes, as evidenced by Bauer (2003a). Furthermore, there is evidence that grammatical gender is often learnt through a combination of suffixing cues and phonological consistencies (Mirković, MacDonald, & Seidenberg, 2005), supporting the idea that reliable succeeding cues are often in the form of bound morphemes.

When looking at these function words it was found that 38.2% of the highest frequency preceding function word cues in child directed speech[8] were *cvc* words, while only 20.6% and 23.5% are *vc* and *cv* words. The distribution is almost identical with the highest frequency succeeding function word cues; 42.1% *cvc,* 23.7% *vc*, and 21.1% *cv.*

---

[8] These high frequency function words were taken from the corpus analysis reported in Chapter 6; the CHILDES corpora that were used were detailed in this chapter.

This finding fits with the observation that the *cvc* marker words in Experiment 5 were by far more "wordlike" than the *cv* and *vc* counterparts. Therefore, it could be similarity to words, or "wordlikeness" that allowed the *cvc* marker words to have increased learning when put in the preceding location. Conversely, when the *cvc* marker words were put in a succeeding location they allowed significantly less learning than the *cv* marker words. This may be because the *cv* (and *vc*) marker words did not contain many "wordlike" properties – they were basically syllables and have more properties of bound morphemes than of independent words. Thus, the difference between the *cvc* and *cv* marker words may have been due to increased learning with the *cv* marker words as they were simply more similar to naturally occurring succeeding cues.

One final consideration is the vowels used in the *vc* and *cv* marker words. The vowels were acceptable in the *vc* form, but the same vowels (lax vowels) in an ending location, as in the *cv* marker words, were dissimilar to the ending vowels found within the affix analysis. More generally, lax vowels are not considered well formed when in the final phoneme position (Norris, McQueen, Cutler, Butterfield, & Kearns, 2001). Thus, the *cv* marker words would be thought less "wordlike" than the *cvc* and even *vc* counterparts. However, it could be that the incongruent position of the ending vowels may have increased the salience of these marker words, which could have led to an increase in the differential learning between the preceding and succeeding cues more than the *cvc* or *vc* marker words. It should be remembered that a salience increase in the marker word in the *aAbB* language should only increase the learning rate, whereas an increase in marker word salience in the *AaBb* language increases the overall asymptote level. Thus, the normal amount of learning in the *aAbB cv* marker word was consistent with this hypothesis, as was the increased learning with the *AaBb cv* marker words.

Chapter 6 directly analysed how well preceding and succeeding words group words of the same grammatical category together. In addition, the combination of preceding and succeeding words will be investigated to attempt to determine whether the combination is better than either individual cue, both in terms of the accuracy and the completeness of the resultant categories. These results will shed light on whether the preceding or succeeding words are more accurate at grammatical categorization, which will aid in interpreting the current results.

6. Usefulness of preceding and succeeding word cues in categorizing grammatical categories

This chapter investigated the usefulness of distributional cues within child directed speech. Six corpora from the CHILDES database (MacWhinney, 2000) were used to investigate how well preceding and succeeding word cues group words of the same grammatical category together. This chapter also investigated the claim by Mintz (2003) that the distributional context of nonadjacent frequent frame words (e.g. you ___ it) are used during language acquisition to aid in the creation of grammatical categories. Mintz (2003) stated that distributional frames may provide the most useful distributional context for grouping words of the same grammatical category together, and thus aid the formation of grammatical categories (Mintz, 2002, 2003, 2006). This argument was based on two main claims; the first was the idea that form classes cannot be learned from bigram distributional cues without recourse to additional correlated cues (e.g., Brooks, Braine, Catalano, Brody, & Sudhalter, 1993; Christiansen & Monaghan, 2006; Frigo & McDonald, 1998; Monaghan, Chater, & Christiansen, 2005; St. Clair & Monaghan, 2005) and the second was that non-adjacent frames (i.e., a_b) were a conceptually relevant linguistic unit that adults and infants have shown to be sensitive to (e.g. Gómez, 2002).

The first claim was partially supported by evidence from the Experiments in Chapters 2 and 3; there was no learning from the bigram distributional cues in any of the previous Experiments without additional correlated phonological cues to category membership *unless* the frequency of the bigram cue and category word was sufficiently high. From this research, it is true that consistent categorization of all category members only occurred with bigram distributional and phonological information combined. Additionally, natural language learning does not rely on one cue alone but rather a mixture of distributional, phonological, prosodic, and semantic cues. The previous Experiments have only combined bigram distributional and phonological cues and have found these combined cues were sufficient to induce categorization of word classes. In fact, any claim that grammatical categories are acquired by only one cue (e.g., by distributional "frequent frames") is inconsistent with what is known about grammatical categories and infants' ability to learn (Christiansen & Monaghan, 2006; Gerken, 1996;

Kelly, 1992; Monaghan, Onnis, Merkx, & Chater, Submitted; Morgan, Shi, & Allopenna, 1996).

The second claim was that the non-adjacent dependencies separated by one word (or frames) are a linguistic unit that adults and infants as young as 18 months are able to manipulate and use in learning both artificial and natural language. It is true that non-adjacent dependencies are vital to language acquisition, as the link between *is* and *ing* in such phrases as *is* go*ing* needs to be learnt in order for grammar to develop normally. However, the overall claim that these non-adjacent dependencies are easily used by infants was only partially supported by the experimental evidence. The learning of non-adjacent dependency information has been found to be dependent on the variability of intervening material (Gómez, 2002; Onnis, Monaghan, Christiansen, & Chater, 2004). Additionally, phonological cues have also been shown to be vital in the learning of non-adjacent dependencies (Onnis, Monaghan, Richmond, & Chater, 2005); in some paradigms non-adjacent dependencies with no phonological cues are unlearnable. Although frequent frame distributional cues may indeed aid in the grammatical categorization process in both adults and infants (Mintz, 2002, 2006), it remains to be proven that frequent non-adjacent frames have distinct advantages over bigram distributional contexts or should be considered apart from additional linguistic cues, such as phonology.

Simplicity has been argued to be one of the main principles to all learning theories, and indeed cognitive science (Chater & Vitanyi, 2003, 2007). The Simplicity Principle argues that in any learning situation, the simplest solution will be the most parsimonious. Many perceptual organization phenomena are consistent with a simplicity preference; for example, the Gestalt laws of closure and good continuation are the simplest solutions to perceptual problems. More relevant for language acquisition, the whole object preference in new word learning is also an indication of an overarching principle of choosing the simplest solution. With this principle in mind, the proposal of non-adjacent frame distribution cues being necessary in the grammatical categorization process may be unnecessarily complicated as bigram distributional cues are sufficient when they are combined with other, known grammatical category cues. Furthermore, Mintz's proposal of the frame distributional cues (aXb) inducing grammatical categorization does not differentiate whether it is the combined influence of preceding and succeeding word working together or if it is the influence of the preceding word added to the influence of the succeeding word that drives the grammatical categorization process. For example, the phrase "you put it" may categorize "put" with all other words to go within the "you_it" frame due to the non-adjacent dependency or it may be due to the combined influence of the "you put" and "put it" bigram contexts.

Chapters 6 and 7 attempted to answer two main questions. The first was whether bigram distributional contexts or frame distributional contexts are more reliable in the grammatical categorization process. The second was whether the frame distributional contexts are driven by the combination of two bigram dependencies or from the higher order combination of the non-adjacent dependencies. Chapter 6 replicated the corpus analysis reported in Mintz (2003), which investigated how well aXb "frequent frames" grouped words of the same grammatical category together. This analysis was directly compared to groupings based solely on the preceding word (aX) and groupings based on the succeeding word (Xb). The preceding and succeeding word analyses will be discussed in terms of how they relate to the aXb frame analysis and also how they relate to the general preference for learning from succeeding cues and the results of Chapter 5. Chapter 7 reported the results of four Experiments that will attempt to separate the two possibilities in regards to learning from distributional aXb frames.

6.1. aXb Frequent frames analysis

6.1.1. Input corpora

The same six corpora from the CHILDES corpus of child-directed speech that Mintz (2003) analysed were used in all three analyses. The child corpora that were used were Anne and Aran (Theakson, Lieven, Pine, & Rowland, 2001), Eve (Brown, 1973), Naomi (Sachs, 1983), Nina (Suppes, 1974), and Peter (Bloom, Hood, & Lightbrown, 1974; Bloom, Lightbrown, & Hood, 1975). Only the sessions in which the child was 2;6 or younger were analysed. All utterances from children were excluded, leaving only adults' speech, the majority of which was child directed.

In the Aran and Nina corpora a few sessions were no longer available for analysis on the CHILDES database, therefore could not be included, and so the corpora here differ slightly from those of Mintz (2003). The Aran corpus contained aran1a to aran20b, with the exception of 14a and 14b. The Nina corpus contained nina01 to nina23, with the exception of nina8. The Anne corpus contained anne01a to anne23b. The Eve corpus contained eve01 to eve20. The Naomi corpus contained n01 to n58. Finally, the Peter corpus contained peter01 to peter12.

The analysis was done from the CHILDES MOR line, where the grammatical category of each word was labelled. This procedure differed slightly from previous analyses that consulted the CELEX database to obtain grammatical category information, as has been done in previous analyses (including Mintz, 2003). However, the two procedures used the same fundamental information, with the current analysis avoiding the lengthy disambiguation procedure that happens with words that have two or more CELEX grammatical category labels. There were several changes made to the corpus before the analysis procedure began. First all punctuation, pause marking, trailing off and interruption markings were replaced with a sentence boundary marker. This was done as all of these markings either signalled the end of a sentence, or signalled a break in the utterance, as in the case of pauses. All words that indicated a grammatical omission were deleted as well as any remaining CHILDES transcription codes. Importantly, in CHILDES any repetitions of words that were marked with "[/]" to indicate the repetition (e.g., "duck [/] duck") were fully transcribed on the normal transcription line, but only one version of the repetition was transcribed on the corresponding grammatical category line (MOR line). All repetitions were inserted on the grammatical category line by hand.

6.1.2. Analysis procedure

Each corpus was analysed separately. The procedure for the aXb "frequent frames" analysis will be covered in detail; a similar process was applied to the aX and Xb analyses as well. A list of all consecutive three word phrases was compiled for the aXb analysis (where a was the preceding word, X numerous medial words, and b was the succeeding word). None of these three word phrases crossed utterances boundaries. The 45 most frequent frames (a_b) were compiled with a list of all the medial words. The overall frequency of the aXb frame and the amount of times each X word occurred within the frame was computed.

The grammatical category label was then used to count the numbers of medial words that belonged to each category for each aXb frame. Two ways of tabulating the count data occurred; one counted the number of word types and one counted the number of word tokens. For the type analysis, each medial word was counted only once whereas the token analysis counted each individual occurrence of the medial words within the frequent frame. In other words, the token analysis was frequency weighted. For both token and type analyses the 45 most frequent frames within the token analysis were assessed.

Mintz (2003) used two types of category distinction in his analysis; standard and expanded labelling. Standard labelling included ten categories which were nouns (inclusive of pronouns), verbs (inclusive of auxiliaries and copula forms) adjectives, prepositions, adverbs, determiner, wh-words, “not”, conjunction and interjection. The expanded labelling counted nouns and pronouns separately, as well as verbs, auxiliaries and copula forms into independent categories. The same distinctions are used in this analysis, with the exception of the copula verb category. As this type of verb was not coded differently within the grammatical category CHILDES coding, these verbs were not differentiated in the expanded labelling.

In order to determine how well the frequent frames were able to categorize the words, accuracy and completeness measures were computed. The accuracy measure assesses categorization success by looking at pairs of medial words (or X words) within the distributional categories (or aXb frames). All possible pairs of words that were included in the analysis (and not the entire corpus) were used in computing accuracy and completeness. In order to compute accuracy the amount of hits and false alarms within the distributional category must be computed. A hit was when both words from the pair were of the same grammatical category. A false alarm was when each word from the pair was from a different grammatical category. Thus, hits gave a measure of correctly grouped words, whereas the false alarms measured the amount of incorrect groupings. Accuracy was the amount of hits divided by the total number of hits and false alarms (accuracy = hits/(hits + false alarms)). Accuracy gives an overall measure of how successful the category was at grouping words of the same grammatical category together.

Completeness measured how well the frequent frames grouped all words from one grammatical category together in the same frequent frame grouping. In order to compute the completeness measure the amount of misses were tabulated. To calculate misses all possible pairs of medial words throughout the whole analysis were computed. A miss was counted when both words in the pair were of the same grammatical category but were not in the same distributional category (i.e., was not categorized in the same aXb frequent frame). The completeness measure was then calculated by taking the hits and dividing by the total amount of hits and misses (completeness = hits/(hits + misses)).

Both accuracy and completeness have values that range from 0 to 1. A value of 1 for accuracy meant that the distributional category contained only one type of grammatical category, such as adjectives. A completeness score of 1 meant that the distributional category contains *all* of the words from a particular grammatical category, e.g., a distributional category that contained all adjectives (but potentially other grammatical categories as well). For the type analysis, only pairs of different word types were computed, whereas the token analysis looked at pairs of individual tokens. Results were computed for each corpus within the type/token distinction as well as for standard and expanded labelling. Thus, for each corpus there are four total analyses.

In addition to the analyses with the frequent frame data, a control condition was created for a random baseline measure of accuracy and completeness. All of the words that were categorized in the above analysis were randomly assigned across the 45 frequent frames to create a random corpus. Each frequent frame category contained the same number of random tokens that were found in the frequent frame based analysis. This baseline provides a measure of accuracy and completeness that would be expected if the frequent frames did not aid in grouping words of the same grammatical category together.

6.1.3. Results

All multiple *t* tests and pairwise comparisons were Bonferonni corrected. The tests that were significant before correction have been marked to give a more complete picture of the data; † indicates .05 level, †† indicates .01 level, and ††† indicates the .001 level. When the Levene's test for homogeneity of variance assumption was not met the test statistic for unequal variance was reported.

6.1.3.1. aXb Frequent frames analysis

Table 6.1 displays the data for the total number of word tokens and types within each of the corpora and then number categorized in the aXb frames. It was found that 6.3% of the total word tokens and 19.4% of the total word types were categorized by the analysis.

Table 6.1 Summary of the total number of word tokens and types in the corpora and the number of tokens and types included in the aXb analysis with the relevant percentage of tokens and types analysed.

Corpus	*Corpus tokens*	*Corpus types*	*Tokens Categorized*	*Types Categorized*	*Percentage of tokens analysed*	*Percentage of types analysed*
Anne	95255	2602	4870	388	5.1%	14.9%
Aran	106931	3249	6041	745	5.6%	22.9%
Eve	60929	2125	3430	386	5.6%	18.2%
Naomi	28979	1877	1725	315	5.9%	16.8%
Nina	70867	1968	6252	463	8.8%	23.5%
Peter	74170	2127	5204	429	7.0%	20.2%
Mean			**4587**	**454**	**6.3%**	**19.4%**

Table 6.2 and Table 6.3 show the accuracy and completeness measures for the token and type analyses with standard and expanded labelling. As can be seen in Table 6.2 the aXb frequent frames analysis had higher accuracy than the random analysis. The average for the token analysis with standard labelling was .93, which was significantly above the baseline of .28, $t(10) = 35.84, p < .005^{\dagger\dagger\dagger}$. The average accuracy decreased slightly for expanded labelling, but was still well above baseline, M = .80 and M = .21, respectively; $t(10) = 40.82, p < .005^{\dagger\dagger\dagger}$. The type analysis for standard labelling was significantly above baseline with an accuracy of .82 compared to .31, $t(10) = 26.07, p < .005^{\dagger\dagger\dagger}$. As with the token analysis, the expanded labelling reduced the accuracy slightly, M = .77 and M = .25 for the aXb and baseline analyses, respectively; $t(10) = 26.48, p < .005^{\dagger\dagger\dagger}$.

Table 6.2 Token and type accuracy measures for standard and expanded labelling for aXb and random aXb corpora with the comparison means from Mintz (2003)

Corpus	*Token Accuracy (Standard)*		*Token Accuracy (Expanded)*		*Type Accuracy (Standard)*		*Type Accuracy (Expanded)*	
	aXb	Random	aXb	Random	aXb	Random	aXb	Random
Anne	.94	.28	.81	.22	.82	.30	.74	.24
Aran	.88	.27	.81	.18	.76	.25	.72	.21
Eve	.95	.32	.78	.22	.80	.32	.74	.25
Naomi	.94	.32	.81	.22	.87	.34	.82	.28
Nina	.96	.32	.84	.25	.84	.30	.81	.25
Peter	.92	.28	.77	.19	.82	.32	.76	.27
Mean	**.93**	**.28**	**.80**	**.21**	**.82**	**.31**	**.77**	**.25**
Mintz	**.98**	**.46**	**.91**	**.27**	**.93**	**.47**	**.91**	**.38**

As with the accuracy data, the completeness measure was well above the random baseline for both token and type analyses, as can be seen in Table 6.3. The aXb completeness measure for the token analysis was .07 for standard labelling and .09 for expanded labelling, both of which were significantly higher than the baseline of .03, $t(10) = 11.44, p < .005^{\dagger\dagger\dagger}$ for standard and $t(10) = 16.21, p < .005^{\dagger\dagger\dagger}$ for expanded labelling. For the type analysis completeness was higher for the aXb frequent frames analysis than baseline, M = .08 and M = .09 for standard and expanded labelling and M = .04 baseline for both labelling conditions, $t(10) = 6.33, p < .005^{\dagger\dagger\dagger}$ for standard and $t(6.79) = 7.39, p < .005^{\dagger\dagger\dagger}$ for expanded labelling.

Table 6.3 Token and type completeness measures for standard and expanded labelling for aXb and random aXb corpora with the comparison means from Mintz (2003)

Corpus	*Token Completeness (Standard)*		*Token Completeness (Expanded)*		*Type Completeness (Standard)*		*Type Completeness (Expanded)*	
	aXb	Random	aXb	Random	aXb	Random	aXb	Random
Anne	.07	.03	.08	.03	.08	.04	.09	.04
Aran	.08	.03	.10	.03	.09	.04	.10	.04
Eve	.06	.03	.08	.03	.07	.03	.08	.03
Naomi	.07	.03	.09	.03	.06	.04	.07	.04
Nina	.08	.04	.09	.04	.10	.05	.11	.05
Peter	.07	.03	.09	.03	.08	.04	.08	.04
Mean	**.07**	**.03**	**.09**	**.03**	**.08**	**.04**	**.09**	**.04**
Mintz	**.07**	**.03**	**.12**	**.03**	**.08**	**.04**	**.10**	**.04**

6.1.3.2. *Discrepancy with Mintz results*

As the aXb analysis was a replication of previous published work, using the same corpora, the results of the current analysis were directly compared to the results reported in Mintz (2003). As can be seen in Table 6.2 and Table 6.3 there were differences between the previous and current analysis. There was no difference in the token and type completeness measures with standard labelling, but the completeness scores were slightly higher in the Mintz (2003) analysis with expanded labelling. With the accuracy scores there were differences in both standard and expanded labelling across token and type analyses, but the differences were more pronounced with the expanded labelling. However, differences with this type of labelling were explained by the exclusion of the copula category.

As four of the six corpora should have been identical between the two analyses the differences within the standard labelling were unexpected and were therefore investigated further. The specific token counts for the frames that were reported in detail for the Peter and Naomi corpora in the Mintz (2003, p. 99) paper were directly compared to the token counts found within the same frames in the current analysis.

All word types and their token frequency were compared for the "you_it", "I_it" and "the_one" frames from the Peter corpus and the "you_it" and "the_is" frames from the Naomi corpus; only the "you_it" Peter frame comparison was described in detail below. There were 93 word types reported within the Mintz "you_it" frame. As can be seen in Table 6.4 40 of the 93 types had discrepancies between the token frequency that were reported in Mintz (2003) and those found in the current analysis. The original Peter CHILDES session files (Peter01-Peter12) were consulted and the frequency numbers for each medial word matched the current analysis.

The precise reason for the discrepancy with the Mintz analysis and the current analysis was next investigated. The original files were consulted and the numbers in parentheses in the Mintz/current discrepancy column of Table 6.4 were the amount of tokens that could possibly be accounted after searching through the original files. As can be seen from this column, 15 of the 40 types were accounted for. The possible reasons for the differences were outlined below. As many discrepancies could not be accounted for, a plausible explanation was that the Mintz input included extra material that was not from the Peter01 to Peter12 CHILDES files.

Table 6.4 Medial words for the "you_it" frame and their frequency in the current analysis, as reported in Mintz (2003) and in the original Peter session files with discrepancy figures.

Frame	*Medial words*	*Mintz*	*Current*	*Mintz/current Discrepancy*	*Original corpus*
You_it	put	52	43	9 (10)	43
	see	28	23	5 (3)	23
	do	27	24	3 (1)	24
	did	25	23	2 (0)	23
	want	23	20	3 (0)	20
	fix	13	12	1 (1)	12
	turned	12	11	1 (0)	11
	get	12	10	2 (2)	10
	got	11	8	3 (0)	8
	turn	10	7	3 (3)	7
	throw	10	4	6 (2)	4
	closed	10	8	2 (0)	8
	open	8	9	-1 (-1)	9
	find	8	6	2 (0)	6
	bring	8	6	2 (0)	6
	took	7	8	-1 (-1)	8
	knocked	6	0	6 (0)	6
	make	4	3	1 (1)	3
	fixed	4	3	1 (0)	3
	try	3	1	2 (2)	1
	opened	3	2	1 (0)	2
	need	3	4	-1 (0)	4
	move	3	1	2 (0)	1
	give	3	2	1 (0)	2
	fixing	3	4	-1 (-1)	4
	drive	3	4	-1 (0)	4
	catch	3	4	-1 (-1)	4
	threw	2	1	1 (1)	1
	taking	2	1	1 (0)	1
	pushing	2	1	1 (0)	1
	had	2	3	-1 (0)	3
	eat	2	1	1 (0)	1
	build	2	1	1 (0)	1
	brought	2	0	2 (0)	0
	wind	1	0	1 (0)	0
	knew	1	0	1 (0)	0
	expected	1	0	1 (0)	0
	drop	1	0	1 (0)	0
	mean	0	1	-1 (0)	1
	use	0	1	-1 (0)	1

The main potential reason for the differences found were the occurrences of phrases such as "you wan(t) (t)a put it …". The Mintz corpora were "…minimally treated before the distributional analysis procedure was preformed. All punctuation was removed and all special CHILDES transcription codes were removed" (Mintz, 2003, p. 95). One conjecture was that the "wan(t) (t)a" words were erroneously removed as they contained parentheses, which might have been taken for extra CHILDES transcription coding. This would leave the utterance "you put it", which would be included in the count of the medial word "put" within the "you_it" frame. This would leave the Mintz analysis with more total tokens than the current analysis and original corpora. This error either accounts for or partially accounts for the discrepancies in the "you put it", "you fix it", "you get it", "you turn it", "you make it", "you try it", "you see it", "you do it", and "you throw it" phrases.

The original Peter files contained an additional "you open it" phrase than what was reported in Mintz (2003). This may be due to the line "<you [!!]> [<] <open it> [<] !" The [!!] indicated that the "you" word was emphasized by the speaker and the < and > symbols indicated that this particular phrase overlapped with the preceding and following utterances. It may be that the underlying "you open it" phrase was interrupted when the special CHILDES coding was removed in the Mintz analysis, but remained unaltered in the current analysis. The "you took it" phrase had similar coding in the original file ("you [!!] took it off") which may again have accounted for the Mintz analysis containing one less "you took it" phrase than the original files. There were no reasons found for the discrepancy in the remaining phrases. The individual type and token data from a further five frames that were detailed in Mintz (2003) was compared in the same manner as above with similar results. See Appendices J to M for table summaries of these analyses.

Though there were some discrepancies between the analyses, the overall results of the Mintz analysis were consistent with the current analysis, though Mintz (2003) reported slightly higher accuracy levels.

6.1.3.3. aX Preceding word analysis

Table 6.5 shows a summary of the total word types and tokens categorized in the aX preceding word analysis. There were overall many more of the total corpus tokens and types accounted for in this analysis than in the aXb analysis, with an average of 42.9% of the tokens and 85.6% of the word types analysed.

Table 6.5 Summary of the total number of word tokens and types in the corpora and the number of tokens and types included in the aX analysis with the relevant percentage of tokens and types analysed.

Corpus	*Corpus tokens*	*Corpus types*	*Tokens Categorized*	*Types Categorized*	*Percentage of tokens analysed*	*Percentage of types analysed*
Anne	95255	2602	39071	2235	41.0%	85.9%
Aran	106931	3249	47822	2810	44.7%	86.5%
Eve	60929	2125	23890	1776	39.2%	83.6%
Naomi	28979	1877	11598	1503	40.0%	80.1%
Nina	70867	1968	34402	1811	48.5%	92.0%
Peter	74170	2127	32384	1813	43.7%	85.2%
Mean			**31528**	**1991**	**42.9%**	**85.6%**

The preceding word analysis produced overall less accurate categories in terms of grammatical category grouping than the aXb analysis. However, as can be seen in Table 6.6, all analyses using only the preceding word were far more accurate at categorizing words from the same grammatical category together than would be expected by chance, as measured by the random corpora. For the token analysis accuracy was at .50 for standard labelling and .41 for expanded labelling, which was well above baseline scores of .17 and .11, $t(5.62) = 18.49, p < .005^{\dagger\dagger\dagger}$ for standard and $t(10) = 15.48, p < .005^{\dagger\dagger\dagger}$ for expanded labelling. The type analysis accuracy was similarly well above the random baseline corpora; standard labelling: $t(10) = 10.90, p < .005^{\dagger\dagger\dagger}$, M = .42 and M = .19 for the aX and random corpora; expanded labelling: $t(10) = 10.66, p < .005^{\dagger\dagger\dagger}$, M = .37 and M = .14, respectively.

Table 6.6 Token and type accuracy measures for standard and expanded labelling for aX and random aX corpora

Corpus	*Token Accuracy (Standard)*		*Token Accuracy (Expanded)*		*Type Accuracy (Standard)*		*Type Accuracy (Expanded)*	
	aX	Random	aX	Random	aX	Random	aX	Random
Anne	.48	.17	.40	.10	.41	.20	.36	.14
Aran	.43	.16	.35	.10	.38	.18	.33	.13
Eve	.51	.17	.38	.11	.42	.19	.37	.14
Naomi	.52	.18	.41	.11	.47	.19	.43	.14
Nina	.55	.19	.49	.12	.47	.23	.43	.17
Peter	.48	.17	.41	.10	.36	.16	.31	.11
Mean	**.50**	**.17**	**.41**	**.11**	**.42**	**.19**	**.37**	**.14**

Table 6.7 shows a summary of the completeness data for the preceding word analysis. As expected the aX analysis produced more complete categories than the random corpora. The token analysis was well above chance, with completeness scores of .09 and .12 for standard and expanded labelling, $t(7.54) = 8.37, p < .005^{\dagger\dagger\dagger}$ for standard and $t(6.47) = 11.66, p < .005^{\dagger\dagger\dagger}$ for expanded; M = .04 for both random analyses. The type analysis showed a similar pattern of results with a standard completeness score of .07 and expanded labelling producing a .09 completeness value. Both of these measures were well above chance level, $t(5) = 15.81, p < .005^{\dagger\dagger\dagger}$ for standard and $t(5) = 14.0, p < .005^{\dagger\dagger\dagger}$ for expanded labelling; M = .04 for both random analyses.

Table 6.7 Token and type completeness measures for standard and expanded labelling for aX and random aX corpora

Corpus	*Token Completeness (Standard)*		*Token Completeness (Expanded)*		*Type Completeness (Standard)*		*Type Completeness (Expanded)*	
	aX	Random	aX	Random	aX	Random	aX	Random
Anne	.08	.04	.11	.04	.07	.04	.08	.04
Aran	.07	.03	.10	.03	.07	.04	.08	.04
Eve	.09	.04	.12	.04	.07	.04	.08	.04
Naomi	.10	.04	.13	.04	.08	.04	.10	.04
Nina	.10	.05	.14	.05	.08	.04	.09	.04
Peter	.08	.04	.14	.04	.07	.04	.09	.04
Mean	**.09**	**.04**	**.12**	**.04**	**.07**	**.04**	**.09**	**.04**

6.1.3.4. Xb Succeeding word analysis

Table 6.8 shows the data for the amount of word tokens and types to be categorized in the Xb succeeding word analysis. As with the aX analysis, there were more word types and tokens categorized by only the succeeding word when compared to the aXb analysis. In total, 38.1% of all tokens in the corpora were captured by this analysis, as were 67.2% of all word types.

Table 6.8 Summary of the total number of word tokens and types in the corpora and the number of tokens and types included in the Xb analysis with the relevant percentage of tokens and types analysed.

Corpus	*Corpus tokens*	*Corpus types*	*Tokens Categorized*	*Types Categorized*	*Percentage of tokens analysed*	*Percentage of types analysed*
Anne	95255	2602	36101	1843	37.9%	70.8%
Aran	106931	3249	45006	2469	42.1%	76.0%
Eve	60929	2125	20807	1268	34.1%	59.7%
Naomi	28979	1877	10156	1082	35.0%	57.6%
Nina	70867	1968	28955	1446	40.9%	73.5%
Peter	74170	2127	28661	1400	38.6%	65.8%
Mean			**28281**	**1585**	**38.1%**	**67.2%**

The succeeding word analysis had overall less accurate grammatical category groupings than the combined aXb analysis. However, word groupings based solely on the succeeding word were more accurate than expected by chance for both the token and type analyses. The token analysis accuracy was .31 and .24 for the standard and expanded labelling, which was significantly above the corresponding .17 and .12 accuracy levels from the baseline corpora; $t(10) = 17.86, p < .005^{\dagger\dagger\dagger}$ for standard and $t(10) = 19.76, p < .005^{\dagger\dagger\dagger}$ for expanded labelling. For the type analysis the accuracy values were .26 and .21 for standard and expanded labelling which were significantly above the random baseline of .18 and .13; $t(10) = 5.62, p < .005^{\dagger\dagger\dagger}$ for standard and $t(10) = 6.09, p < .005^{\dagger\dagger\dagger}$ for expanded labelling.

Table 6.9 Token and type accuracy measures for standard and expanded labelling for Xb and random Xb corpora

Corpus	*Token Accuracy (Standard)*		*Token Accuracy (Expanded)*		*Type Accuracy (Standard)*		*Type Accuracy (Expanded)*	
	Xb	Random	Xb	Random	Xb	Random	Xb	Random
Anne	.31	.17	.24	.12	.25	.18	.19	.13
Aran	.32	.17	.24	.11	.27	.19	.22	.14
Eve	.29	.16	.23	.11	.25	.18	.20	.12
Naomi	.31	.16	.24	.11	.24	.17	.20	.12
Nina	.34	.17	.27	.12	.30	.21	.25	.15
Peter	.29	.17	.23	.12	.22	.17	.17	.12
Mean	**.31**	**.17**	**.24**	**.12**	**.26**	**.18**	**.21**	**.13**

The completeness values for the succeeding word analysis were generally lower than the preceding word and frequent frames analysis, but the groupings were still more complete than would be expected by chance. The token analyses had higher completeness scores than the baseline corpora, M = .07 and M = .04 for standard labelling and M = .08 and M = .04 for expanded labelling; $t(10) = 8.0, p < .005^{\dagger\dagger\dagger}$ and $t(10) = 7.42, p < .005^{\dagger\dagger\dagger}$, respectively. The completeness scores were slightly lower for the type analysis, with M = .04 and M = .05 for standard and expanded labelling. However, these score were still above the random baseline scores of .03 (for both baseline analyses), $t(5) = 6.33, p < .005^{\dagger\dagger\dagger}$ for standard labelling and $t(5) = 7.91, p < .005^{\dagger\dagger\dagger}$ for expanded labelling.

Table 6.10 Token and type completeness measures for standard and expanded labelling for Xb and random Xb corpora

Corpus	*Token Completeness (Standard)*		*Token Completeness (Expanded)*		*Type Completeness (Standard)*		*Type Completeness (Expanded)*	
	Xb	Random	Xb	Random	Xb	Random	Xb	Random
Anne	.06	.04	.07	.04	.04	.03	.04	.03
Aran	.07	.04	.08	.04	.04	.03	.04	.03
Eve	.07	.04	.07	.04	.05	.03	.05	.03
Naomi	.07	.04	.08	.04	.04	.03	.05	.03
Nina	.08	.05	.10	.05	.05	.03	.05	.03
Peter	.07	.05	.08	.05	.04	.03	.05	.03
Mean	**.07**	**.04**	**.08**	**.04**	**.04**	**.03**	**.05**	**.03**

6.1.3.5. *Combined analysis*

The frequent frames, preceding and succeeding word analyses were directly compared with a three way ANOVA in terms of the amount of word tokens and types that were categorized. When the results do not differ between the token and type data, only the token ANOVA will be reported.

When looking at the overall numbers of tokens categorized, there was a significant main effect indicating that the aXb, aX and Xb analyses differed, $F(2,15) = 12.74, p = .001, \eta_p^2 = .63$. Pairwise comparisons indicated that the aXb analysis categorized fewer tokens than the aX or Xb analyses ($p < .01$ for both comparisons) but the aX and Xb analyses did not differ ($p = 1.0$).

The aXb, aX and Xb accuracy and completeness scores were directly compared with one way ANOVAs with analysis (aXb, aX and Xb) as the factor. The dependent measures were the accuracy and completeness data. Standard labelling was used for all the analyses, as the results did not differ between standard and expanded labelling.

When investigating the token accuracy data, there was a significant main effect of analysis, $F(2, 18) = 635.38, p < .001, \eta_p^2 = .99$. This main effect was broken down into pairwise comparisons between the aXb, aX and Xb groups. There were significant differences between all three groups such that the aXb analysis induced the most accurate grouping of grammatical categories, followed by the aX group with the Xb groupings as the least accurate, M = .93 for aXb, M = .50 for aX and M = .31 for Xb; all ps $< .001^{\dagger\dagger\dagger}$.

When looking at the token completeness measure there was a significant main effect of analysis, $F(2, 18) = 6.23, p < .05, \eta_p^2 = .45$. Pairwise comparisons indicated that there was significant difference between the aX and aXb analysis ($p < .05^{\dagger}$) and between the aX and Xb analysis ($p < .05^{\dagger}$), but there was no difference between the Xb and aXb analysis ($p = 1.0$). This result showed that the aX analysis produced distributional categories that captured more total words of the same grammatical category than the aXb and Xb analysis, M = .09 for aX and M = .07 for aXb and Xb.

The analysis main effect for the type completeness measures was also significant, $F(2, 18) = 29.01, p < .001, \eta_p^2 = .80$. This significant effect was different from the previous analysis, as the differences were due to the low completeness in the Xb analysis which differed from the higher completeness in the aXb and aX analyses, $p < .001^{\dagger\dagger\dagger}$ for both comparisons; M = .05 for Xb, M = .08 for aXb and aX. There were no differences between the aX and aXb analyses, $p = .64$.

6.1.4. Discussion

The differences between the results of the aXb analysis reported in Mintz (2003) and the results reported in this Chapter were found to be due to a combination of slight differences in the input files, analysis procedure and differences in the Mintz aXb output and the original files. However, the overall findings of the Mintz (2003) paper were replicated. The aXb frequent frame analysis was very accurate at grouping words of the same grammatical category, with average accuracy ranging from .76 to .93. The completeness values of the resulting groupings with standard labelling were precisely replicated; the Mintz analysis produced higher completeness values for the expanded labelling, but this was explained by the slight differences in expanded labelling between Mintz and the current analysis.

The results indicated that the aXb analyses were much more accurate at grouping words of the same grammatical category together than both of the bigram analyses. Overall token accuracy for the aXb analysis was .87, but this dropped to .45 when only looking at preceding word cues and .28 for succeeding word cues. This result was as expected, as the resulting words in the groupings were much more constrained when obtained with a two word context, as in the aXb analysis.

However, the high accuracy in the aXb analysis came at a price; the amount of total word types and tokens that were categorized was much lower in the aXb frequent frame analysis than in the aX and Xb analysis. The amount of tokens and types analysed with the aXb distributional cue was only 6.3% and 19.4%, respectively. The corresponding scores were 42.9% and 85.6% with the aX analysis and 38.1% and 67.2% with the Xb analysis. Therefore, although the aXb analysis produced grouping that were more homogenous in terms of grammatical category, there were far fewer words actually categorized than the bigram analyses.

This trade off between accuracy and amount of words categorized was found in the completeness scores. The aX analysis produced distributional groupings that had higher completeness than the aXb and Xb analysis for the token data. This result indicated that the preceding distributional cues induced groupings that accounted for more of the total amount of grammatical category members than either the aXb or Xb analyses. The Xb analysis did not differ from the aXb analysis, which was interesting as the Xb analysis captured the same amount of total word types and tokens as the aX analysis. The pattern changes for the type data, in that the aXb and aX analysis had the same completeness score; the Xb analysis was still far below the aX analysis. This indicated that the lower completeness score in the aXb token analysis was influenced by the overall lower number of tokens categorized.

When looking at only the bigram analyses it was clear that the preceding word cue provided both more accurate and more complete grammatical category groupings than the succeeding word cue. This result was both expected and unexpected. When considering the general pattern for higher learning from succeeding cues that has been found throughout this thesis and predicted by the R-W model the result was surprising as it may have been expected that succeeding word cues would induce more accurate and complete groupings. However, the results of Chapter 5 indicated that the succeeding cue advantage was increased when the marker words were similar to suffixes, indicating that the succeeding cue advantage perhaps has roots at the inflectional level, not the word level. Chapter 5 also indicated that there was increased learning in the preceding cue condition when the marker words were more similar to words than affixes (the *cvc* marker words). Furthermore, analysis on the linguistic level provides support for the idea that consistent preceding cues are at the word level while consistent succeeding cues are at the morpheme level, as demonstrated by the suffixing preference across languages (Dryer, personal communication as cited in Hall, 1988). Therefore, the results that the aX groupings were more accurate and complete than the Xb groupings were consistent with the preceding word/succeeding affix hypothesis.

As discussed above the frequent frame analysis induced more accurate grammatical category groupings than either of the bigram analyses. However, the total amount of words grouped must also be taken into account when weighing the claim that the frequent frames are instrumental in the process of grammatical categorization. The results of the aX analysis indicated that the preceding word not only produced an adequate level of accuracy in grammatical category grouping but also produced groupings that were more complete in terms of capturing more of the total number of nouns/verbs in the whole analysis. In addition to this, the aX analysis captured 6.9 times the total words that were included in the aXb analysis. The decrease in accuracy may have been more than made up for the increase in the amount of the corpora that were accounted for. This result was in line with previous research that showed neural networks could learn categories from aX bigram information, but not aXb trigram information (Monaghan & Christiansen, 2004); it may be that trigrams induce very accurate categories but are not sufficiently usable to aid category learning. However, when the trigrams were allowed to use the initial and ending bigram information separately, the neural networks readily leant the categories (Monaghan & Christiansen, 2004).

This possibility that the aXb frequent frame advantage was due to the combination of the aX and Xb bigrams could not be directly tested in this analysis. However, the amount of overlap in the beginning and ending cue words that were used in the frequent frames (i.e., the "a" and "b" in aXb) and the preceding cue words (i.e., aX analysis) and succeeding cue words (i.e., Xb analysis) was computed for all corpora. It was found that an average of 95% of the "a" words in the aXb analysis were also used as beginning word cues in the aX analysis. For the "b" words in the aXb analysis, there was an average of 93% overlap with the succeeding word cues in the Xb analysis. This analysis, while not conclusive, was suggestive that categorization could occur through aXb frequent frames simply due to the influence of the aX and Xb information separately.

In order to test the above hypothesis Experiments 8 and 9 were conducted. Experiment 8 was an aXb Experiment which tested whether learning can occur through trigram distributional information alone and whether this learning was increased with additional phonological cues. Experiment 9 used a similarly structured aXb Experiment, but the a_b dependency was broken while the initial and ending bigram distributional information remained informative. The hypothesis that learning would occur if the aX and Xb dependencies were intact, but the a_b dependency was broken was investigated in this Experiment.

7. Comparison of trigram and bigram distribution cues in grammatical categorization

This chapter directly investigated the useability of the frame distributional cues discussed in the previous chapter. Whether these frequent frames could induce grammatical categorization without addition phonological cues was investigated first through an extension of a study conducted by Mintz (2002)[9]. The Mintz artificial language learning experiment used distributional frames as the only cue to aid categorization. Mintz found evidence of categorization, but this evidence was relatively weak due to the small number of test sentences that directly tested categorization. Experiment 8 investigated a similar artificial language that was composed of trigram (or framing) distributional cues; however, the design of this Experiment greatly increased the power by increasing the amount of critical test items overall.

In addition to the change in the experimental design, the presence or absence of phonological cues within the category words was manipulated. The additional test for phonological cues was included to increase the relevance of the word categorization process within the Experiment to natural language acquisition; substantial evidence points to the conclusion that children make use of both phonological and distributional correlates of grammatical categories during the acquisition process (Cassidy & Kelly, 1991, 2001; Kelly, 1992, 1996; Monaghan, Chater, & Christiansen, 2005; Morgan, Shi, & Allopenna, 1996). The inclusion of the phonological coherence manipulation also allowed comparisons between the bigram distributional cues in Experiments 1 to 4 and the trigram distributional cues in the current Experiment.

Experiment 9 directly tested whether any learning found within the trigram distributional languages was due to the a_b framing words combined or whether it was due to learning from the initial bigram (aX) and learning from the ending bigram (Xb) separately. This Experiment retained the usefulness of the bigram cue co-occurrences, but the trigram non-adjacent cues were broken; the first word no longer perfectly predicted the third word. If equal learning was found within this design, any learning within the trigram Experiment may perhaps be attributable only to learning from bigrams.

---

[9] Two separate replications of the Mintz (2002) experiment were conducted and the second replication used the precise stimuli that were in the original study, although different recordings. Both attempts failed to replicate the original study.

7.1. *Experiment 8*

Experiment 8 investigated learning from trigram distributional information, or the a_b framing cues. This Experiment also included phonological cues within the category words to provide a direct comparison between the bigram Experiments in Chapters 2 and 3 and also to provide a more realistic category learning language. The phonological correlates to grammatical category have been reviewed in detail and the evidence point to their importance during the grammatical categorization process (e.g., Cassidy & Kelly, 1991, 2001; Kelly, 1992; Monaghan, Chater, & Christiansen, 2005). The design of Experiment 8 also allowed a greater number of critical test items than in Mintz (2002). This was done by using a 2 alternative forced choice method, in which correctly categorized sentences were directly compared to incorrectly categorized sentences. It was predicted that the correctly categorized sentences would be endorsed more often that the incorrectly categorized sentences. The remaining change was the inclusion of a generalization test at the end of the Experiment. This consisted of the same 2 alternative forced choice format, but testing novel category words that contained the same phonological characteristics of the trained category words. This tested whether the participants only learned the category words, or if the general phonological form of the category words could be applied to new instances of category words.

7.1.1. Participants

Thirty-four University of York undergraduate students participated in this Experiment. Participants had not taken part in similar Experiments. Two participants were excluded; one due to uncertainty about whether English was their native language and the other due to technical faults. The average age was 19.25 years; range was 18 to 22. There were 28 female and four male participants in the included sample. All included participants were native English speakers and were paid £4 or given course credit.

7.1.2. Materials

An artificial language was created that contained trigram distributional cues, which consisted of "frames" between the first and the third word in each three word sentence. The first word perfectly predicted the third word and the middle word was one of 16 category words (eight words per category). There were four frames which were divided evenly between two categories. Within each category, one of the frames contained all eight category words and was designated the "complete category" frame. The other frame was designated the "partial category" frame as it contained only five of the category words, with the remaining sentences withheld until the test session. This created 13 unique training sentences per category. See Table 7.2 for a summary of the training items.

The individual framing and category words are listed in Table 7.1. All training and test sentences were synthesized with the Festival Speech Synthesizer (Black, Clark, Richmond, King, & Zen, 2004). Each framing word was randomly assigned to one of the framing word positions (i.e. A, B, C, D, E, F, G, or H) and the combination of individual framing words to each position varied between participants. The advantage of this arrangement was that any learning could not be attributed to particularly salient combinations of framing words.

Table 7.1 Words used in Experiment 8

	Phonologically Coherent		*Phonologically Incoherent*	
Frame words	*Category A*	*Category B*	*Category A*	*Category B*
jiv	beng	thorsh	vawse	thorsh
pim	dimp	vawse	foth	beng
sij	gemb	joov	gemb	joov
vot	pihnt	foth	pihnt	dimp
rud	keend	zodge	zodge	keend
guf	gihnk	shufe	gihnk	pemk
nep	timp	suwch	timp	suwch
lan	pemk	choz	shufe	choz

As in Experiments 1 to 4, there was a phonologically coherent condition, with both phonological and distributional cues, and a phonologically incoherent condition, with only distributional cues. The specific words used in both conditions are given in Table 7.1. The category A words in the coherent condition all contained nasals, stops, offset consonant clusters, unrounded high vowels, and no fricatives whereas the words in category B contained no consonant clusters, no nasals or stops, generally low vowels and fricatives. For the incoherent condition half of the words from category B were exchanged with half of the words from category A, as shown in Table 7.1.

The categorization test session consisted of a 2 alternative forced choice between compatible and incompatible test sentences in which participants had to judge which sentence was most similar to training sentences. The compatible test sentences conformed to the regularities of the language and were the remaining sentences from the partial category frame not used as training sentences. These compatible test sentences were paired with incompatible test sentences, which were composed of the same frames as the compatible test sentences, but the category words were of the opposing category. Each compatible test sentence was paired with every incompatible test sentences, creating a total of 18 test contrasts. See Table 7.2 for a summary of the test items.

Table 7.2 Summary of training and test sentences in Experiment 8.

Training			
Category A		Category B	
Complete	*Partial*	*Complete*	*Partial*
AX_1B	CX_1D	EX_9F	GX_9H
AX_2B	CX_2D	$EX_{10}F$	$GX_{10}H$
AX_3B	CX_3D	$EX_{11}F$	$GX_{11}H$
AX_4B	CX_4D	$EX_{12}F$	$GX_{12}H$
AX_5B	CX_5D	$EX_{13}F$	$GX_{13}H$
AX_6B		$EX_{14}F$	
AX_7B		$EX_{15}F$	
AX_8B		$EX_{16}F$	
Categorization Test			
Category A		Category B	
Compatible	*Incompatible*	*Compatible*	*Incompatible*
CX_6D	$CX_{14}D$	$GX_{14}H$	GX_6H
CX_7D	$CX_{15}D$	$GX_{15}H$	GX_7H
CX_8D	$CX_{16}D$	$GX_{16}H$	GX_8H
Generalization Test			
Compatible “new”	*vs. compatible “trained”*	*Compatible “new”*	*vs. incompatible “trained”*
CY_1D	CX_6D	CY_1D	$CX_{14}D$
CY_2D	CX_7D	CY_2D	$CX_{15}D$
CY_3D	CX_8D	CY_3D	$CX_{16}D$
Incompatible “new”	*vs. compatible “trained”*	*Incompatible “new”*	*vs. incompatible “trained”*
CY_4D	CX_6D	CY_4D	$CX_{14}D$
CY_5D	CX_7D	CY_5D	$CX_{15}D$
CY_6D	CX_8D	CY_6D	$CX_{16}D$

A second generalization test session was created to test whether participants were learning general characteristics of the category words. Six new words were created, half of which conformed to the general phonological characteristics of category A words (Y_1 Y_2 Y_3; keng, geend, pimb) and the other half to the phonological characteristics of category B (Y_4 Y_5 Y_6; shoov, zuwth, fodge). With each test trial, the "new" category word sentence was compared with a test sentence from the first test session. There were four types of test contrasts: the "new" word was either in compatible or incompatible distributional context and was compared to either a compatible or incompatible test sentence with trained category words. See Table 7.2 for a summary of the test contrasts. There were 24 total test contrasts in the second test session, with half utilizing the category A frame and the other half using the category B frame. See Appendix N for complete details of the generalization test items.

For the categorization test session, the test sentences were counterbalanced such that the compatible sentence was presented first for half of the contrasts and second for the other half. In addition, as each test sentence occurred in multiple contrasts, the amount of times each sentence appeared as the first sentence was counterbalanced as evenly as possible. However, as each test sentence occurred three times, a between subjects counterbalancing condition was created for complete counterbalancing. The order in which the test sentences were presented within the contrasts was reversed for half of the participants. For the generalization test the sentences were counterbalanced so that the position of the sentences with the "new" category words were presented equally often as the first and second sentence. Additionally, the pairings of individual category words were not repeated irrelevant of the framing context. A between subjects counterbalance of the order in which the test sentences were presented occurred with the generalization test as well.

7.1.3. Procedure

Before training, participants heard each of the nonwords for familiarisation with the synthesized speech materials. Participants were then instructed that they had to learn the structure of a language with no meaning, followed by 20 presentations of each training sentence in a randomised order. There was a 30 millisecond pause between the individual words in the sentences and a 500 millisecond pause between the sentences. For the test session, they were told to indicate the sentence that sounded most similar to the training sentences by pressing 1 (for the 1st sentence) or 2 (for the 2nd sentence) on the keyboard. After the first test session a further three presentations of the training sentences occurred after which the participants were tested on the generalisation sentences, with the same instructions detailed above.

7.1.4. Results

All multiple *t* tests were Bonferonni corrected. The tests that were significant before correction have been marked to give a more complete picture of the data; † indicates .05 level, †† indicates .01 level, and ††† indicates the .001 level.

The dependent variable from the first test session was the number of times the participants correctly chose the compatible test item. The coherent and incoherent conditions were directly compared on the results of the first test session. There was a significant difference between the coherency conditions in that the group with both distributional and phonological cues showed evidence of grammatical categorization, $t(30) = 4.07, p < .001$, M = 11.0 for the coherent condition and M = 8.06 for the incoherent condition. This replicated the main findings from Chapters 2 and 3 that phonological and distributional cues show more evidence of grammatical categorization than distributional cues alone.

The coherency conditions were next compared to the chance level of nine. The coherent condition was significantly better than expected by chance, $t(15) = 3.41, p < .01^{\dagger\dagger}$, while the incoherent condition was not significantly different from chance level, $t(15) = -2.33, p = .07^{\dagger}$. This also replicated Chapters 2 and 3 by extending the main finding that categorization only occurs when the two types of cues are combined to a trigram distributional language.

The dependent variable from the second test session was the amount of times the participant chose the "new" category word sentence as similar to the artificial language. As the analysis was testing for generalization from the phonologically coherent cues, only the participants who were in the coherent condition were considered; therefore the data from the incoherent condition was excluded. A repeated measures ANOVA was conducted with the status of the "new" words (compatible or incompatible) and the status of the trained category words (compatible or incompatible) as the within subjects factors. A dummy variable (counterbalancing) was included as between subjects factor.

The new word status main effect was significant, $F(1,14) = 8.93, p = .01, \eta_p^2 = .39$. The participants were more likely to choose the new word sentences when they were in a compatible, rather than incompatible, distributional context, M = 2.28 for compatible context and M = 1.69 for incompatible context. This indicated that the participants were sensitive to the general phonological characteristics of the category words and were able to apply this knowledge when encountering novel, but phonological similar category words in compatible and incompatible circumstances. There were no other significant main effects or interactions: trained word status: $F(1,14) = 1.16, p = .30$; all others $F < 1$.

7.1.5. Discussion

This Experiment replicated Experiments 1 through 4 in that there was only evidence of grammatical categorization when both distributional and phonological cues were combined to predict category membership. Trigram distributional cues were no different than the bigram cues in the previous Experiments and could not induce grammatical categorization without the additional phonological cues.

Unique to this Experiment was the evidence that participants could generalize to new category words that were phonologically similar to the known category words. The participants were more likely to say that the sentence with the novel category word was similar to the training language when the novel word was in a consistent distributional context. When the phonological form of the new category word and distributional context were incongruent the participants were less likely to endorse the sentence as being from the training language.

The ability to generalize from specific, known instances to novel instances is an important aspect of language acquisition. Grammar cannot be learnt without the ability to generalize beyond known phrases and words. In learning a new noun, such as *goose*, the child must be able to generalize from other, known nouns to this new word. If this generalization takes place, the new noun *goose* can be used in many of the same grammatical contexts as other nouns. For instance, if the new noun is generalized to the general category that includes the word *zoo* and the child knows that the phrase "*I like the zoo*" is permissible, the phrase "*I like the goose*" should also be permissible. This process is not infallible, as Pinker (1994) points out, as many words belong to more than one category. However, the general conclusion that is emerging from the language learning field is that by taking into account the sheer frequency of co-occurrences that happen in everyday language this problem ceases to be problematic (e.g. Mintz, 2003;

Monaghan, Chater, & Christiansen, 2005). Importantly, these two words are also linked by phonological cues that have been shown to differentiate nouns and verbs; in other words, cues that give coherence to the noun category. *Goose* and *zoo* both contain high, back vowels and coronal consonants, which Monaghan et al. (2005) demonstrated are characteristics of the noun category.

Therefore, the generalization from the known category words to phonologically similar novel category words indicated that the categorization within the Experiment was similar to grammatical categorization during language acquisition. The participants not only learned the trained category words but were able to apply their knowledge of the phonological cues within those categories to new instances of category words. They made the leap from their known words to new instances and knew in what contexts these words were acceptable.

Experiment 10 further investigated this generalization process and also attempted to assess exactly what type of distributional information was being used in both Experiment 8 and the Mintz (2002) study. The one to one correspondence between the first and last word in all the trigram languages was broken, but the individual bigram co-occurrences remained a reliable source of categorization information. In addition to this, the generalization test was expanded by testing the novel words in compatible and incompatible distributional contexts directly as well as testing them against the trained category words.

7.2. Experiment 9

7.2.1. Participants

Twenty-five University of York undergraduate and graduate students participated in this Experiment. Participants had not taken part in similar Experiments. One participant was excluded as they produced unusable card sorting results (unequal groups)[10]. The average age was 23.38 years; range was 19 to 42. There were 14 female and ten male participants in the sample. All included participants were native English speakers and were paid £4 or given course credit.

---

[10] All results were conducted with this participant excluded and included. There were no differences between the two sets of results.

7.2.2. Materials

The artificial language used in this study was based on the trigram distributional language in Experiment 8. However, the non-adjacent dependency between the first and third word was broken while the individual preceding and succeeding bigram information was retained. This led to a situation where the bigram dependencies were more reliable than the non-adjacent dependencies. The total amount of "frames" increased from two frames per category to four frames per category. As in Experiment 8, only one frame contained all eight category words; the remaining three frames contained only five of the category words, as can be seen in Table 7.3. Thus, the three critical category words for testing (X_6, X_7 and X_8) never occurred in the bigram contexts which were used during the test sessions. The frequency of each individual training sentence was reduced to ten occurrences; however, the critical test sentences for categorization (AX_6B, AX_7B, and AX_8B for Category 1) still occurred 20 times, as in Experiment 8.

Table 7.3 Summary of training and the second generalization test sentences in Experiment 9

Training			
Category A		Category B	
Complete	*Partial*	*Complete*	*Partial*
AX_1B	CX_1D	EX_9F	GX_9H
AX_2B	CX_2D	$EX_{10}F$	$GX_{10}H$
AX_3B	CX_3D	$EX_{11}F$	$GX_{11}H$
AX_4B	CX_4D	$EX_{12}F$	$GX_{12}H$
AX_5B	CX_5D	$EX_{13}F$	$GX_{13}H$
AX_6B	CX_1B	$EX_{14}F$	GX_9F
AX_7B	CX_2B	$EX_{15}F$	$GX_{10}F$
AX_8B	CX_3B	$EX_{16}F$	$GX_{11}F$
AX_1D	CX_4B	EX_9H	$GX_{12}F$
AX_2D	CX_5B	$EX_{10}H$	$GX_{13}F$
AX_3D		$EX_{11}H$	
AX_4D		$EX_{12}H$	
AX_5D		$EX_{13}H$	
Generalization Test 2			
Category A		Category B	
Compatible	*Incompatible*	*Compatible*	*Incompatible*
CY_1D	CY_4D	GY_4H	GY_1H
CY_2D	CY_5D	GY_5H	GY_2H
CY_3D	CY_6D	GY_6H	GY_3H

The test contrasts for the categorization and generalization test sessions were exactly the same as in Experiment 8. In a second generalization test session which followed immediately after the first generalization test session, the new category words in compatible and incompatible contexts were directly compared. Each new word compatible sentence was paired with every incompatible sentence, which created a total of 18 additional test items. The counterbalancing of these test items was exactly the same the categorization test session, which was described in detail in Experiment 8. These training items are detailed in Table 7.3.

In addition 16 category word cards were created for this Experiment in order to test participants' ability to explicitly sort the category words into their respective categories.

7.2.3. Procedure

The procedure was the same as in Experiment 8 with the addition of the third test session and the card sorting task at the end of the Experiment. The card sorting was the same task as in Experiments 1 through Experiment 7, in which the participants were given the 16 shuffled category word cards and simply had to sort the words into two equal piles according to which words the participants felt belonged together.

7.2.4. Results

All multiple *t* tests were Bonferonni corrected, with significance before correction as indicated for Experiment 8.

The dependent variable from the first test session with the remaining partial paradigm sentences was the number of times the participants correctly chose the compatible test item. Contrary to Experiment 8, there was no difference between the phonologically coherent and incoherent conditions in this Experiment, $t(22) = 1.01, p = .33$, M = 9.92 and M = 8.83, respectively. When compared against chance level of responding, the coherent and incoherent conditions showed no evidence of being able to differentiate the compatible and incompatible test items, $t(11) = 1.13, p = .28$ for the coherent condition and $t(11) = -.24, p = .82$ for the incoherent condition. These results provide no evidence that the category words were categorized into separate groups during exposure to the training language.

The dependent variable from the second test session was the amount of times the participant decided the sentence that contained the "new" category word was more similar to the artificial language than the sentence that contained the known category words. As in Experiment 8, only the data from the participants who were in the coherent condition were considered in this analysis. A repeated measures ANOVA was conducted with the status of the "new" words (compatible or incompatible) and the status of the trained category words (compatible or incompatible) as the within subjects factors. The order in which the test contrasts were presented (counterbalancing) was included as a dummy variable. There were no significant main effects or interactions; new word status by trained word status interaction: $F(1,22) = 2.06, p = .17$; all others $F < 1$.

The dependent variable from the sentences that directly compared the new category words in compatible and incompatible distributional contexts was the number of times the compatible test sentences were chosen. As with the previous analysis only participants in the coherent condition were considered. The coherent condition was therefore compared to chance level of responding, which was nine. Surprisingly, given the previous results, there was a small but significant effect of differentiating the compatible and incompatible test items over the level expected by chance, $t(11) = 2.46, p < .05$, M = 10.17. This indicated that some learning of the general phonological characteristics of the two categories occurred over the course of the Experiment and was expressed when directly comparing the phonologically consistent category words in compatible and incompatible distributional contexts.

The card sorting results support the slight indication of categorical learning found in the previous result. There was a significant difference between the phonologically coherent and incoherent conditions when looking at the distributional categories, $t(22) = 3.26, p < .01$ with the coherent condition sorting significantly more category words together than the incoherent condition; M = 6.0 and M = 4.58, respectively. Furthermore, the coherent condition was marginally significantly above the chance level of 5.09, $t(11) = 2.34, p = .08^{\dagger}$. The incoherent condition was significantly below what was expected by chance, $t(11) = -2.63, p < .05^{\dagger}$. This below chance performance was due to the phonological interference effect, which has been discussed extensively before; $t(11) = 1.73, p = .11$, M = 5.67 for the incoherent condition phonological groupings when compared to chance level. There was no difference between the two coherency groups when looking at the phonologically defined categories, $t(22) = .65, p = .52$, which indicated that the sorting in both groups may simply be due to the phonological and orthographic similarities within the phonologically coherent category words.

7.2.5. Discussion

There was no evidence that the participants were able to differentiate the compatible and incompatible test items, which indicated that no categorization occurred in this Experiment. Consistent with this finding, no generalization of the phonological structure to the novel category words occurred when compared to the trained category words. Interestingly, some knowledge of the general phonological structure associated with the bigram distributional contexts must have occurred as when faced with the choice between "new" category words in compatible and incompatible contexts the participants choose the compatible test sentence more often than the incompatible test sentences and more often than expected by chance.

This last finding indicated that the phonological characteristics of the categories may have been learnt and grouped together in a category, although the actual category words themselves were not categorized. As discussed in Chapter 2, the associations between the phonological cues and the distributional marker words should be stronger than the category words themselves, as the phonological characteristics were present during each category word-marker word pairing. It appeared that the training sessions allowed the general phonological characteristics to be learnt, but there was no conclusive evidence that the actual category words were learnt. However, prior exposure to the trained category words may have interfered with the application of the phonological cues to category membership in the first test session and the first generalization test. Only when confronted with the direct comparison of the phonologically consistent novel category words within compatible and incompatible context could this more abstract knowledge influence the decisional process. Alternatively, it could be that the abstract phonological cues influenced both the categorization and the generalization test, but that the application of this knowledge in the generalization test was not influenced by interference from prior exposure. This could have resulted in the impact of the phonological cues being greater with the novel category words, and slightly lesser with the trained category words.

It also appears that the knowledge of the phonological characteristics was able to influence performance in the card sorting task. The coherent condition sorted more category words than the incoherent condition and was above chance level, indicating that by the end of the experiment some categorical knowledge was present. However, this result may have been driven by the phonological cues themselves, as there was no difference between the coherent and incoherent groups when both conditions were coded based on the phonological categories.

Thothathiri and Snedeker (2005) conducted a similar study that investigated the use non-adjacent frame cues versus initial and ending bigram cues. The training stimuli were very similar to Experiment 8 stimuli, but there were only two category words. The test stimuli were also highly similar to the categorization test sentences used in Experiments 8 and 9 with the exception of "frame violation" test sentences. These sentences mixed the initial and ending bigrams from both the complete and partial category. In other words, the same manipulation that was used within the training stimuli of Experiment 9 was conducted during the test session in this study. The results indicated that responses to the grammatical and "frame violation" sentences were identical, providing evidence that learning from frames was not dependent on the non-adjacent dependencies and indicating that the bigram dependencies may be at the root of learning from trigram distributional cues. These results, although suggestive, failed to reach statistical significance, which was likely due to a very low number of participants. However, the converging evidence suggests that bigram learning at the very least plays a partial role in learning from non-adjacent dependencies.

One methodological issue that might have influenced the finding of categorical knowledge in Experiment 8, but insufficient evidence in Experiment 9 was the different number of participants between the two Experiments. Experiment 8 had 32 total participants, whereas Experiment 9 had only 24 participants. The original N number was increased above the level for Experiments 1 to 7 for Experiment 8 as there were 40 participants in the Mintz (2002) study; therefore an increase beyond 24 participants was decided. However, due to the increased strength of the experimental design, it was decided that 32 participants should be sufficient to show categorical learning in this new paradigm. Power analyses conducted on the results of Experiment 8 using the mean and standard deviation statistics indicated that this increased N number was perhaps unnecessary. For a 50% chance of rejecting the null hypothesis correctly only three participants would be needed per coherency condition for the Experiment 8 results. For an 80% chance of correctly rejecting the null hypothesis only six participants were predicted to be needed per condition. However, power analyses of Experiment 9 reveal a different pattern. There would need to be 64 participants in the entire Experiment for a 50% chance of a significant result, while 146 would be needed for an 80% chance. These results show that an increasing the numbers to 32 for Experiment 9 would not have only increased the probability of getting a true significant result from around 27% to about 32%. Increasing the N number to a level that would likely show a significant difference was not practical.

The overall conclusion from this Experiment was mixed in that the general phonological characteristics of the two categories and their connection to the embedded bigram cues was learned but the actual category words were not learned. The reason for the lack of evidence for category word learning may be due to the fact that the language structure of this Experiment was the most complex of all of the Experiments in this thesis. Future work could include replicating this Experiment with substantial amounts of further training – this would increase the co-occurrences of the bigram cues and category words which may lead to categorization of the phonological cues and the category words.

The trigram and bigram learning were directly compared to determine whether the categorization from the trigram cues was higher than the levels of performance from the bigram cues in the following analysis.

7.3. Experiment 8 and 9 combined analysis

7.3.1. Combined Results

In order to directly compare learning from bigram versus trigram distributional cues the two Experiments were analysed together. For the first test session, a two way ANOVA was conducted with coherency (coherent and incoherent) and distributional cue (trigram and bigram) as the between subjects factor. A dummy variable (counterbalance) was entered into the ANOVA as a between subjects factor. The dependent variable was the number of times the participants correctly chose the compatible test item.

The coherency main effect was significant, with the coherent condition choosing the compatible test sentences more often than the incoherent condition, $F(1,48) = 9.75, p < .005, \eta_p^2 = .17$; M = 10.46 and M = 8.45, respectively. There was no main effect indicating a difference between trigram and bigram distributional cues, $F < 1$, and no interaction between the two variables, $F(1,48) = 2.07, p = .16$.

As only the coherency group showed any categorical knowledge in Experiment 9, the performance in the first test session in the coherent conditions were compared across Experiments. There was no significant difference between the trigram and bigram distributional cues, $t(26) = 1.10, p = .28$; M = 11.0 and M = 9.92, respectively.

For the second test session, a mixed ANOVA was conducted with new word status (compatible and incompatible) and old word status (compatible and incompatible) as the within subjects factors and distributional cue (trigram and bigram) as the between subjects factor. A dummy variable (counterbalance) was entered into the ANOVA as a between subjects factor. The dependent variable was the number of times the participants correctly chose the new category word sentence.

The new word status main effect was significant, $F(1,24) = 7.35, p < .05, \eta_p^2 = .23$. As in Experiment 9, the participants were more likely to choose the new word sentences when they were in a compatible, rather than incompatible, distributional context, M = 2.20 for compatible context and M = 1.76 for incompatible context. There was no difference between the trigram and bigram distributional cue conditions, $F < 1$. There were no other significant main effects or interactions; old word status: $F(1,24) = 2.95, p = .07$; all others: $F < 1$.

7.4. General Discussion

From the results of the combined analysis there was no evidence that the trigram distributional language induced higher learning than the bigram distributional language. The lack of any interactions with the distributional cue factor indicated that the pattern of results was consistent across both Experiments. The non-significant difference in the coherent conditions across the trigram and bigram cues does not provide support to the argument put forward by Mintz (2002, 2003, & 2006) that trigrams are more useful than bigram distributional cues in the grammatical categorization process. However, the results also do not provide conclusive support to the idea that learning in trigram distributional cue languages was due to the combination of beginning and ending bigram cues, as learning from bigram cues was not above chance levels.

The case is different with the trigram language when the a_b non-adjacent dependencies were unbroken as the participants were able acquire significant knowledge about the two categories and were also able to generalize the phonological cues within the language to novel category words. However, when the a_b non-adjacent dependencies were broken leaving only the aX and Xb bigrams to provide reliable distributional cues, the participants were only able to learn the general phonological characteristics of the categories. There was no evidence of categorization of the category words themselves, but evidence of the ability to generalize beyond the category words to phonological consistent novel words was present. Interestingly, there were semi-consistent trigram cues within the "bigram" language. The trigram dependencies between the first and third word were not perfectly predictive but still could have been a semi-reliable cue. The situation of trigram dependencies being partially predictive in this manner parallels trigrams found in natural language; "you_it" is a common example of a natural trigram, but it is certainly not true that each time the word "you" is heard, the word "it" follows after an intervening word. This result indicated that the trigrams, in this experimental situation, have to be perfectly predictive for categorization to occur.

Although only the trigram distributional cues were able to induce true grammatical categorization, the performance on the critical categorization task was not significantly above the level achieved by the participants who received only consistent bigram cues. These results indicated that learning from bigram cues may have simply been slower than learning from trigram cues; this idea is supported by the evidence that the participants who received bigram distributional cues were able to learn the general phonological characteristics. As discussed in previous chapters, the general phonological characteristics are predicted to be easier to learn than the category words themselves; with more exposure to the training language, learning of the category words from these bigram distributional cues would be expected. Alternatively, it could be that the complexity of the artificial language may have also interfered with the learning of the co-occurring bigram cues and the category words. It should be pointed out that the category words used during the test sessions occurred half as often as the remaining five category words; this may not have been enough to categorize all eight category words together. Future work could investigate simpler designs that equate the frequency of all category words.

More generally, the results of Experiment 8 supported the general finding of Experiments 1 to 4 that the combination of phonological and distributional cues aids the categorization process. The results of Experiment 8 allow the conclusion that phonological cues aid the categorization process even with more informative trigram distributional cues. This does not rule out the possibility that learning from trigram cues alone could occur; it has been found in at least two separate investigations (Mintz, 2002; Thothathiri & Snedeker, 2005). The current lack of evidence for categorical knowledge in the incoherent condition may have been due to a lower frequency of the category words than has been found to be necessary for distributional cue only learning. Specifically, the studies that have found learning from only distributional cues have had much greater frequency than in the current Experiment. As frequency was not manipulated in these Experiments, the role of frequency with trigram distributional cues was not specifically tested.

One issue that has not yet been mentioned is that the general frequency of trigram frame information within natural languages is relatively infrequent. For instance, for the frame "you X it" the likelihood of the frame "you_it" occurring next time the initial word "you" occurs is very small. In a sense, the experiments displaying learning from frequently occurring trigram cues are not simulating natural language learning, as the "frequent frames" do not occur within natural language anywhere near the frequency found in the artificial language learning Experiments. An interesting future test of the frequent frames idea is to insert the frame cues within an artificial language that has the same distribution of frame cues as is found in natural language to determine whether these frame cues induce categorization under more natural circumstances.

8. General Discussion

This thesis was an investigation into the process of grammatical categorization. Much research has investigated how distributional cues can aid the process of grammatical category acquisition (e.g., Braine, 1966; Green, 1979; Maratsos & Chalkley, 1980; Mintz, 2002, 2003, 2006; Mintz, Newport, & Bever, 2002; Monaghan, Chater, & Christiansen, 2005; Monaghan & Christiansen, 2004). More recently, the role of phonological cues in differentiating grammatical categories has come into focus (e.g., Cassidy & Kelly, 1991, 2001; Kelly, 1988, 1992, 1996; Monaghan & Christiansen, 2004; Shi, Morgan, & Allopenna, 1998; Shi, Werker, & Morgan, 1999). Initial work by Monaghan et al. (2005) investigated how distributional and phonological cues could combine to provide more accurate grammatical categorization. It was found in corpus work and a written artificial language experiment that distributional cues aided the categorization of high frequency words more than low frequency words, but the opposite pattern was found for phonological cues; low frequency words were able to utilise the phonological cues that define grammatical categories to a larger extent than high frequency words.

This combination of phonological and distributional cues was further investigated in St. Clair and Monaghan (2005) with an auditory artificial language. Interestingly, with preceding distributional cues no learning occurred at all, but with succeeding distributional and phonological cues categorization of the two word classes occurred. The differing ability of preceding and succeeding cues to aid the grammatical categorization process was fully investigated in this thesis. The combination of preceding and succeeding cues with phonological information and how this influenced the grammatical categorization process was also a main theme of this thesis.

While investigating these two related topics, the further hypothesis that prior linguistic experience influences the effectiveness of the distributional cues was developed, as it was noticed that the phonological form of the distributional cues influenced how easily participants were able to acquire categorical knowledge. Phonological similarity of the distributional cue words to typical English preceding and succeeding grammatical category cues seemed to modulate these differences. This was also fully investigated within this thesis.

Different accounts of grammatical categorization have hypothetically suggested that distributional frame information (aXb) may be key in the grammatical categorization process and operate independently of any other linguistic cues (Mintz, 2002, 2003, 2006). This thesis investigated this possibility by analysing the two possibilities of learning from a_b frames which are that learning occurs from the non-adjacent dependency between *a* and *b* or that learning occurs through the combined influence of the aX and Xb bigrams. Corpus analysis and experimental work have investigated these two accounts of learning from distributional frames as well as directly comparing frame learning with and without additional phonological cues.

Before fully discussing all of these issues in turn, a brief review of the work and results presented in each chapter will be outlined.

8.1. Thesis summary

8.1.1. Chapter 2

Chapter 2 consisted of two Experiments which investigated learning from bigram distributional information with and without additional phonological cues. Experiment 1 investigated preceding distributional cues while Experiment 2 investigated succeeding distributional information. It was found that succeeding bigram distributional cues, but not preceding cues, could induce grammatical categorization. However, both types of distributional cue failed when presented without additional phonological cues. Although the difference between Experiment 1 and 2 was not significant, two potential explanations for learning with succeeding, but not preceding cues were discussed. The R-W model predicted that associations in the *aAbB* preceding cue language would take longer to form; therefore the language should take longer to learn. However, in the *AaBb* succeeding cue language the associations between the category words and marker word will form relatively quickly, thus leading to substantial learning quickly. Additionally, the specific marker words (*alt* and *ong*) may have been treated as inflections by the participants; *alt* and *ong* may have seemed more natural in the succeeding location to category words, rather than in the preceding location (*dreng-ong* versus *ong-dreng*). This may have increased the salience of the marker words in the succeeding location. This hypothesis was tested in Chapter 3.

8.1.2. Chapter 3

Chapter 3 consisted of two Experiments. These Experiments were essentially replications of Experiments 1 and 2 using *cvc* marker words. It was thought that the phonological form of these marker words would equate any possible salience increases the phonological form might provide for either preceding or succeeding cues. The preceding cue Experiment found substantial learning in the phonologically coherent condition, as well as with high frequency words in the incoherent condition. This provided support that word categorization can occur from only distributional cues with high frequency words but that additional phonological cues are necessary for lower frequency words. Experiment 4 also found substantial learning when both types of cues were available and the conclusion of high frequency word categorization with only distributional cues from Experiment 3 was supported, although the result was only marginally above chance level. Additionally, succeeding cues induced higher grammatical category learning than preceding cues, providing conclusive evidence of the succeeding cue advantage. However, this result seemed to be driven by the phonological coherent condition; the hypothesis put forward was that it was not only the category words that formed associations with the marker words, but the abstract phonological characteristics as well. Indeed, the phonological characteristics were predicted to have higher associative strength as these cues were present in all category/marker word pairings, while pairings of individual category words and marker words were relatively infrequent by comparison.

8.1.3. Chapter 4

Chapter 4 was an investigation into a hypothesis that was derived from the comparison of Experiment 1 and 3, which had preceding cues. With these preceding cues, no learning was found with *vc/vcc* marker words but there was learning with *cvc* marker words. The possibility that *vc/vcc* marker words were congruent in a succeeding location, but incongruent in a preceding location, due to similarity with succeeding cues within English was investigated. All prefixes and suffixes in the English language were analysed both in terms of the general phonological form and also in terms of the specific phonological details of the initial and ending phonemes. The general results indicated that *cvc* and *cv* affixes were likely to be prefixes and *vc* affixes were likely to be suffixes.

This finding was backed up by an analysis of the more general phonological form of beginning and ending phoneme types (C_C, C_V, and V_C). Overall, this analysis provided an explanation for the differing amounts of learning in Experiments 1 to 4. The *vc*/*vcc* succeeding marker words may have had increased salience due to similarity with suffixes, whereas the *cvc* preceding marker words may have had increased salience due to similarity with prefixes. Additionally, this analysis provided the hypothesis that *cv* marker words may induce increased learning with preceding marker words due to similarity with prefixes.

8.1.4. Chapter 5

Chapter 5 further investigated the hypothesis that the salience of the marker word may increase depending on phonological similarity to English preceding and succeeding distributional cues (specifically, prefixes and suffixes). There were three Experiments included in this Chapter which directly compared the succeeding and preceding marker word locations with categories that also contained phonological cues. From the results of Chapter 4 it was predicted that *cvc* and *cv* marker words would increase learning in the preceding location and that the *vc* marker words would increase learning in the succeeding location. The results partially supported the hypothesis; *cvc* marker words increased learning in the preceding location and seemed to depress learning in the succeeding location, which acted to equate learning in the two conditions. There was similarly no difference between the two conditions in the *vc* Experiment. Only in the *cv* Experiment was there a significant succeeding word advantage. These results did not confirm the prefix/suffix similarity hypothesis. The depressed learning with succeeding cues in the *cvc* Experiment seemed to suggest that *cvc* marker words were particularly inconsistent in the succeeding location. The possibility that was proposed that preceding cues in natural language are usually high frequency functions words, but that succeeding cues are usually in the form of the suffixes. Learning was increased in the preceding location as the *cvc* marker words were the most “wordlike” of the three marker word phonological structures, as these marker words could phonotactically have been considered a word and also the *cvc* structure was the most common structure in preceding function word cues. The succeeding cue learning was increased in the *cv* Experiment as these marker words were less likely to have been considered as a pseudoword and therefore potentially more similar to bound morphemes.

8.1.5. Chapter 6

Chapter 6 further investigated the possibility that preceding word cues are a natural cue for grammatical categorization. The 45 most frequent preceding, succeeding and frame cues (i.e., aX, Xb, and aXb) were investigated and measures of accuracy and completeness were taken on how well the distributional categories were grouped according to grammatical category. It was found that the aXb grouping were the most accurate at grouping words of the same grammatical category. The preceding cue groupings were less accurate than the aXb groupings, but more accurate than the succeeding cue groupings. In terms of how well the distributional category captured all words of the same grammatical category together (completeness), it was generally found that the Xb succeeding word groupings were significantly less complete than the aXb and aX groupings. Although this result seemed to suggest that the aXb groupings were far superior to either of the bigram groupings, the amount of words classified by each analysis indicated that the aX and Xb analyses were able to group up to seven times the amount of words as the aXb trigram groupings. Thus, the aX groupings proved to be the best cue from the three investigated as the accuracy was acceptably high, completeness score was the best of all three analyses, and this analysis also categorizing the highest amount of words. This result supports the account of the results in Chapter 5 detailed above.

8.1.6. Chapter 7

Chapter 7 investigated learning from distributional frame (aXb) information that had been found in previous research (Mintz, 2002). Specifically Experiments 8 and 9 aimed to determine whether the learning that results from aXb frames is due to the a_b non-adjacent dependency or due to the combination of the beginning and ending bigram (e.g., aX and Xb). Additionally, Experiments 8 and 9 included both a phonologically coherent and incoherent condition to attempt to determine whether additional phonological cues to category membership would aid learning from frame information, or whether the frame distributional information was sufficient for categorization without additional cues. Experiment 8 had frame (aXb) distributional cues and it was found that participants not only categorized the words within the language, but (in the coherent condition) could extend their knowledge of the phonological characteristics of the categories to novel category words that were phonological similar to the trained category words. When the non-adjacent dependency between the first and third word was broken, but the usefulness of the individual preceding and succeeding bigrams was retained (Experiment 9) there was no evidence of categorization of the trained category words; however, there was evidence of the ability to apply the phonological characteristics of the categories to the novel, but phonologically consistent category words. This indicated that the phonological characteristics were grouped into categories, but either the language design was too complicated or there was insufficient training for the category words themselves to be categorized.

8.2. Phonological and Distributional cues

All Experiments within this thesis found evidence of grammatical categorization only when distributional and phonological cues were combined, with the exception of high frequency category words. The theory put forward by Monaghan and Christiansen (2005) predicted that there would be a difference between high and low frequency words such that phonological cues would be necessary for the categorization of low frequency words, but distributional information should be sufficient for the grammatical categorization given the category words of sufficiently high frequency. This was the pattern of results found in Experiments 3 and 4; learning of high frequency category words in the incoherent condition of Experiment 3 was no different than learning in the coherent condition. Similarly, learning of the incoherent high frequency words in Experiment 4 was marginally above chance level.

There was no evidence of increased learning in the incoherent high frequency words in Experiments 1 and 2, but this was due to the presence of the syntax-incompatible test sentences. These test sentences had a repeated marker word (*aAaB*), and thus an obvious syntax violation. The evidence indicated that the incoherent participants focused on whether the test sentences had one or two marker words which interfered with their ability to respond according to any categorical knowledge they may have acquired. The results of Experiment 3 and 4 confirmed this hypothesis, as the categorization of the high frequency words increased to approximately the level of the coherent condition with these syntax-incompatible sentences removed.

Experiments 1 to 4 used bigram distributional cues; the evidence so far has only investigated the combination of bigram distributional and phonological cues. Experiment 8 extended this finding to trigram, or frame, distributional cues. It was found that only when the trigram distributional cues were combined with additional phonological information was any evidence of categorization found. The experimental design of Experiment 8 did not manipulate frequency, but it was clear from the findings that the frequency of the category words was not sufficient to induce categorization from distributional cues alone. Previous findings from distributional cues alone (Mintz, 2002; Thothathiri & Snedeker, 2005) had category words with higher frequencies than in Experiment 8. Nonetheless, the finding that phonological cues aid the acquisition of word classes when combined with trigram *and* bigram distributional cues provides substantial support for the idea that a combination of linguistic cues is preferential to individual cues.

The underlying implications for grammatical categorization is that distributional cues alone may indeed be sufficient for categorizing the most frequent words in a child's linguistic environment. For instance, within the Anne corpus (which has 95255 total word tokens) the phrase "to do" occurs 156 times, "to get" 96 times, "to go" 177 times, "to have" 113 times and "to put" 139 times. The relative frequent co-occurrences of the verbs do, get, go, have and put with the high frequency function word "to" may indeed be sufficient for categorization with the majority of words that occur after "to"; in other words, verbs.

However, lower frequency verbs, such as peel, muddle, live, and kneel only occurred once after the word "to" in the Anne corpus. These verbs (along with numerous others that did not occur at all in the Anne corpus) do not co-occur frequently enough with the distributional cues to be correctly and easily categorized in the verb category. In these cases, the use of the phonological correlates that are known to characterize grammatical categories would be invaluable (Kelly, 1992; Monaghan, Chater, & Christiansen, 2005; St. Clair & Monaghan, 2005). Experimental findings indicate that the phonological characteristics of words do indeed guide the grammatical categorization process, as children are more likely to decide a novel word is referring to an object if this novel word has the general phonological characteristics found within the noun category (Cassidy & Kelly, 1991, 2001). These five Experiments highlight the importance of phonological cues, specifically for the categorization of low frequency words.

This work adds to the growing literature advocating the importance of combined phonological and distributional cues in learning many aspects of language during initial language acquisition (Christiansen & Monaghan, 2006; Gerken, 1996; Monaghan, Chater, & Christiansen, 2005; Monaghan, Onnis, Merkx, & Chater, Submitted; Reali, Christiansen, & Monaghan, 2003). In particular, Gerken (1996) advocated that the combination of these two types of cues may aid in the initial segmentation of words, in the categorization of words into distinct grammatical categories, and in higher order syntax learning. This thesis certainly provides evidence for the combined usefulness of phonological and distributional cues in the categorization process while complementary work supports the ideas of segmentation (e.g., Bonatti, Peña, Nespor, & Mehler, 2005; Onnis, Monaghan, Richmond, & Chater, 2005) and syntactic structure (e.g., Gerken, Jusczyk, & Mandel, 1994) similarly benefiting from these two combined cues.

8.3. Succeeding cue advantage

Increased categorization from succeeding distributional cues has been one of the main findings of this thesis. The differential effect of marker word location was first found in St. Clair and Monaghan (2005). The finding of more learning with succeeding marker words has been replicated four times within this thesis, although the phonological form of the marker word also exerts an influence on learning from preceding and succeeding distributional cues. Differences due to the marker word phonology will be fully investigated in the next section.

As discussed in previous Chapters the Rescorla-Wagner model of associative learning (Rescorla & Wagner, 1972) predicts that there should be higher learning from succeeding cues than from preceding cues. This is illustrated in Figure 8.1; the associative strength between the category words and the succeeding distributional cue will be higher as each time you hear a particular category word, the relevant marker word is always the succeeding word; thus, the contingency between these cues will be very high, which increases to the asymptote level. However, the situation is different with preceding cues: each time you hear a preceding marker word, six possible category words could be the next word; this situation leads to a much lower contingency with the category words, which leads to a lower asymptote (Pearce, 1997). Therefore, although the number of pairings of the category and marker words was equated in the preceding and succeeding Experiments, the associations between the marker and category words will be stronger with succeeding cues.

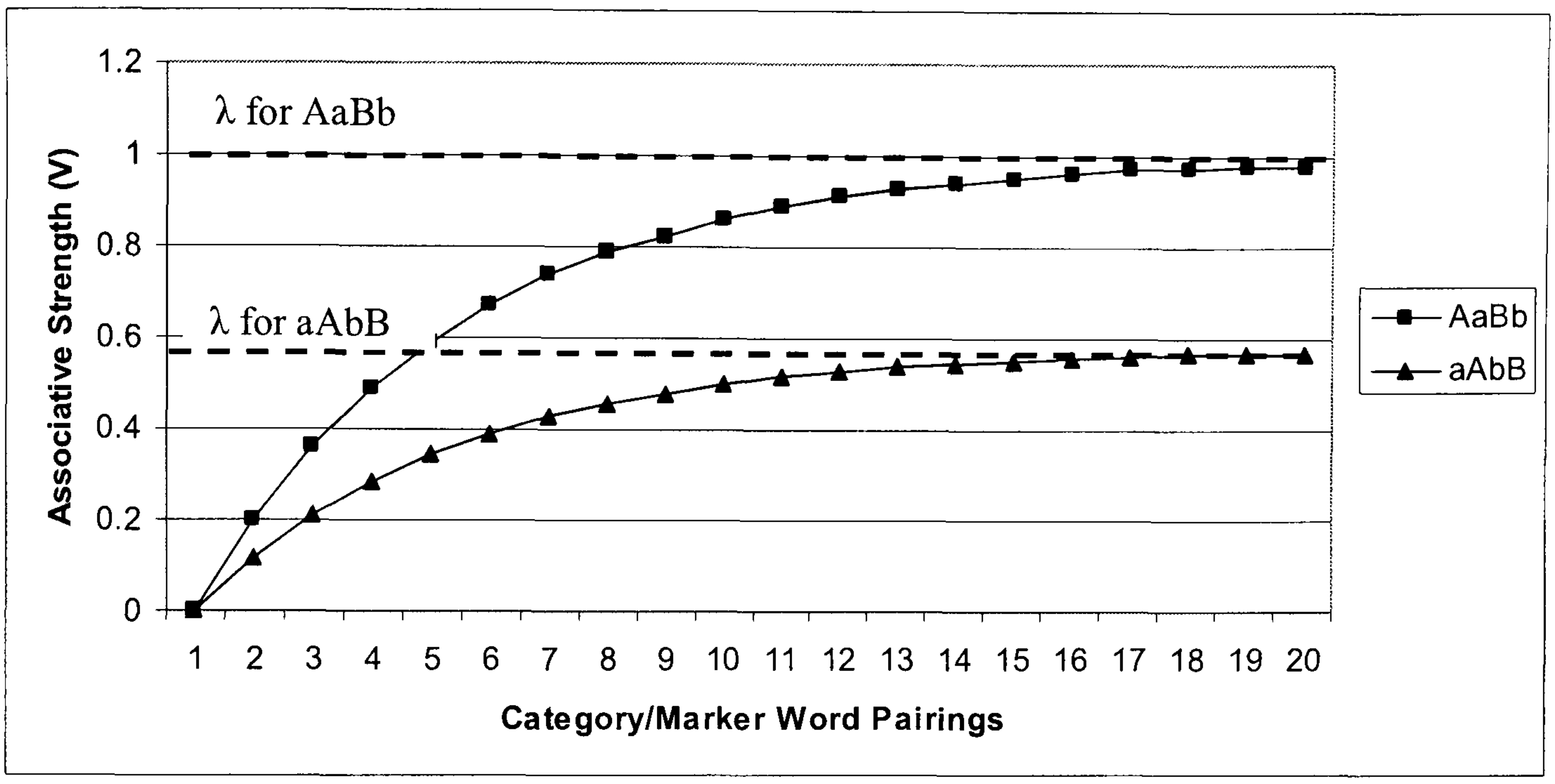

Figure 8.1 Theoretical associative strength increases with preceding (aAbB) and succeeding (AaBb) distributional cues.

Interestingly, the difference between preceding and succeeding marker words was only expressed with co-occurring phonological cues to category membership. In other words, there was very little difference within the incoherent conditions but a large difference in the coherent conditions. This indicated that the phonological cues within the category words themselves were a key factor in the succeeding cue advantage.

As discussed in Chapter 3 and illustrated in Figure 8.2 this result was explained by the idea that it was not only the category words but also the abstract shared phonological features within the category words that were forming associations with the distributional cues that varied in associative strength depending on the position of the marker words. Additionally, it was predicted that the associations between the abstract phonological features and the marker words would be *higher* than the category words themselves. This was due to the fact that during every A_i and a pairing, the phonological cues were constant even though the individual category words varied. The associations between the actual category and marker words should be relatively weaker in comparison to the associations between the phonological cues and marker words.

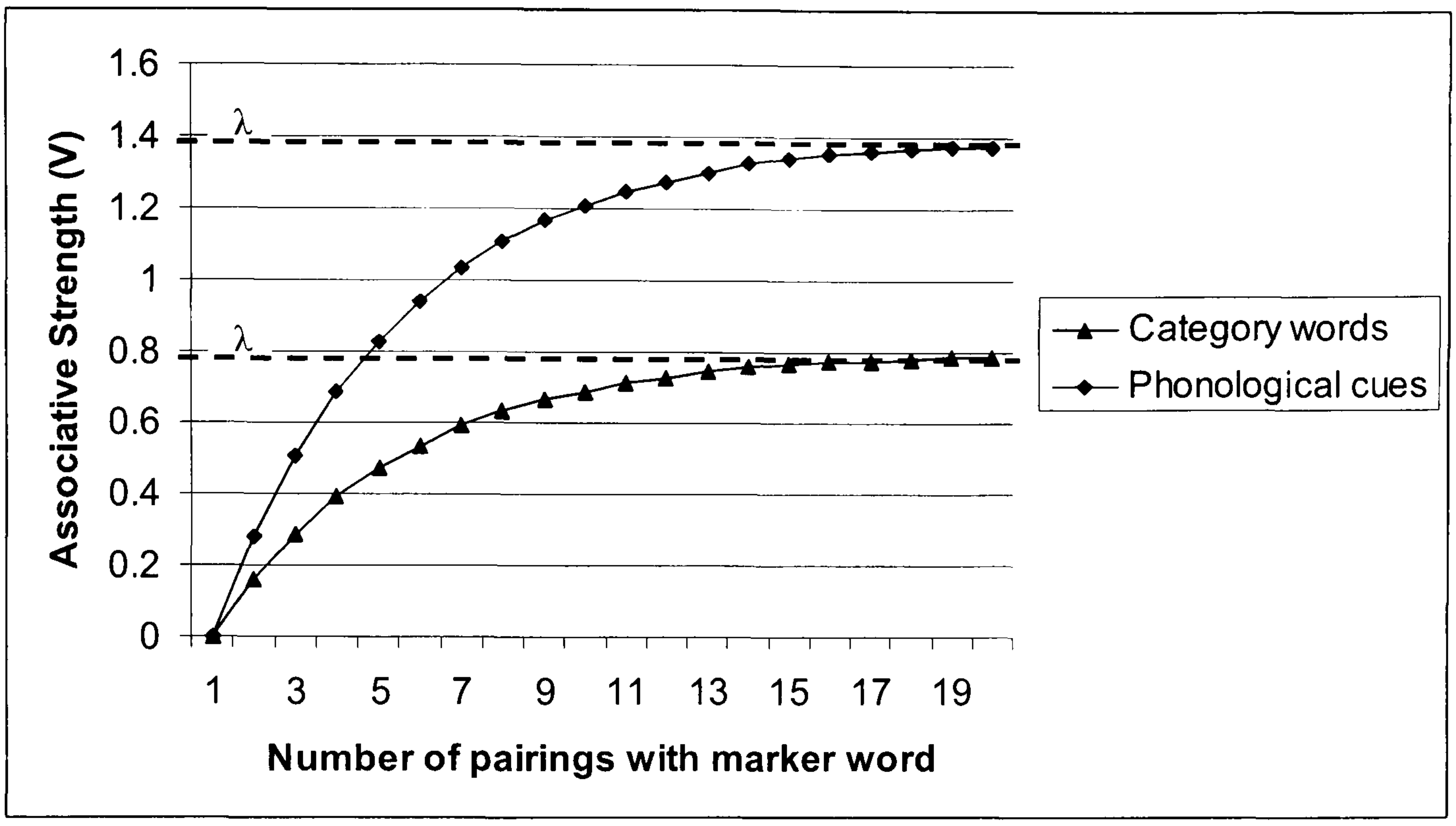

Figure 8.2 Theoretical overall associative strength increases between the marker word distributional cues and the phonological cues and category words.

Support for this theory came from Experiment 9. In this Experiment there was no definitive evidence of *word* categorization, but there was evidence that the phonological characteristics were grouped together and associated with the distributional cues. This finding is perfectly in line with the above account of categorical learning within the coherent conditions. As the associations are stronger between the abstract phonological cues and the marker words, it is not surprising that the phonological characteristics may become reliably associated with the distributional cues even when the category words fail to achieve the required levels of associative strength for substantial learning.

Experiment 9 also disproves an alternative hypothesis to the premise that the category words are grouped together by the combination of distributional and phonological cues. The alternative hypothesis states that the actual category words themselves were not being categorized; rather the associations between abstract phonological cues and distributional cues were being learnt and this knowledge was guiding the responses during the test sessions. Thus, when faced with the test sentences, the participants simply applied their knowledge of the phonological features that should "go" with the distributional cues when deciding on individual category-marker word pairings. However, the results of Experiment 9 indicate that when the phonological cues alone are reliably associated with the distributional cues this process only worked for novel category words. The abstract phonological cues can be applied to novel category words, but there was no evidence these phonological cues were applied to the category words that the participants heard during the training sessions. This result clearly indicated that the categorization from phonological and distributional cues found in the previous Experiments was actual *word* categorization, not the application of knowledge of the abstract phonological structure of the categories.

As mentioned in Chapter 3, there was a potential difference between the *aAbB* and *AaBb* languages and the ability to correctly sort the category words during the card sorting task. The fact that the category words were predictive of the succeeding marker word in the *AaBb* language was thought to aid in the ability to correctly group the category words together when separated from the marker word. With the *aAbB* language, it was the marker words that were predictive of the following category words; this situation would perhaps not lead to an ability to correctly sort the category words together *without* the co-occurring marker word. However, when analysing all coherent participants there was no significant difference between the *aAbB* and *AaBb* language structure conditions, $t(118) = -.58, p = .56$, M = 4.37 for *aAbB* and M = 4.48 for *AaBb*. With all of the data combined both conditions were above chance level, $t(59) = 3.20, p < .005$ for *aAbB*, $t(59) = 4.05, p < .005$ for *AaBb*. The non-significant difference may be due to the partially predictive succeeding marker word that occurs in half of the A_i presentations (the *Ab* dependency in the *aAbB* language). Another possibility is that through the shared association with the marker words, whether in the *aAbB* or *AaBb* languages, the category words themselves developed associative links and it was these associative links that guided the categorization found within the card sorting task. The shared phonological cues could also potentially aid the forming of associative links between the category words. This possibility is highly likely; however, the evidence for these inter-category associative links was not conclusive and further research would have to investigate these potential links.

This associative learning account of grammatical categorization is very similar to the feature theory of general categorization detailed in Chapter 1. Categorization according to the feature theory occurs based on the shared features of category members

(Pearce, 1997). In the case of the grammatical categorization within these studies, the shared phonological and distributional cues were associated together to form the two categories. The higher associative strength with succeeding cues led to generally higher levels of categorization than with preceding distributional cues. Interestingly, the feature theory predicts that the categories should be hard to form with only distributional cues; the only shared feature in this case is the distributional cue. However, categorization should progress much faster when there are concurrent phonological cues within the category members. This is precisely what was found; categorization only occurred without shared phonological cues when the frequency of category/marker word pairings was high enough to allow the association between the defining features to form a category of these high frequency words.

The succeeding cue advantage can be seen within natural language; specifically in the phonological correlates with grammatical gender. Within French, *-ais* and *–oi* tend to indicate masculinity and *–ssion* and *–stion* femininity; within Hebrew /a/ and /t/ ending phonemes indicate femininity; within Russian *–a* ending nouns are generally feminine, *-o* ending nouns are generally gender neutral and masculine nouns tend to end with a consonant; within German, masculine nouns are more often monosyllabic and end with a fricative and /t/ phoneme combination while feminine nouns tend to end in vowels (Kelly, 1996). Across most languages which have grammatical gender, it seems a common feature to have phonological suffixing cues that provide an indication the status of the gender.

Frigo and McDonald (1998) directly investigated prefixing and suffixing cues within a word categorization paradigm. They integrated preceding and succeeding "phonological markers" into their category words and also gave a succeeding word cue to their participants; there was an additional succeeding word cue to indicate categorical membership as well. Only when both prefix and suffix cues were present within the category words was full categorization found. However, a similar study that used only suffixing within word cues found that categorization did occur when these suffix cues were combined with succeeding word cues (Brooks, Braine, Catalano, Brody, & Sudhalter, 1993). The main difference between these studies was the phonological form of the suffix and prefix cues; Frigo and McDonald used only *cvc*/*ccvc* prefixes and suffixes, whereas Brooks et al. used suffixes that were either a single vowel or of the *vc* phonological form. Both of these studies demonstrates learning from succeeding cues, as learning from the succeeding word cues was essential in both studies, but indicates that the phonological form of suffix cues may influence learning ability. These findings support the succeeding cue learning advantage while also providing support for the preceding word/succeeding affix ideas discussed within the next section.

8.4. Preceding word and succeeding affix

The result that *vc/vcc* marker words induced categorization with succeeding cues but not preceding cues prompted the hypothesis that the *vc/vcc* phonological form was perhaps more natural in the succeeding location. The results from Chapter 3 with *cvc* marker words indicated that learning occurred with both preceding and succeeding cues, although the succeeding cues induced more categorical knowledge. These results provided further support that perhaps the *vc/vcc* marker words were dissimilar to naturally occurring preceding cues, and thus the salience of these preceding cues was lower than with the potentially more neutral *cvc* marker words. This prototype hypothesis originated in the associative learning literature, in which categories of objects are learnt by comparing the features of new exemplars to a general "prototype" of the category (Pearce, 1997). Chapters 4 and 5 investigated whether similarity to the phonological form of naturally occurring English preceding and succeeding cues could predict learning rates with different marker words. The saliency of the marker word cues was predicted to increase or decrease depending on phonological similarity to prefixes and suffixes; increased salience would translate either into higher associative strength and higher learning, in the case of succeeding cues, or increased rate of learning, in the case of preceding cues.

The results of Experiments 5, 6 and 7 did not precisely conform to the expectations from the prototype hypothesis; however, the results did fit the pattern that would be expected if the prototypical preceding cue was a high frequency *word* cue and the prototypical succeeding cue was a *suffix* cue. A general tendency for languages to have more suffixing than prefixing has been observed and documented (see Cutler, Hawkins, & Gilligan, 1985 for a review). A processing account of this suffixing preference was proposed by Cutler et al., such that stems are processed before affixes, therefore affixes at the end of words would facilitate this process. For example, with the word+suffix *walked* the stem word *walk* would be recognized just as easily as when there is no suffix. However, prefixed words such as *undone* have a delay in processing as time for processing the prefix *un* will delay the uniqueness point of the word *done*. Additionally, resources would initially go to searching for stems that have the beginning phonemes of the prefix, which would further delay processing the stem word *done*.

This processing account is perfectly compatible with the associative learning idea that learning from succeeding cues is easier than learning from preceding cues. The Cutler et al. (1985) processing account indicated that affixes were processed entirely separately from the stem; this separate processing would be necessary if affixes were to be used as distributional cues in the grammatical categorization process. The separate processing of suffixes and root words (and the order of processing) would allow the associations between the grammatical categories of the root words and suffix cues to form. Thus, the suffixing preference inherently indicates that suffixes will be a more reliable affix cue than prefixes. In regards to the *cv* marker words (and during the second test session with *vc* marker words), the succeeding cues induced higher learning than the preceding cues due to the fact that the *cv* (and *vc*) phonological form was more similar to bound morphemes than individual words (in particular, function words).

The results of Chapter 6 supported the theory that preceding word cues are more useful cues to grammatical categorization than succeeding word cues. This was shown with both accuracy and completeness measures; both types of bigram cue categorized the same amount of total words but the preceding word cues were significantly more consistent in terms of grammatical categories. Additionally, Bauer (2003a) documented evidence of a change in the prefixes in Old English to modern English, such that many prefixes have disappeared or changed into forms that can function as independent words. Specifically, Bauer notes that English has "move[ed] away from the obligatorily bound status of prefixes towards a situation where the elements which are added to the left-hand edge of English words are in themselves more wordlike … movement away from prefixation and towards something more like compounding" (2003a, p. 37). This finding supports the preceding word/succeeding affix hypothesis by demonstrating that English prefixing is now uncommon and that preceding cues have a word like structure.

Generally, preceding word cues are in the form of high frequency function words. As mentioned in Chapter 4, there is evidence from patients that function words and affixes (specifically suffixes) induce the same types of errors, which indicate that perhaps these two types of linguistic cues perform similar grammatical functions (Patterson, 1980). The possibility that function words and affixes might be processed in a similar manner is consistent with the preceding word/succeeding affix cue hypothesis.

The results of Chapter 5 generally conform to this hypothesis; there is higher learning from "wordlike" preceding marker words while higher learning is found in "affix-like" succeeding marker words. However, the lack of significant differences between the different types of preceding cues suggested these general tendencies in English did not influence learning from preceding cues to a great extent. Overall, the results indicated that succeeding cues were more prone to influences from the normal patterns in English. While there was no specific evidence that the *cv* phonological form should be particularly similar to suffixes, this particular phonological form had fewer of the phonotactic characteristics that make up individual words. In particular, the ending vowels were not consistent with the ending vowels found in all suffixes investigated in Chapter 4. One possibility was that these particular marker words were more likely to be treated as bound morphemes as these marker words could not stand alone as word cues; this would increase salience (and learning) in the succeeding but not preceding location. Another possibility, as discussed in Chapter 5, was that the lack of congruence in the ending lax vowels within the *cv* marker words may have increased the salience of these marker words and led to the larger differences in learning between the preceding and succeeding conditions. The learning in the preceding *cv* condition was not different to the *cvc* and *vc* conditions, but the learning in the succeeding condition was significantly higher than the *cvc* succeeding cue condition. This would be the results expected by the R-W model if the salience of these incongruent marker words was increased in both conditions. However, it is possible that both explanations hold some validity and it was a combination of these two factors that influenced the results.

The difference between the coherent condition in Experiment 3, which had preceding *cvc* marker words and the corresponding condition in Experiment 5 was unexpected. As mentioned previously, participants in Experiment 5 scored approximately eight more test items correctly than the participants in Experiment 3. Although the exact reason for this difference is unknown, the card sorting results appeared to indicate that the results from the Experiment 5 replication were more representative of learning from *cvc* preceding cues. Participants in the coherent condition of Experiment 3 had a much lower average score in the card sorting task than any of the other Experiments, while the participants in the corresponding Experiment 5 condition were consistent with all other Experiments. Additionally, the succeeding cue condition of Experiment 5 almost exactly replicated the result found in the corresponding coherent condition of Experiment 4.

This conclusion may call into question the significantly higher learning found with succeeding cues in the combined analyses of Experiment 3 and 4. However, this does not invalidate the overall finding that succeeding cues induce higher categorization than preceding cues, as the significantly higher learning with succeeding cues was replicated in the second test session of Experiment 6 and in the combined data from both test sessions of Experiment 7. The finding of categorization within the high frequency words of the incoherent condition in Experiment 3 may also be questioned. However, the participants in the incoherent condition had card sorting results that were at expected levels from all other similar Experiments, which indicated that the unknown extraneous variable in Experiment 3 may have only influenced participants who received both phonological and distributional cues. Additionally, evidence of marginally significant word categorization with the high frequency words within the incoherent condition of Experiment 4 provided support that the conclusions reached in Experiment 3 were valid.

Overall, the conclusion from these findings is that prior linguistic experience influences the effectiveness of preceding and succeeding distributional cues in the grammatical categorization process. This can be equated to stimulus generalization studies conducted with pigeons; when trained to respond to a tone of 1200 Hz, then tested on tones varying from 400 to 2000 Hz the pigeon will respond optimally to the 1200 Hz tone, but will still show high levels of responding to 800 and 1600 Hz tones (Moore, 1973). In other words, the pigeon can generalize to similar sounding tones, but responding decreases the further away the test stimulus deviates from the "prototype" stimulus. These findings show the same process for linguistic stimuli. The closer the distributional cues within the artificial language approximated the natural preceding and succeeding distributional cues within English, the higher the salience of the preceding and succeeding marker words became.

The effectiveness of succeeding cues seemed especially susceptible to increases (or decreases) in categorical learning. Increasing the saliency of the preceding cues would increase the rate of associative learning between the preceding cue and the category words; however, increasing the saliency of the succeeding cues would increase the overall level of associative strength between the succeeding cue and the category words. When considering the differential effects of increased salience on preceding and succeeding cues, it is perhaps not surprising that prior linguistic experience has a larger influence with succeeding cues than with preceding cues. An overall increase in the associative strength would lead to higher learning, whereas the increased learning rate with preceding cues would lead to faster learning but overall the same associative strength levels. This finding may not relate directly to initial word categorization as the knowledge young language learners have about typical preceding and succeeding cues may be limited. However, the situation will change with increasing linguistic knowledge as the most useful distributional cues will increase in saliency, which in turn aids the learning and categorization processes.

These results allow further discussion of the different finding of Frigo and McDonald (1998) and Brooks et al. (1993). As mentioned in the previous section, the interword prefix and suffix cues used within the Frigo and McDonald study were *cvc*/*ccvc* while the suffix cues used with the Brooks et al. (1993) study were either a simple vowel or *vc*. These results seem to corroborate the findings in this thesis by pointing to the idea that learning from succeeding cues is aided by marker elements that are consistent with the general phonological form of suffixes. The fact that learning from prefixes did not occur in the Frigo and McDonald study was curious, as the phonological form of the preceding cues was more consistent with the *cvc* marker words used within this thesis. The answer may be due to the fact that the preceding cues in Frigo and McDonald were used as prefixes, and not as preceding word cues. Intriguingly, there seems to be an above chance trends with the prefix conditions in the Frigo and McDonald data, whereas the suffix only condition did not show any trend of differing from chance level.

The combined consideration of higher learning with succeeding cues and the suffixing preference across languages raises an interesting possibility with regard to the usefulness of the suffixing preference in higher order language structure learning.

Experiment 7 indicated that learning was higher with succeeding cues when the marker word was similar in phonological form to suffixes, which indicated that suffixes may be a highly useful cue for grammatical categorization within natural language. The increased ability to learn language structure from suffix information rather than prefix information may be another factor in the suffixing preference. Cutler et al.'s (1985) processing account explains the suffixing preference in terms of the preference of processing stems before suffixes and prefixes. However, the evidence from this thesis suggests that the suffixing preference may be influenced by how well language structure can be learned from succeeding and preceding cues. Higher learning ability with succeeding cues may influence the preference for suffixing, rather than prefixing, within language itself.

8.4.1. Marker word phonology and learning over time

Only two Experiments found significantly more learning in the second test session than in the first when both phonological and distributional cues were available to aid categorization. In Experiment 1, which had *vc/vcc* preceding marker words, and in the succeeding condition of Experiment 6, which had *vc* marker words, participants were able to learn additional information during the second training session that aided in differentiating the compatible and incompatible test items. See Table 8.1 for a summary of the Experiments and status of whether there was a difference between the test sessions.

Table 8.1 Experiment name, marker word location and phonology, and status of time difference within the coherent conditions.

Experiment	*Cue location*	*Phonology*	*Time?*
Experiment 1	Preceding	*vc/vcc*	Yes
Experiment 2	Succeeding	*vc/vcc*	No
Experiment 3	Preceding	*cvc*	No
Experiment 4	Succeeding	*cvc*	No[‡]
Experiment 5	Preceding	*cvc*	No
Experiment 5	Succeeding	*cvc*	No
Experiment 6	Preceding	*vc*	No
Experiment 6	Succeeding	*vc*	Yes
Experiment 7	Preceding	*cv*	No
Experiment 7	Succeeding	*cv*	No

[‡] There was a main effect of time in Experiment 4, but the advantage for the second test session was in both the coherent and incoherent conditions. The coherent condition alone was not significant.

The difference in Experiment 1 may have been due to the influence of the syntax-incompatible test sentences that may have interfered with the expression of categorical knowledge, especially in the incoherent condition. Although there was no evidence to support the conclusion that these test items interfered with learning in the coherent condition, there is still a possibility the syntax-incompatible sentences may have interfered slightly, as the associative strength between the category words (and phonological cues) and the preceding marker word was not of considerable strength. Therefore, the expression of the categorical information learned from the preceding cues may have been expressed more in the second, rather than the first test session. There was no corresponding increase in the second test session of Experiment 2, which used the same marker words and also contained the syntax-incompatible test sentences, but had succeeding distributional cues. This may have been due to the higher associative strength between the phonological cues and category words with the succeeding marker word; the increased associative strength may have prevented any influence from the presence of the syntax-incompatible test sentences.

The succeeding condition of Experiment 6 also had an increase in categorical knowledge in the second test session. Within the first test session of this Experiment there was no difference between the preceding and succeeding condition, but in the second test session the succeeding condition showed an increase in categorical knowledge. This was explained in detail in Chapter 5; the increased saliency of the succeeding marker words increased the overall asymptote level, allowing stronger associative links between the marker and category words. However, it might be that additional marker/category word pairings were needed for this higher asymptote to be reached.

While this explanation is perfectly consistent with the R-W account of categorical learning given in this thesis, the reason for this time difference to only occur with *vc* marker words is not as clear. It may be due to the fact that the *vc* marker words were most similar to the general phonological form of suffixes. However, this does not explain why there was no difference between the test sessions with the *cv* marker words, when the overall level of learning was higher. It is true that the *cv* phonological form was not very common in suffixes, but the higher learning still would have been expected to require additional category/marker word pairings and display significantly higher learning in the second test session. Although no firm conclusions can be drawn regarding the sporadic finding of higher learning with more training, it would be an interesting line of investigation for future work to investigate the conditions in which learning increases over the course of the Experiment.

8.4.2. Marker word phonology and task correlations

The significant positive correlations between the similarity and the card sorting task have sometimes been present within the Experiments, but this varied across Experiments. The presence of a significant positive correlation between the two tasks indicated that the higher the participants scored on the similarity task, the higher the likelihood they would be able to explicitly group the two categories of words together. See Table 8.2 for a summary of the results for all Experiments and marker word location conditions.

Table 8.2 Experiment name, marker word location and phonology, status of the similarity/card sorting task correlations within the coherent conditions.

Experiment	*Cue location*	*Phonology*	*Correlation?*
Experiment 1	Preceding	*vc/vcc*	No ($p = .11$)
Experiment 2	Succeeding	*vc/vcc*	Yes
Experiment 3	Preceding	*cvc*	No ($p = .18$)
Experiment 4	Succeeding	*cvc*	No ($p = .73$)
Experiment 5	Preceding	*cvc*	Yes
Experiment 5	Succeeding	*cvc*	No ($p = .18$)
Experiment 6	Preceding	*vc*	Yes
Experiment 6	Succeeding	*vc*	No ($p = .16$)
Experiment 7	Preceding	*cv*	Marginal ($p = .053$)
Experiment 7	Succeeding	*cv*	Yes

The first interesting finding when comparing the correlational results across all Experiments was that Experiment 3 (with preceding *cvc* marker words) again differed from the preceding condition of Experiment 5, also with *cvc* marker words. This is easily explained by the fact that the card sorting results in Experiment 3 were below average for all other Experiments. This would obviously interfere with the correlation between the two tasks.

In Experiments 5, 6 and 7 there was a general trend for significant correlations between the similarity and card sorting tasks in the preceding cue condition. The only preceding cue Experiment not to follow this pattern (exempting Experiment 3) was Experiment 1; however, the results of Experiment 1 indicated that there was a moderate correlation (r = .49); it was likely this was not significant due to a lack of power in the test (as there were only 12 participants included in this analysis). Generally, the findings indicate that high ability to differentiate compatible and incompatible test sentences resulting from categorical knowledge induced from a combination of preceding distributional and phonological cues was related to high abilities in explicitly sorting the two categories of words.

The results were not the same for the succeeding distributional and phonological cue combinations. Only two of the seven succeeding cue conditions had a positive correlation between the two tasks indicating that preceding distributional cues lead to higher association with the ability to differentiate the two categories of words. This may be because the backward association between the marker word and the category words allowed more explicit access to the associations between the relevant marker word and the category words. Alternatively, the higher associative strength between the category and marker words in the succeeding conditions may have allowed participants who did not perform well at the similarity task to still perform well at explicitly sorting the words according to their categories. Another possibility is that there was less noise within the succeeding cue condition as the learning was stronger than in the preceding cue condition. This reduced noise could have led to lower overall correlations between the two measures simply because there was less variation in the scores. However, a consultation of the data indicated that the second and third options are unlikely to be the cause of this difference.

8.5. Frame advantage: Learning from non-adjacent dependencies or combined bigram information

The last main topic in this thesis relates to learning from trigram and bigram distributional cues. Mintz (2002) claimed that frame cues were able to induce categorization on the basis of distributional information alone, whereas bigram distributional cues were unable to aid categorization without additional correlated cues. However, evidence in this thesis argues against this premise, as bigram distributional cues are sufficient to categorize high frequency words.

Additionally, the underlying nature of learning from trigram "frame" cues was one of the main questions that the last two chapters aimed to answer. Specifically, whether learning from trigram “frame” distributional cues was due to the influence of the non-adjacent frame elements (a_b) or whether it was due to the beginning and ending bigram working in combination was investigated. This relates fundamentally to what kind of distributional information is most useful and usable for grammatical categorization during language acquisition. The Simplicity Principle states that the simplest solution to any cognitive problem should be preferable over more complex solutions (Chater & Vitanyi, 2003). The use of the non-adjacent frame cues in preference to the highly informative bigram cues would perhaps violate this principle.

Evidence from simulations in Monaghan and Christiansen (2004) support the theory that the combined influence of the aX and Xb bigrams within the aXb frame constitutes learning from aXb frame cues. No learning with the neural network simulation was found when the non-adjacent cues were submitted as the only cue for categorizing the intervening X word, but when the initial and ending bigrams were allowed to also contribute separately significant learning within the simulation was found.

Chapters 6 and 7 further investigated this issue. Chapter 6 directly looked at the usefulness of trigram and initial and ending bigram cues. Although trigram cues were able to group words of the same grammatical category more consistently than either of the bigram cues this was done at the cost of severely restricting the total number of words classified. Overall it was found that the initial bigram analysis may be the best cue for capturing large amounts of words of the same grammatical category together with reasonable accuracy. The ending bigram cue was more accurate at categorizing words of the same grammatical category than the random baseline level, but the resulting categories were less accurate in terms of grammatical categories than the initial bigram and trigram cues.

This analysis was consistent with previous discussed preceding word cue and succeeding affix cue. The preceding word cues were found to be much better than the succeeding word cues. However, this analysis did not answer the question of whether learning from trigram frames, which have been demonstrated previously (Mintz, 2002, 2006; Thothathiri & Snedeker, 2005), occurs through the a_b non-adjacent frame or from the combination of the bigram information. However, it was found that vast majority of words in the *a* and *b* locations within the a_b frame were also used as initial and ending bigram word cues. This finding at least provided preliminary support to the idea that the combination of the two bigram cues may drive trigram distributional learning.

The findings of Experiment 8 and 9 provide conflicting evidence with the previous work. Only when trigram cues were combined with phonological cues did word categorization take place, whereas previous work has found categorization with only distributional cues (Mintz, 2002, 2006; Thothathiri & Snedeker, 2005). The discrepancy with previous work mainly lies with frequency; in the previous studies there were only a small amount of category words used. Experiment 8 had eight words per category which were not of sufficiently high frequency to obtain learning without co-occurring phonological cues. This finding was at odds with the claim by Mintz that trigram (or frame) distributional information can induce categorization without recourse to additional correlated cues; in light of these new findings, it seems that this statement is only consistent with high frequency words. Just as with bigram distributional cues, lower frequency words would still need the assistance of additional phonological cues to aid the categorization process.

Experiment 9 aimed to directly test whether learning from frame information was due to the non-adjacent information or whether it was due to the combination of the two bigram cues. There was no sign of word categorization within this Experiment, which only had consistent bigram cues. However, there was evidence that the participants categorized the phonological characteristics of the category words and associated these characteristics with the bigram cues which allowed the participants to generalize beyond the category words used in the training session to novel category words. This indicated, as mentioned before, that the associations between the abstract phonological cues within the category words and the bigram distributional cues formed, but that the associations with the actual category words were not sufficiently strong to induce word categorization.

Although no word categorization was found with only consistent bigram cues, the theory that trigram learning occurred through the combination of the two bigrams cannot be discounted. Thothathiri and Snedeker (2005) tested the same non-adjacent trigram versus bigram learning and found evidence to support the idea that learning from frames occurs through the combination of the two bigram cues. Additionally, in the current bigram study the ability to generalize the phonological form indicated that some categorical learning did occur. One prediction to come from this account is that with substantially more training word categorization would occur. Within the current experimental design, which is very complex, substantially more training is predicted to be required; however, if the design were to be simplified slightly, by perhaps reducing the number of category words in each category from eight to six, word categorization might occur with the current levels of training.

The last issue to discuss with regards to the final Experiments was the finding of generalization beyond the actual words used during the training sessions. This ability is vital in the word categorization process, and in grammar learning in general. Both Experiments that tested for generalization found that the participants could apply their knowledge of the phonological characteristics of the category words to novel category words. They were able to differentiate when a novel category word was in the "correct" distributional context from when the distributional context was "incorrect". In other words, they applied their knowledge of the general characteristics of the category and their associations with the distributional cues to determine whether the novel category word was within the correct distributional context.

This finding has direct implications for grammatical categorization in natural language acquisition. When a language learner encounters a new word, the combination of the phonological characteristics and the distributional context (and perhaps other linguistic or extra-linguistic cues) co-occurring with the word will aid the learner in categorizing the new word into the correct grammatical category. Once this categorization is complete, the language learner will then be able to generalize beyond the original syntactic context with which the new word was originally paired; all syntactic contexts which the learner knows are allowable for other category words are now potentially useable with this new word. This is perhaps a simplified view of this process; however, the essential components in the process of generalization are present within these Experiments and can be explained in terms of associative learning principles.

This thesis implies more generally that associative learning may be able to account for many aspects of language acquisition. The research presented in this thesis provides an associative learning account of the acquisition of grammatical categories. Smith (2000) provided evidence for an associative learning account of initial vocabulary acquisition. Ellis (2006a) has argued that language acquisition in general can be explained in terms of associative learning, in particular maintaining that "the driving forces of language learning ... are frequency, conditioned by contingency, conditioned by selection" (p. 15). Indeed, the many connectionist accounts of language learning phenomenon have underlined the importance of associations between the contingencies inherent within language. Additionally, many language learning shortcomings found in second language acquisition have even been explained in terms of associative learning (Ellis, 2006b). From these findings it is clear that associative learning plays an important role in language learning; however, a more complete and likely mechanism for explaining language acquisition is distributional and statistical mechanisms that perhaps rely on low level associative learning. The question remains whether there is room for innate language constructs within the associative language learning paradigm; there certainly seems to be no room for the principle and parameter approach to language acquisition. There is no need for innate *knowledge* with a powerful learning mechanism that accounts for many of the initial objections to language learning. However, as mentioned in Chapter 1 there is an innate basis to our underlying cognitive architecture. Therefore, it is highly likely that our ability to *learn* from frequency and contingency differences within language is innately specified. While associative learning may not account for all of language acquisition, the combined power of associative and statistical learning may be a sufficient explanation for the language acquisition problem.

8.6. Limitations

One main limitation to this work not yet mentioned is the use of synthesized speech within the Experiments. It could be argued that the use of unnatural speech may have restricted the applicability of the findings from all experimental work in this thesis. However, synthesized speech is widely used in artificial language learning studies to control for extraneous factors that are inherent in natural speech (e.g., Monaghan, Onnis, Merkx, & Chater, Submitted; Newport & Aslin, 2004; Peña, Bonatti, Nespor, & Mehler, 2002; Perruchet, Tyler, Galland, & Peereman, 2004; Saffran, Newport, & Aslin, 1996). In the Experiments presented in this thesis, the phonological cues within the category words were precisely controlled and there were no extraneous influences from prosody or accent variation that would inevitably occur with natural language. A reasonable solution to the problems of prosody and accent differences without recourse to synthesized speech would have been to record the stimuli with the same speaker throughout all Experiments. However, the availability of the same speaker for all recordings could not be guaranteed throughout the entire time frame of this thesis; therefore, synthesized speech was used in all Experiments.

Although synthesized speech was used throughout this thesis, Experiment 9 was replicated with recorded speech. The possibility investigated in this further study was that the synthesized speech may have influenced the learning rate within the Experiments, and thus was a potential cause of the lack of word categorization in Experiment 9. However, this further Experiment using recorded speech did not differ in results from the corresponding Experiment with synthesized speech (both coherent and incoherent conditions were not significant different between the Experiments; $t(26) = -0.59$, $p = .56$ and $t(26) = -0.52$, $p = .61$, respectively for main categorization test). This provides evidence that the use of synthesized speech within the Experiments presented in this thesis did not cause any difference from the results which could have occurred with natural speech stimuli.

A further limitation was the inseparability of embedded trigrams in the *aAbB* and *AaBb* languages. The learning in Experiments 1 to 7 which investigated the preceding and succeeding cue difference has been discussed as bigram learning throughout this thesis. However, there were also less reliable trigram cues in both the *aAbB* and *AaBb* structured languages. For instance with the *aAbB* language half of the presentations of category A words were within the trigram a_b, while the other half had only the perfectly predictive marker word *a* cue. Although the categorization found within these Experiments could theoretically be argued to be at least partially due to these embedded trigram cues, it would also follow that categorization would be found within Experiment 9, which had similarly perfectly predictive bigram cues but only partially predictive trigram cues. Since there was no grammatical categorization found in Experiment 9, the feasibility of substantial learning from trigram cues in the "bigram" experiments is unlikely. When looking at only bigrams there was also a succeeding marker word cue in the *aAbB* structure languages in half of the category word presentations. The possible influence of these dependencies cannot be discounted as easily as the trigram cues discussed above. Future experiments could separate the two phrases ("*aA*" and "*bB*") with a significant pause to interfere with these extraneous dependencies.

An additional limitation of this work is the fact that all languages were spoken without a context. There were no referents for any of the category words and the two categories of words did not directly refer to nouns, verbs, or any other grammatical category, although some participants reported that they imagined one phrase referred to verbs and the other phrase to nouns. Within the context of word learning, cross situational referent learning has been shown to occur within experimental settings (Yu & Smith, 2006, in press). Given only six minutes of exposure to pairings of multiple words to multiple picture referents participants show a remarkable ability to learn from the statistical co-occurrences of the word-picture referent across all of the training trials; only by comparing the referent-word combinations across all exposure trials could the words be mapped to the correct referent (Yu & Smith, in press). This gives evidence that young language learners may be able to map the co-occurrences of real word referents across multiple instances; in other words, the statistical co-occurrence mechanism postulated for analysing transitional probabilities in segmentation and distributional information in grammatical categorization may indeed work for mapping words to their real world referents as well. These findings allow the possibility that grammatical categorization may benefit from cross-situational referent learning as well; although referent context was not investigated in this thesis, the possibility of further learning with increased grammatical category cues is consistent with the findings.

The artificial language learning Experiments presented in this thesis share common limitations to all ALL studies. One limitation is the influence of the first language on the artificial language tasks; this was well documented and investigated in this thesis. More learning occurred than when the marker word constructs were similar to the corresponding English distributional cues than when they were slightly dissimilar. In the context of this thesis, interference due to the participants' native language was perhaps less of a limitation than a line of empirical inquiry. The small size of the artificial languages used within these studies is also a major limitation. However, the sheer scale of natural language requires language acquisition researchers to scale down necessarily. As learning from phonological and distributional cues was under investigation in this thesis, these cues were isolated away from all other linguistic cues to determine whether humans can learn from the combination of these cues. Small categories and a small scale language are necessary components of this process, as learning from larger languages occurs along a much longer time frame and would require experiments of a length that the average participant would not agree to. Inherent in this small scale nature of the languages are much smaller categories of words than are found in natural languages. Although the Experiments within this thesis suffer from limitation, they have improved upon some previous grammatical categorization literature (e.g., Mintz, 2002 had four words per category; Thothathiri & Snedeker, 2005 had two words per category) while falling short of some studies (e.g., Brooks, Braine, Catalano, Brody, & Sudhalter, 1993 had 15 words per category; Frigo & McDonald, 1998 had ten words per category).

The other limitations within this thesis have mainly been discussed throughout the thesis. The presence of the syntax-incompatible test sentences within Experiments 1 and 2 cause difficulties in interpreting the results from the incompatible conditions. Additionally, the lack of direct comparability due to these test sentences between Experiments 1 and 2 with subsequent Experiments also caused a slight problem in interpreting results across Experiments.

8.7. Future work

Further investigations into the non-adjacent versus bigram learning from distributional frames are warranted from the discussion above. Increased training or a simplified language may aid in determining whether word categorization resulting from aXb frames could be the result of the combined influence of the bigram cues. Additionally, future experiments could potentially investigate whether the learning from the aXb frames is mostly due to the initial bigram, ending bigram or is truly a combination of the two bigrams. This could be done by constructing an artificial language similar to the language used in Experiment 9 but with only informative initial bigrams or informative ending bigrams (i.e., aXY and YXb, where Y is a random word with no informative value).

Future work looking at how the suffixing preference may be driven by the increased ease of learning from succeeding cue information. This may be done by creating languages that contain the distributional cues within the category words themselves, similar to the design of the Frigo and McDonald (1998) study. This would allow direct tests of the difference between prefixes and suffixes. However, unlike Frigo and McDonald, who paired the "marked" words with an ending bigram distributional cue, the suffix and prefix like markers would be the distributional cues, with additional phonological cues within the category words themselves.

A further replication of the preceding *cvc* marker words would aid in differentiating which level of learning is representative from these marker words in the preceding location, although all current evidence suggests the learning in Experiment 5 is more representative than learning in Experiment 3. A replication of the corpus analysis presented within Chapter 6, but using within word preceding and succeeding cues would also be an interesting further direction this work could take. It is unlikely using only prefixes and suffixes would lead to comprehensive evaluation, as there are so few prefixes within the English language that could give reliable grammatical category cues, but this may be a first step towards evaluating the effectiveness of within word cues. This could also have implications for whether the suffixing preference could have been influenced by increased learning from suffix cues.

Future investigations could also integrate additional correlated cues that aid in differentiating the categories. For instance, semantics could be included by integrating the presentation of referent objects (or actions) during the presentation of the auditory language. A similar paradigm as is used in word referent mapping studies could be implemented on a category basis (Yu & Smith, 2006, in press). The combination of distributional, phonological and referent cues would provide a more realistic category (and word) learning paradigm. Similar categorization tests could be included as are found in the Experiments presented in this thesis, but additional tests to determine whether the specific category word-referent object associations have been formed would aid in determining what role referent objects in the environment play in the word categorization task (specifically in noun acquisition). Generalization tests similar to those used in Experiments 8 and 9 may be used to determine whether phonological plus distributional cues allow generalizations to novel referent objects or actions. Success in this type of generalization test would strengthen the evidence of grammatical category generalization found in Experiments 8 and 9.

Perhaps the most exciting avenue for future work is to investigate how sleep influences the processes involved in language acquisition. Initial evidence from Gómez, Bootzin, and Nadel (2006) indicate that the abstraction of language stimuli occurs after a period of sleep, but not after an equivalent period of wakefulness. Initial pilot results with the language design in Experiments 1 to 7 indicated that the same was true for abstraction to novel category words (as used in Experiments 8 and 9). There was generalization to these novel category words after a 12 period that included sleep but not after an equivalent 12 hour period that did not include sleep. The finding of generalization beyond training material after periods of sleep has been found throughout a range of tasks in many modalities (Fischer, Drosopoulos, Tsen, & Born, 2006; Gómez, Bootzin, & Nadel, 2006; Karni, Tanne, Rubenstien, Askenasy, & Sagi, 1994; Mednick et al., 2002; Wagner, Gais, Haider, Verleger, & Born, 2004). Specific language findings include sleep facilitation in phoneme generalization (Fenn, Nusbaum, & Margoliash, 2003), new word learning (Dumay & Gaskell, 2007), and grammar learning (Gómez, Bootzin, & Nadel, 2006). Extending the pilot findings and fully investigating the role of sleep in language learning would be an exciting extension of the work presented in this thesis.

8.8. Conclusions

There are five main contributions from this thesis. To summarize:

1. Grammatical categorization occurs through a combination of correlated cues; the combination of distributional information across grammatical utterances and phonological information within the category words has been found to be important in the grammatical categorization process.
2. Grammatical categorization can be explained by the associations between phonological cues and the category words with co-occurring distributional information according to the laws of associative learning from the Rescorla-Wagner model.
3. Succeeding distributional cues allow higher levels of categorical knowledge than preceding distributional cues; this is predicted by the R-W model of associative learning.
4. Prior linguistic experience influences the ease of learning from preceding and succeeding distributional cues; the evidence suggests that preceding words are the most useful preceding categorization cue in English, while suffixes are the most useful succeeding cue.

5. Distributional frame cues are highly useful and usable cues during grammatical categorization; however, the evidence suggests that the frame cues usefulness lies within two bigram cues embedded within the distributional frame, although the evidence is not conclusive.

The first finding indicates that true grammatical categorization of both high and low frequency words cannot occur through distributional cues alone; the combination of at least phonological and distributional cues is necessary. However, it is likely that further correlates of grammatical categories also aid grammatical categorization. The associative learning account of grammatical categorization advocated within this thesis predicts that all cues that consistently co-occur with grammatical category members could potentially aid the grammatical categorization process. What is clear from this thesis is that language acquisition (or at least the acquisition of grammatical categories) is driven by multiple cue learning; no one single cue can account for grammatical categorization.

This thesis also provides strong evidence for the succeeding cue advantage, which is explained in terms of the increased salience (and higher associative strength) with succeeding distributional cues. The findings that indicate that suffixes may be the most consistent succeeding cue in English allows some interesting ideas relating to the general suffixing preference across all human languages. While previous accounts provide a logical explanation for this preference (Cutler et al., 1985), the current findings indicate potential learning constraints on prefixing and suffixing. This thesis indicates that language learning should progress faster within a suffixing language, as it will, at the very least, aid grammatical categorization more than a prefixing language. The preferential learning from suffix information over prefix information may have been a learning constraint that tended to shape languages towards suffixing rather than prefixing.

The last major point in this thesis relates to whether the basis of learning from frame distributional cues lies within the non-adjacent (a_b) cue or with the initial and ending bigram cues combined. Although the results do not allow any firm conclusions, the pattern of results from this thesis and other labs (Thothathiri & Snedeker, 2005) as well as previous work (Monaghan & Christiansen, 2004) provide converging evidence that bigram learning may be the basis of learning from the aXb frame cues. This conclusion, if validated, indicates that bigram distributional cues are the main distributional cue that aids grammatical categorization during language acquisition.

In summary, this thesis has extended previous work by demonstrating that phonological and distributional cues are both necessary in the grammatical categorization process. An associative learning explanation of grammatical categorization has also been developed that will hopefully lead to further applications of associative learning to the language learning field in general. Finally, this work indicates that bigram distributional cues may be the fundamental source of distributional information in the grammatical categorization process.

Appendix A

Table A Experiment 1 training sentences for the dialect 1 coherent condition.

Experiment 1 Training Sentences
alt tweand ong foth
ong vawse alt tweand
alt tweand ong zodge
ong thorsh alt tweand
ong foth alt dreng
alt dreng ong suwch
alt dreng ong thorsh
ong shufe alt dreng
alt klimp ong vawse
ong suwch alt klimp
ong zodge alt klimp
alt klimp ong shufe
alt gwemb ong foth
ong vawse alt gwemb
alt prienk ong vawse
ong suwch alt prienk
ong foth alt blint
alt blint ong suwch

Appendix B

Table B. Experiment 1 test 1 and 2 test sentences for the dialect 1 coherent condition.

Coherent Condition Dialect 1 Test 1	*Coherent Condition Dialect 1 Test 2*
Compatible High Frequency	Compatible High Frequency
alt tweand ong suwch	alt dreng ong foth
alt dreng ong vawse	alt tweand ong vawse
ong foth alt klimp	alt klimp ong suwch
ong vawse alt dreng	ong foth alt tweand
ong suwch alt tweand	ong vawse alt klimp
alt klimp ong foth	ong suwch alt dreng
Compatible Low Frequency	Compatible Low Frequency
alt gwemb ong zodge	alt gwemb ong thorsh
alt prienk ong shufe	alt prienk ong zodge
alt blint ong thorsh	alt blint ong shufe
ong zodge alt blint	ong zodge alt gwemb
ong thorsh alt prienk	ong thorsh alt blint
ong shufe alt gwemb	ong shufe alt prienk
Syntax-Incompatible High Frequency	Syntax-Incompatible High Frequency
alt klimp alt vawse	ong suwch ong dreng
alt suwch alt dreng	alt tweand alt foth
ong thorsh ong tweand	ong klimp ong shufe
Syntax-Incompatible Low Frequency	Syntax-Incompatible Low Frequency
alt blint alt zodge	alt thorsh alt gwemb
ong prienk ong shufe	alt zodge alt blint
ong foth ong gwemb	ong prienk ong vawse
Incompatible High Frequency	Incompatible High Frequency
ong dreng alt foth	alt vawse ong tweand
alt vawse ong klimp	ong dreng alt suwch
ong tweand alt suwch	alt foth ong klimp
Incompatible Low Frequency	Incompatible Low Frequency
alt shufe ong blint	ong gwemb alt zodge
ong gwemb alt thorsh	alt shufe ong prienk
alt zodge ong prienk	ong blint alt thorsh

Appendix C

Table C Test sentences in the coherent condition for Experiments 3 and the preceding conditions of Experiments 5, 6, and 7.

Coherent Condition Test 1	*Coherent Condition Test 2*
Compatible High Frequency	Compatible High Frequency
markerA tweand markerB vawse	markerA tweand markerB suwch
markerA dreng markerB foth	markerA dreng markerB vawse
markerA klimp markerB suwch	markerA klimp markerB foth
markerB foth markerA klimp	markerB foth markerA tweand
markerB vawse markerA dreng	markerB vawse markerA klimp
markerB suwch markerA tweand	markerB suwch markerA dreng
Compatible Low Frequency	Compatible Low Frequency
markerA gwemb markerB zodge	markerA gwemb markerB thorsh
markerA prienk markerB shufe	markerA prienk markerB zodge
markerA blint markerB thorsh	markerA blint markerB shufe
markerB zodge markerA blint	markerB zodge markerA gwemb
markerB thorsh markerA prienk	markerB thorsh markerA blint
markerB shufe marker A gwemb	markerB shufe markerA prienk
Incompatible High Frequency	Incompatible High Frequency
markerB tweand markerA suwch	markerB tweand markerA vawse
markerB dreng markerA vawse	markerB dreng markerA foth
markerB klimp markerA foth	markerB klimp markerA suwch
markerA foth markerB tweand	markerA foth markerB klimp
markerA vawse markerB klimp	markerA vawse markerB dreng
markerA suwch markerB dreng	markerA suwch markerB tweand
Incompatible Low Frequency	Incompatible Low Frequency
markerB gwemb markerA shufe	markerB gwemb markerA zodge
markerB prienk markerA thorsh	markerB prienk markerA shufe
markerB blint markerA zodge	markerB blint markerA thorsh
markerA zodge markerB prienk	markerA zodge markerB blint
markerA thorsh markerB gwemb	markerA thorsh markerB prienk
markerA shufe markerB blint	markerA shufe markerB gwemb

Appendix D

Table D Beta values (with S.E.), significance levels, and 95% CI for exp b for the C_C logistic regression analysis (type analysis; Step 4 variables only shown)

				95% CI for exp b		
Step 4		*B (S.E.)*	*Significance*	*Lower*	*exp b*	*Upper*
Constant		40.96 (41909.02)	1.0		6.15	
Beginning consonant manner of articulation	*Overall*		.116			
	(1)	-21.55 (11870.32)	1.0		.00	
	(2)	-22.72 (11870.32)	1.0		.00	
	(3)	-17.48 (11870.32)	1.0		.00	
	(4)	-20.71 (45790.53)	1.0		.00	
Beginning consonant place of articulation	*Overall*		.191			
	(1)	-21.12 (40192.84)	1.0		.00	
	(2)	-22.22 (40192.84)	1.0		.00	
	(3)	-19.73 (40192.84)	1.0		.00	
	(4)	-17.56 (40192.84	1.0		.00	
	(5)	.28 (26663.78)	1.0	.00	1.33	--
Ending consonant manner of articulation	*Overall*		.113			
	(1)	2.96 (1.28)	.021	1.57	19.32	237.86
	(2)	.32 (1.16)	.781	.14	1.38	13.31
	(3)	1.89 (1.23)	.125	.59	6.61	73.91
	(4)	1.03 (1.35	.444	.20	2.80	39.09

Appendix E

Table E Beta values (with S.E.), significance levels, and 95% CI for exp b for the C_C logistic regression analysis (token analysis)

				95% CI for exp b		
Step 1		*B (S.E.)*	*Significance*	*Lower*	*exp b*	*Upper*
Constant		24.80 (4091.38)	1.0		587520931	
Beginning consonant manner of articulation	*Overall*		< .001			
	(1)	-25.61 (1054.11)	.98		.00	
	(2)	-30.89 (1054.11)	.98		.00	
	(3)	-14.35 (1054.11)	.99		.00	
	(4)	-26.27 (4081.32)	1.0		.00	
Beginning consonant place of articulation	*Overall*		< .001			
	(1)	-23.74 (3791.48)	1.0		.00	
	(2)	-30.67 (3791.48)	.99		.00	
	(3)	-27.07 (3791.48)	.99		.00	
	(4)	-9.08 (3791.48)	1.0		.00	
	(5)	.50 (1439.14)	1.0		1.65	
Beginning consonant voicing	*(1)*	3.24 (.212)	< .001	16.77	25.42	38.54
Ending consonant manner of articulation	*Overall*		< .001			
	(1)	9.21 (.36)	< .001	4922.5	10027.75	20427.8
	(2)	.47 (.23)	.04	1.01	1.60	2.51
	(3)	5.65 (.28)	< .001	164.8	285.41	494.28
	(4)	-8.23 (.20)	< .001	.00	.00	.00
Ending consonant place of articulation	*Overall*					
	(1)	-17.17 (1119.35)	.99		28528215	
	(2)	-17.06 (1119.35)	.99		25733024	
	(3)	-17.031 (1119.35)	.99		32860889	
Ending consonant voicing	*(1)*	.91 (.18)	< .001	1.76	2.50	3.54

Appendix F

Table F Beta values (with S.E.), significance levels, and 95% CI for exp b for the C_V logistic regression analysis (token analysis; Step 2 variables only shown)

				95% CI for exp b		
Step 2		*B (S.E.)*	*Significance*	*Lower*	*exp b*	*Upper*
Constant		85.30 (8916.81)	.99		1.11e+37	
Beginning consonant manner of articulation	*Overall*		1.0			
	(1)	-42.37 (466.48)	.93		.00	
	(2)	-43.30 (1154.84)	.97		.00	
	(3)	40.01 (12050.47)	1.0		2.38e+17	
	(4)	-21.75 (314.98)	.95	.00	.00	4.7e+258
Beginning consonant place of articulation	*Overall*		1.0			
	(1)	-64.07 (8937.80)	.99		.00	
	(2)	-64.13 (8925.58)	.99		.00	
	(3)	15.89 (11841.60)	1.0		7924453	
	(4)	-63.12 (16274.45)	1.0		.00	
Beginning consonant voicing	*(1)*	-42.34 (645.30	.95		.00	
Ending vowel height	*Overall*		1.0			
	(1)	42.32 (949.53)	.96		2.39e+18	
	(2)	.992 (1377.03)	1.0		2.70	
	(3)	42.32 (13997.47)	1.0		2.39e+18	
	(5)	-80.02 (7812.02)	.99		.00	

Appendix G

Table G Beta values (with S.E.), significance levels, and 95% CI for exp b for the V_C logistic regression analysis (type analysis; Step 4 variables only shown)

				95% CI for exp b		
Step 4		*B (S.E.)*	*Significance*	*Lower*	*exp b*	*Upper*
Constant		39.44 (32583.06)	1.0		1.35e+17	
Beginning vowel position of articulation	*Overall*		.739			
	(1)	-21.26 (50386.23)	1.0		.00	
	(2)	.00 (44935.33)	1.0		1.0	
	(3)	.86 (44935.33)	1.0		2.35	
	(4)	.88 (44935.33)	1.0		2.4	
	(5)	19.23 (46092.18)	1.0		4.5e+08	
Beginning vowel roundedness	*Overall*		1.0			
	(1)	20.98 (48360.35)	1.0		1.29e+09	
	(2)	-20.02 (40191.60)	1.0		.00	
Ending consonant place of articulation	*Overall*		.019			
	(1)	-18.24 (25647.78)	1.0		.00	
	(2)	-20.54 (25647.78)	1.0		.00	
	(3)	19.62 (25647.78)	1.0		.00	
	(4)	-20.01 (25647.78)	1.0		.00	

Appendix H

Table H Beta values (with S.E.), significance levels, and 95% CI for exp b for the V_C logistic regression analysis (token analysis)

				95% CI for exp b		
Step 1		*B (S.E.)*	*Significance*	*Lower*	*exp b*	*Upper*
Constant		43.23 (6149.93)	.99		6.24e+18	
Beginning vowel height	*Overall*		< .001			
	(1)	16.45 (6295.09)	1.0		13905254	
	(2)	-24.41 (5940.66	1.0		.00	
	(3)	-4.77 (6277.81)	1.0		.008	
	(4)	-5.26 (6023.94)	1.0		.005	
	(5)	-2.95 (6277.81)	1.0		.052	
	(6)	-6.12 (6277.81)	1.0		.002	
	(7)	.287 (6335.18)	1.0		1.33	
	(8)	-5.76 (6051.80)	1.0		.003	
Beginning vowel position of articulation	*Overall*		1.0			
	(1)	-27.58 (4386.83)	1.0		.00	
	(2)	3.54 (4244.96)	1.0		.029	
	(3)	-5.79 (4348.39)	1.0		.003	
	(4)	15.49 (4360.78)	1.0		5307227.8	
	(5)	15.87 (4224.55)	1.0		7830549.2	
Beginning vowel roundedness	*Overall*		1.0			
	(1)	22.48 (4223.02)	1.0		5.77e+09	
	(2)	-16.52 (4111.24)	1.0		.00	
Ending consonant manner of articulation	*Overall*		< .001			
	(1)	4.03 (.03)	< .001	53.05	56.26	59.67
	(2)	2.35 (.035)	< .001	9.83	10.53	11.28
	(3)	.035 (.019)	.057	.999	1.04	10.8
	(4)	1.08 (.018)	< .001	2.82	2.93	3.04
Ending	*overall*		<.001			

consonant place of articulation	*(1)*	-18.64 (645.21)	.98		.00	
	(2)	-25.57 (645.21)	.97		.00	
	(3)	-23.15 (645.21)	.97		.00	
	(4)	-20.94 (645.21)	.97		.00	
Ending consonant voicing	*(1)*	1.34 (.027)	< .001	3.64	3.83	4.04

Appendix I

Below are detailed the results from the beginning and ending consonant and vowel phonological features which were reviewed in Chapter 4 (see Section 4.1.2.9).

Beginning consonant manner of articulation

When looking at the beginning consonant manner of articulation there was a significant difference between prefixes and suffixes with the type analysis, $\chi^2(4) = 19.79$, $p < .05^{\dagger\dagger}$, V = .39 and with the token analysis, $\chi^2(4) = 51079.70$, $p < .05^{\dagger\dagger\dagger}$, V = .47. The majority of prefixes in both analyses began with either a plosive or a fricative. Beginning nasal and approximant consonants occurred in a small number of prefixes in the type analysis, but there were virtually no beginning nasal phonemes in the token analysis. Plosive, fricative and nasal consonants occurred with approximately equal frequency as the first phoneme of suffixes with both analyses. Additionally, there were more suffixes than prefixes that began with approximants and a substantial minority of lateral approximant consonants began suffixes, whereas there were no lateral approximant beginning consonants of prefixes.

Table I1 Beginning consonant phoneme manner of articulation for prefixes and suffixes with type and token analyses

		Beginning Consonant Manner of Articulation				
		Plosive	*Fricative*	*Nasal*	*Approximant*	*Lateral Approximant*
Type Analysis	*Prefix (N = 30)*	53.3%	36.7%	6.7%	3.3%	0%
	Suffix (N = 99)	22.2%	23.2%	29.3%	11.1%	14.1%
Token Analysis	*Prefix (N = 101356)*	64.3%	23.2%	.6%	11.9%	0%
	Suffix (N = 128937)	31.4%	23.9%	16.5%	7.2%	21%

Beginning manner of articulation contributed significantly in differentiating prefixes and suffixes when combined with other variables. The importance of this variable in differentiating prefixes and suffixes may have been influenced by the general phonotactic constraints of word initial phonemes and word internal phonemes, which would relate to the beginning consonant phonemes of prefixes and suffixes, respectively. It may also be that beginning prefix plosives and fricatives allow higher salience, which may be a beneficial addition for grammatical category cues in the preceding location. Cues in the succeeding location (suffixes) do not necessary need this additional salience as associative learning mechanisms, and the previous experimental results, allow faster learning from succeeding cues.

Beginning consonant voicing

There was no significant difference in beginning consonant voicing across prefixes and suffixes with the type analysis, $\chi^2(1) = 6.85, p = .14^{\dagger\dagger}$, V = .23 but there was a significant difference with the token data, $\chi^2(1) = 17060.99, p < .05^{\dagger\dagger\dagger}$, V = .27. The beginning prefix consonants were more likely to be unvoiced while the majority of beginning consonants of suffixes were voiced. The prevalence of unvoiced beginning prefix consonants was higher with the type analysis than the token analysis, although the difference within the suffixes was more pronounced in the token analysis. The difference in beginning consonant voicing can be partially explained by beginning consonant manner of articulation; for the type analysis 54.4% of suffix and 10% of prefix consonants were, by the nature of the manner of articulation, voiced phonemes. With the token analysis 44.7% of beginning suffix consonants and 12.5% of beginning prefix consonants were inherently voiced.

Table I2 Beginning consonant voicing for prefixes and suffixes with type and token analyses

		Beginning Consonant Voicing	
		Voiced	*Unvoiced*
Type Analysis	*Prefix (N = 30)*	36.7%	63.3%
	Suffix (N = 99)	63.6%	36.4%
Token Analysis	*Prefix (N = 101355)*	44.7%	55.3%
	Suffix (N = 128937)	71.6%	28.4%

The difference may also be due to the phonotactic differences between word initial/ending and word internal phonemes; however, an inspection of ending consonant voicing does not fully support this hypothesis. The majority of ending consonants for prefixes (thus, word internal) are voiced, but so are the ending consonants of suffixes, which occur as the word ending phoneme. Beginning consonant voicing was more important with the C_V analysis than with the C_C analysis, which indicated that perhaps this variable was not as individually important and should be considered alongside beginning consonant manner of articulation, as these two variables are inherently linked.

Beginning consonant place of articulation

Beginning consonant place of articulation did not differ between the two types of affixes for the type analysis, Fisher's Exact test = 5.84, $p = 1.0$, V = .23, although there was a significant difference between prefixes and suffixes in the token analysis, $\chi^2(7) = 82667.76$, $p < .05^{\dagger\dagger\dagger}$, V = .60. The majority of beginning suffix consonants were alveolar consonants, whereas there was a more equal spread amongst alveolar, bilabial, velar, and labiodental consonants for beginning prefix consonants.

Table I3 Beginning consonant phoneme place of articulation for prefixes and suffixes with type and token analyses

		Beginning Consonant Place of Articulation							
		Alveolar	*Bilabial*	*Velar*	*Labiodental*	*Labial-Velar*	*Palatal*	*Glottal*	*Dental*
Type Analysis	*Prefix (N = 30)*	60.0%	23.3%	13.3%	3.3%	0%	0%	0%	0%
	Suffix (N = 99)	41.4%	21.2%	14.1%	11.1%	6.1%	4.0%	1.0%	1.0%
Token Analysis	*Prefix (N = 101356)*	41.8%	21.4%	26.7%	10.0%	0%	0%	0%	0%
	Suffix (N = 128937)	84.8%	4.3%	.2%	1.0%	1.3%	.5%	.1%	7.8%

A slight prevalence for alveolar places of articulation was to be expected as six of the top nine most frequent consonant phonemes have alveolar articulation (/l/, /ɹ/, /s/, /n/, /d/, and /t/; Fry, 1947). However, the overwhelming prevalence of 84.8% of beginning suffix consonants having alveolar articulation was unlikely to be only due to these frequent phonemes, although the prefix 41.8% prevalence of alveolar articulation may be explained by these frequent phonemes. Alternatively, the difference between the affixes may be due to a tendency to have more varied place of consonant articulation within words than at the end of words. However, this explanation was not backed up as the ending consonants of suffixes were also likely to be alveolar consonants, indicating that both beginning and ending consonants of suffixes are frequently alveolar consonants.

Beginning vowel height

There was no significant difference between prefixes and suffixes in the type analysis, Fisher's Exact Test = 6.01, p = 1.0, V = .21. However, there was a significant difference in the token analysis, $\chi^2(8) = 72442.19$, $p < .05^{\dagger\dagger\dagger}$, V = .47. Both prefixes and suffixes had the highest prevalence of beginning vowels midway between close-mid and open-mid height; this is where the highly frequent /ə/ vowel is located. The prefixes also had a high prevalence of /æ/ vowels, which are midway between open-mid and open on the low end of the vowel height scale. The suffixes had fewer vowels on the lower end but more on the higher end of the scale with approximately 29.3% close to close-mid vowels (e.g., /I/). Overall, there seemed to be a larger spread in the beginning vowels of suffixes, with more heterogeneity, and a smaller spread in prefixes, with a tendency for vowels to be of medium or low height. See Table I4 for a summary of the data.

Table I4 Beginning vowel height for prefixes and suffixes with type and token analyses

		Beginning Phoneme - Vowel Height								
		Close (1)	*1.5*	*1.75*	*Close-mid(2)*	*2.5*	*2.75*	*Open-mid(3)*	*3.5*	*Low (4)*
Type Analysis	*Prefix (N = 29)*	3.4%	24.1%	6.9%	10.3%	44.8%	0%	6.9%	3.4%	0%
	Suffix (N=128)	11.7%	22.7%	7%	3.1%	35.2%	1.6%	12.5%	3.1%	3.1%
Token Analysis	*Prefix (N = 94993)*	0%	14.6%	0%	8.8%	35.4%	0%	7.4%	33.8%	0%
	Suffix (N = 232816)	3.6%	29.3%	4.1%	2.0%	46.3%	0%	4.2%	4.8%	5.6%

Beginning vowel position

There was no significant difference between prefixes and suffixes on initial vowel position of articulation with the type analysis, Fisher's Exact Test = 4.78, p = 1.0, V = .18. However, there was a significant difference for the token analysis, $\chi^2(6) = 44389.50$, $p < .05^{\dagger\dagger\dagger}$, V = .37. Centrally positioned vowels were most common in both prefixes and suffixes; this was to due to the high prevalence of the /ə/ beginning vowel again. Front vowels were also common as beginning vowels of both prefixes and suffixes, but they were more prevalent in prefixes than in suffixes. Vowels between the front and central designations (i.e., /I/) were more common as beginning vowels in suffixes than in prefixes. Overall, the beginning vowels of suffixes were slightly more centrally located than the prefixes. However, the V_C combined analysis indicated that beginning vowel position was important in differentiating prefixes and suffixes; it was likely that the usefulness of this measure was increased when combined with other beginning vowel measures, particularly vowel roundedness.

Table I5 Beginning vowel position for prefixes and suffixes with type and token analyses

		Beginning Phoneme – Vowel position						
		Front (1)	*1.25*	*1.5*	*Central (2)*	*2.25*	*2.75*	*Back (3)*
Type Analysis	*Prefix (N = 29)*	20.7%	3.4%	24.1%	44.8%	0%	3.4%	3.4%
	Suffix (N=128)	24.2%	11.7%	25.0%	28.1%	.8%	2.3%	7.8%
Token Analysis	*Prefix (N = 94993)*	45.7%	0%	14.6%	35.4%	0%	0%	4.3%
	Suffix (N = 232816)	13.8%	6.5%	29.6%	42.8%	.3%	.9%	6.1%

Beginning vowel roundedness

As Table I6 shows, the majority of beginning vowels were unrounded. There was no difference between the two affix groups with the type analysis, Fisher's Exact test = 2.17, $p = 1.0$, V = .13, although there was a significant difference with the token analysis, $\chi^2(2) = 7499.74$, $p < .05^{\dagger\dagger\dagger}$, V = .15. The majority of beginning vowels were unrounded in both the type and token analysis, with more variability in the suffix condition across both analyses. Although there was not much difference between prefixes and suffixes within the individual analysis, this variable was significant in the V_C combined analysis.

Table I6 Beginning vowel roundedness for prefixes and suffixes with type and token analyses

		Beginning Vowel Roundedness		
		Rounded	*Unrounded*	*Diphthong-Rounded/unrounded*
Type Analysis	*Prefix (N=29)*	0%	96.6%	3.4%
	Suffix (N=128)	7.8%	88.3%	3.9%
Token Analysis	*Prefix (N=94993)*	0%	100%	0%
	Suffix (N=232816)	6.2%	92.5%	1.3%

Ending consonant manner of articulation

There were no differences in the type analysis between the two affix groups when looking at the ending consonant manner of articulation, $\chi^2(4) = 1.13, p = 1.0$, V = .07. However, the token analysis did show a significant difference across prefixes and suffixes, $\chi^2(4) = 41161.62, p < .05^{\dagger\dagger\dagger}$, V = .34. Approximately half of all consonant ending prefixes ended with a nasal, with the other half split fairly evenly across the other four possible articulating consonants. The suffixes, on the other hand, had no skew in ending consonant type. Plosives were found most frequently and the remaining four types of consonants were approximately equally frequent as an ending consonant, with the exception of fewer lateral approximant ending consonants.

Table I7 Ending consonant manner of articulation for prefixes and suffixes with type and token analyses

		Ending Consonant Manner of Articulation				
		Plosive	*Fricative*	*Nasal*	*Approxi-mant*	*Lateral Approxi-mant*
Type Analysis	*Prefix (N = 47)*	29.8%	23.4%	19.1%	17.0%	10.6%
	Suffix (N = 168)	32.1%	28.0%	17.9%	11.9%	10.1%
Token Analysis	*Prefix (N = 127663)*	15.5%	11.7%	48.3%	17.4%	7.1%
	Suffix (N = 230941)	35.9%	17.0%	17.7%	20.4%	9.0%

This result was consistent with prefixes attaching onto the following words; nasal consonants often aid in adhering prefixes to the root words (Bauer, 2003b). The result that plosives were the most frequent ending phoneme manner of articulation for suffixes mirrored the finding that the majority of prefixes also began with plosive consonants, although the prevalence was stronger for beginning consonant prefix phonemes. This supports the possibility that plosives may be more common as beginning and ending phonemes, although the hypothesis of higher saliency in beginning prefix plosives cannot be discounted.

Ending consonant voicing

There was no significant difference in ending consonant voicing between prefixes and suffixes with the type analysis, $\chi^2(1) = .435, p = 1.0$, V = .05. However, the token analysis did show a significant difference between prefixes and suffixes, $\chi^2(1) = 1706.42$, $p < .05^{\dagger\dagger\dagger}$, V = .07. The general pattern showed that ending consonant phonemes for both prefixes and suffixes tend to be voiced; this preference was stronger with prefixes than suffixes. It is likely that ending consonant voicing is more predictive when combined with ending consonant place and manner of articulation, as has been found in the combined analyses.

Table I8 Ending consonant voicing for prefixes and suffixes with type and token analyses

		Ending Phoneme Voicing	
		Voiced	*Unvoiced*
Type Analysis	*Prefix (N = 47)*	59.6%	40.4%
	Suffix (N = 168)	54.2%	45.8%
Token Analysis	*Prefix (N = 127662)*	79.9%	20.1%
	Suffix (N = 230940)	73.8%	26.2%

Ending consonant place of articulation

There was a significant difference between prefixes and suffixes when looking at the ending consonant place of articulation with the type analysis, $\chi^2(4) = 18.82, p < .05^{\dagger\dagger}$, V = .30 and the token analysis, $\chi^2(4) = 37904.40, p < .05^{\dagger\dagger\dagger}$, V = .33. For both prefixes and suffixes the majority of ending consonants were alveolar, but the prefixes had significant numbers of bilabial, velar and labiodental consonants as the final phoneme. For the suffixes, velar consonants occurred at approximately the same level as they did in prefixes in the type analysis. However, over 95% of the suffixes in the token analysis had alveolar articulation.

Table I9Ending consonant place of articulation for prefixes and suffixes with type and token analyses

		Ending Consonant Place of Articulation				
		Alveolar	*Bilabial*	*Velar*	*Labiodental*	*Postalveolar*
Type Analysis	*Prefix (N = 47)*	61.7%	17.0%	10.6%	10.6%	0%
	Suffix (N=168)	81.0%	4.2%	11.9%	1.8%	1.2%
Token Analysis	*Prefix (N = 127663)*	77.6%	12.2%	5.8%	4.3%	0%
	Suffix (N = 230942)	95.4%	0.3%	1.4%	1.5%	1.3%

A slight prevalence for alveolar places of articulation was to be expected as six of the top nine most frequent phonemes have alveolar articulation, as mentioned previously. A further reason for the difference may be due to a tendency to have more varied places of consonant articulation within words than at the end of words. Regardless of the underlying reason, this measure was useful in differentiating prefixes and suffixes. However, it may be more appropriate to suggest that ending consonant place of articulation would be more useful when combined with other phonological measures.

Ending vowel height

There was no significant difference between prefixes and suffixes when looking at ending vowel height in the type analysis, Fisher's Exact Test = 8.95, $p = 1.0$, V = .48. However, there was a significant difference in the token analysis, $\chi^2(5) = 16424.55, p < .05^{\dagger\dagger\dagger}$, V = .40. The majority of ending vowel suffixes ended with a high vowel with the majority of these vowels being /i/ or /I/. Prefixes showed a similar, but slightly different pattern. There were still a significant amount of the high vowels indicated above, but there also appeared to be a several high diphthongs (which are indicated by the average of the individual vowels values, in this case /eI/ = (2 + 1.5)/2 = 1.75). Ending vowel height was particularly useful in differentiating prefixes and suffixes when combined with beginning consonant information. See Table I10 for a summary of the data.

Table I10 Ending vowel height for prefixes and suffixes with type and token analyses

		Ending Phoneme - Vowel Height						
		Close (1)	*1.5*	*1.75*	*Close-mid (2)*	*2.5*	*Open-mid (3)*	*Open (4)*
Type Analysis	*Prefix (N = 8)*	62.5%	12.5%	12.5%	12.5%	0%	0%	0%
	Suffix (N=48)	79.2%	12.5%	0%	0%	4.2%	0%	4.2%
Token Analysis	*Prefix (N = 68687)*	57.5%	11.7%	24.5%	6.3%	0%	0%	0%
	Suffix (N = 40042)	89.5%	10.1%	0%	0%	.4%	0%	.1%

Ending vowel position

There was no significant difference between prefixes and suffixes on initial vowel position with the type analysis, Fisher's Exact Test = 9.67, $p = 1.0$, V = .48. However, there was a significant difference for the token analysis, $\chi^2(6) = 16572.3$, $p < .05^{\dagger\dagger\dagger}$, V = .39. As with vowel height, this measure was influenced by the majority of ending vowel phonemes being either /i/ or /I/ for suffixes which meant that the front and mid front to central categories were the only categories that had substantial vowels for suffixes. As expected there were differences for the prefixes, which still showed a high amount of /i/ and /I/ vowels, but also several back vowels as well. The main difference between prefixes and suffixes was a wider distribution of position of articulation for ending vowel prefixes and a much smaller distribution for suffixes. The combined analysis indicated that this feature was not useful in differentiating prefixes and suffixes when combined with ending vowel height information; this was likely due to redundancy between the two sources.

Table I11 Ending vowel position for prefixes and vowels with type and token analyses

		Ending vowel position						
		1 (Front)	*1.25*	*1.5*	*2 (Central)*	*2.25*	*2.75*	*3 (Back)*
Type Analysis	*Prefix (N = 8)*	62.5%	0%	12.5%	0%	12.5%	12.5%	0%
	Suffix (N=48)	79.2%	2.1%	10.4%	4.2%	0%	0%	4.2%
Token Analysis	*Prefix (N = 68687)*	57.5%	0%	11.7%	0%	6.3%	24.5%	0%
	Suffix (N = 40042)	89.5%	.4%	9.9%	.2%	0%	0%	.1%

Ending vowel roundedness

The type analysis looking at ending vowel roundedness did not show a significant difference between prefixes and suffixes, $\chi^2(1) = 6.11, p = 1.0$, V = .33, but the token analysis was significant, $\chi^2(1) = 11596.29, p < .05^{\dagger\dagger\dagger}$, V = .33. The majority of ending vowels in prefixes were rounded, but approximately 25% were diphthong vowels that contained both rounded and unrounded vowels. All suffix ending vowels were rounded; there were no unrounded ending vowels at all, across both prefixes and suffixes.

Table I12 Ending vowel roundedness for prefixes and suffixes with type and token analyses

		Ending Vowel Roundedness	
		Rounded	*Diphthong-Rounded/unrounded*
Type Analysis	*Prefix (N=8)*	87.5%	12.5%
	Suffix (N=48)	100%	0%
Token Analysis	*Prefix (N=68686)*	75.5%	24.5%
	Suffix (N=40043)	100%	0%

The lack of unrounded vowels was easily explained for the suffix group, as the majority of ending vowels were the unrounded /i/ or /I/ vowels. This was also the case for all single vowels in the prefix condition; however, the prefixes also contained ending diphthongs, which were coded as combined rounded/unrounded when the vowels' roundedness did not coincide.

Appendix J

Table J Medial words for the "I_it" frame and their frequency in the current analysis, as reported in Mintz (2003) and in the original Peter session files with discrepancy figures (with potential discrepancy accounted for by analysis described in Chapter 6 in parentheses).

Frame	*Medial words*	*Mintz*	*Current*	*Mintz/current Discrepancy*	*Original corpus*
I_it	see	18	17	1 (0)	17
	think	9	8	1 (0)	8
	got	8	5	3 (0)	5
	thought	5	4	1 (0)	4
	do	4	3	1 (0)	3
	open	3	2	1 (0)	2
	did	3	2	1 (0)	2
	use	2	1	1 (0)	1
	give	2	1	1 (0)	1
	took	1	0	1 (1)	0
	leave	1	0	1 (0)	0
	knew	1	0	1 (0)	0
	close	1	0	1 (0)	0

Appendix K

Table K Medial words for the "the_one" frame and their frequency in the current analysis, as reported in Mintz (2003) and in the original Peter session files with discrepancy figures (with potential discrepancy accounted for by analysis described in Chapter 6 in parentheses).

Frame	*Medial words*	*Mintz*	*Current*	*Mintz/Current Discrepancy*	*Original corpus*
The_one	red	11	6	5 (1)	6
	yellow	8	7	1 (0)	7
	green	8	5	3 (0)	5
	orange	6	2	4 (0)	2
	blue	5	4	1 (0)	4
	right	4	2	2 (0)	2
	little	3	2	1 (0)	2
	light	1	0	1 (0)	0
	empty	1	0	1 (0)	0

Appendix L

Table L Medial words for the "you_it" frame and their frequency in the current analysis, as reported in Mintz (2003) and in the original Naomi session files with discrepancy figures (with potential discrepancy accounted for by analysis described in Chapter 6 in parentheses).

Frame	*Medial words*	*Mintz*	*Current*	*Mintz/Current Discrepancy*	*Original corpus*
You_it	like	11	12	-1 (-1)	2
	guessed	1	0	1 (0)	0
	ate	1	2	-1 (-1)	2
	pop	0	1	-1 (0)	1

Appendix M

Table M Medial words for the "the_is" frame and their frequency in the current analysis, as reported in Mintz (2003) and in the original Naomi session files with discrepancy figures.

Frame	*Medial words*	*Mintz*	*Current*	*Mintz/Current Discrepancy*	*Original corpus*
The_is	sleeper	0	1	-1 (0)	1
	floor	0	1	-1 (0)	1
	horse	0	1	-1 (0)	1
	zoo	0	1	-1 (0)	1

Appendix N

Table N Experiment 9 and 10 Generalization test sentences.

Compatible “new” vs. Compatible “old”	Compatible “new” vs. incompatible “old”
$CY_1D - CX_6D$	$CX_{14}D - CY_1D$
$CY_2D - CX_7D$	$CX_{15}D - CY_2D$
$CX_8D - CY_3D$	$CY_3D - CX_{16}D$
$GY_4H - GX_{14}H$	$GX_6H - GY_4H$
$GX_{15}H - GY_5H$	$GY_5H - GX_7H$
$GX_{16}H - GY_6H$	$GY_6H - GX_8H$
Incompatible “new” vs. compatible “old”	*Incompatible “new” vs. incompatible “old”*
$CX_7D - CY_4D$	$CY_4D - CX_{14}D$
$CY_5D - CX_8D$	$CX_{16}D - CY_5D$
$CX_6D - CY_6D$	$CY_6D - CX_{15}D$
$GY_1H - GX_{15}H$	$GX_8H - GY_1H$
$GY_2H - GX_{16}H$	$GY_3H - GX_6H$
$GX_{14}H - GY_3H$	$GX_7H - GY_2H$

References

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality Independence of Implicitly Learned Grammatical Knowledge. *Journal of Experimental Psychology: Learning, 21*, 899.

Baayen, R. H., Pipenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bates, E., Elman, J., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1998). Innateness and emergentism. In W. Bechtel & G. Graham (Eds.), *A Companion to Cognitive Science* (pp. 590-601). Oxford: Basil Blackwell.

Bauer, L. (2003a). English prefixation - A typological shift? *Acta Linguistica Hungarica, 50*, 33-40.

Bauer, L. (2003b). The phonotactics of some English morphology. In H. G. Jacobsen, D. Bleses, T. O. Madson & P. Thomsen (Eds.), *Take Danish for Instance: Linguistic studies in honour of Hans Basbøll presented on the occasion of his 60th birthday 12 July 2003* (pp. 1-8): University Press of Southern Denmark.

Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances? *Developmental Psychology, 29*, 711-721.

Billman, D. (1989). Systems of correlations in rule and category learning: Use of structured input in learning syntactic categories. *Language and Cognitive Processes, 4*, 127-155.

Black, A. W., Clark, R., Richmond, K., King, S., & Zen, H. (2004). Festival Speech Synthesizer (Version 1.95). Edinburgh: University of Edinburgh.

Bloom, L., Hood, L., & Lightbrown, P. (1974). Imitation in language development: If, when and why. *Cognitive Psychology, 6*, 380-420.

Bloom, L., Lightbrown, P., & Hood, L. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development, 40*, Serial No. 160.

Bloomfield, L. (1933). *Language*. New York: Henry Holt and Company.

Bonatti, L. L., Nespor, M., Peña, M., & Mehler, J. (2006). How to Hit Scylla Without Avoiding Charybdis: Comment on Perruchet, Tyler, Galland, and Peereman (2004). *Journal of Experimental Psychology: General, 135*, 314-321.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic Constraints on Statistical Computations: The Role of Consonants and Vowels in Continuous Speech Processing. *Psychological Science, 16*, 451-459.

Bouton, M. E. (2007). *Learning and behavior: A contemporary synthesis*. Sunderland, MA: Sinauer Associates.

Braine, M. D. S. (1966). Learning the positions of words relative to a marker element. *Journal of Experimental Psychology, 72*, 532-540.

Braine, M. D. S., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., et al. (1990). Exploring Language Acquisition in Children with a Miniature Artificial Language: Effects of Item and Pattern Frequency, Arbitrary Subclasses and Correction. *Journal of Memory and Language, 29*, 591-610.

Brooks, P. J., Braine, M. D., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32*, 76-95.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Cartwright, T., & Brent, M. (1997). Syntactic categorization in early language acquisition: formalizing the role of distributional analysis. *Cognition, 63*, 121-170.

Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language, 30*, 348-369.

Cassidy, K. W., & Kelly, M. H. (2001). Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin & Review, 8*, 519-523.

Cassidy, K. W., Kelly, M. H., & Sharoni, L. J. (1999). Inferring gender from name phonology. *Journal of Experimental Psychology: General, 128*, 362-381.

Chater, N., & Vitanyi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences, 7*, 19-22.

Chater, N., & Vitanyi, P. (2007). 'Ideal learning' of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psychology, 51*, 135-163.

Chomsky, N. (1959). Review of Verbal Behavior, by B.F. Skinner. *Language, 35*, 26-57.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1980). *Rules and Representations*. Oxford: Basil Blackwell Publisher.

Chomsky, N. (2006). *Language and Mind* (3rd ed.). New York: Cambridge University Press.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23*, 157-205.

Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences, 5*, 82-88.

Christiansen, M. H., Hockema, S. A., & Onnis, L. (2006). Using phoneme distributions to discover words and lexical categories in unsegmented speech. In *Proceedings of the 28th Annual Cognitive Science Society Conference* (pp. 172-177). Mahwah, NJ: Lawrence Erlbaum.

Christiansen, M. H., & Monaghan, P. (2006). Discovering Verbs Through Multiple-Cue Integration. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 88-107). NY, US: Oxford University Press.

Conway, C. M., & Christiansen, M. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences, 5*, 539-546.

Conway, C. M., & Christiansen, M. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31*, 24-39.

Crain, S., & Pietroski, P. (2001). Nature, Nurture and Universal Grammar. *Linguistics and Philosophy, 24*, 139-185.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant Melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology, 30*, 1119-1130.

Curtin, S., Mintz, T. H., & Christiansen, M. H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition, 96*, 233-262.

Cutler, A. (1993). Phonological cues to open- and close-class words in the processing of spoken sentences. *Journal of Psycholinguistic Research, 22*, 109-131.

Cutler, A., Hawkins, J. A., & Gilligan, G. (1985). The Suffixing Preference: A Processing Explanation. *Linguistics, 23*, 723-758.

Dienes, Z., Altmann, G. T. M., & Gao, S.-J. (1999). Mapping across domains without feedback: A neural network model of transfer of implicit knowledge. *Cognitive Science, 23*, 53-82.

Dienes, Z., Altmann, G. T. M., Kwam, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 1322-1338.

Dienes, Z., & Perner, J. (2002). A theory of the implicit nature of implicit learning. In R. M. French & A. Cleeremans (Eds.), *Implicit Learning and Consciousness: An Empirical, Philosophical, and Computational Consensus in the Making?* London: Psychology Press.

Dumay, N., & Gaskell, M. G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychological Science, 18*, 35-39.

Echols, C. H. (1993). A perceptually-based model of children's earliest productions. *Cognition, 46*, 245-296.

Ellis, N. C. (2006a). Language Acquisition as Rational Contingency Learning. *Applied Linguistics, 27*, 1-24.

Ellis, N. C. (2006b). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*, 164-194.

Elman, J. L. (1993). Learning and Development in Neural Networks: The Importance of Starting Small. *Cognition, 48*, 71.

Everitt, B. S. (1992). *The Analysis of Contingency Tables* (2nd Edition ed.). London: Chapman & Hall.

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature, 425*, 614-616.

Fischer, S., Drosopoulos, S., Tsen, J., & Born, J. (2006). Implicit Learning-Explicit Knowing: A Role for Sleep in Memory System Interaction. *Journal of Cognitive Neuroscience, 18*, 311-319.

Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39*, 218-245.

Fudge, E. (1984). *English Word-Stress.* London: George Allen & Unwin.

Garzón, F. C. (2005). Non-adjacent transitional probabilities and the induction of grammatical regularities. In *Proceedings from the 26th Annual Conference of the Cognitive Science Society* (pp. 767-772). Mahwah, NJ: Lawrence Erlbaum Associates.

Gerken, L. (1996). Phonological and distributional information in syntax acquisition. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition.* (pp. 411-425). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Gerken, L., Jusczyk, P. W., & Mandel, D. R. (1994). When Prosody Fails to Cue Syntactic Structure: 9-Month-Olds' Sensitivity to Phonological versus Syntactic Phrases. *Cognition, 51*, 237-265.

Gerken, L., Wilson, R., Gómez, R. L., & Nurmsoo, E. (2002). Linguistic category induction without reference: The importance of correlated cues: Unpublished manuscript, University of Arizona.

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language, 32*, 249-268.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431-436.

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps Promote Abstraction in Language-Learning Infants. *Psychological Science, 17*, 670-674.

Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science, 7*, 567-580.

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependecy learning. *Infancy, 7*, 183-206.

Green, T. R. G. (1979). The Necessity of Syntax Markers: Two Experiments with Artificial Languages. *Journal of Verbal Learning and Verbal Behavior, 18*, 481-496.

Hall, C. J. (1988). Integrating diachronic and processing principles in the suffixing preference. In J. A. Hawkins (Ed.), *Explaining Language Universals*. Oxford: Basil Blackwell.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition, 78*, B53-B64.

Hauser, M. D., Weiss, D., & Marcus, G. (2002). Rule Learning by Cotton-Top Tamarins. *Cognition, 86*, B15-B22.

Herbranson, W. T., & Shimp, C. P. (2003). "Artificial grammar learning" in pigeons: A preliminary analysis. *Learning & Behavior, 31*, 98-106.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*, 548-567.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences, 3*, 323-328.

Karni, A., Tanne, D., Rubenstien, B. S., Askenasy, J. J. M., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science, 265*, 679-682.

Kelly, M. (1988). Phonological biases in grammatical category shifts. *Journal of Memory and Language, 27*, 343-358.

Kelly, M. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99*, 349-363.

Kelly, M. (1996). The Role of Phonology in Grammatical Category Assignments. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 249-261). Mahwah, NJ: Lawrence Erlbaum Associates

Kelly, M., & Bock, J. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception & Performance, 14*, 389-403.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*, B35-B42.

Köpcke, K. M. (1982). *Untersuchungen zum Genussystem der deutschen Gegenwartssprache*. Tübinger: Niemeyer.

Lany, J., Gomez, R. L., & Gerken, L. (2007). The role of prior experience in language acquisition. *Cognitive Science, 31*, 481-507.

Lewis, J. D., & Elman, J. (2002). Learnability and the statistical structure of langauge: Poverty of the stimulus arguments revisited. In *Proceedings of the 26th Annual Boston University Conference on Language Development* (pp. 359-370). Somerville, MA: Cascadilla Press.

MacWhinney, B. (2000). The CHILDES Project: Tools for analyzing talk, Vol 2: The database (3rd ed.).

Maratsos, M. P., & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. E. Nelson (Ed.), *Children's Language* (Vol. 2, pp. 125-214). New York: Gardner Press.

Mattys, S., & Jusczyk, P. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*, 91-121.

McDonald, J. L. (1997). Language acquisition: The acquisition of linguistic structure in normal and special populations. *Annual Review of Psychology, 48*, 215-241.

Mednick, S. C., Nakayama, K., Cantero, J. L., Atienza, M., Levin, A. A., Pathak, N., et al. (2002). The restorative effect of naps on perceptual deterioration. *Nature Neuroscience, 5*, 677-681.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition, 30*, 678-686.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*, 91-117.

Mintz, T. H. (2006). Finding the Verbs: Distributional Cues to Categories Available to Young Learners. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 31-63). NY, US: Oxford University Press.

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science, 26*, 393-424.

Mirković, J., MacDonald, M. C., & Seidenberg, M. S. (2005). Where does gender come from? Evidence from a complex inflectional system. *Language and Cognitive Processes, 20*, 139-167.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition, 96*, 143-182.

Monaghan, P., & Christiansen, M. H. (2004). What distributional information is useful and usable for language acquisition? In *Proceedings of the 26th Annual Conference for the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

Monaghan, P., Onnis, L., Merkx, M., & Chater, N. (Submitted). Multiple-cue-driven computations in speech processing.

Moore, B. R. (1973). The role of directed Pavlovian reactions in simple instrumental learning in the pigeon. In R. A. Hinde & J. Stevenson-Hinde (Eds.), *Constraints on Learning* (pp. 159-186). London: Academic Press.

Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition.* (pp. 263-283). NJ, England: Lawrence Erlbaum Associates, Inc.

Mori, K., & Moeser, S. D. (1983). The Role of Syntax Markers and Semantic Referents in Learning an Artificial Language. *Journal of Verbal Learning and Verbal Behavior, 22*, 701.

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language, 54*, 283-299.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127-162.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology, 49*, 85-117.

Norris, D., McQueen, J. M., Cutler, A., Butterfield, S., & Kearns, R. (2001). Language-universal constraints on speech segmentation. *Language and Cognitive Processes, 16*, 637-660.

Onnis, L., & Christiansen, M. (2005). New beginnings and happy endings: Psychological plausibility in computational models of language acquisition. In *Proceedings of the 27th Annual Conference for the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Onnis, L., Monaghan, P., Christiansen, M., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 1047-1052). Mahwah, NJ: Lawrence Erlbaum.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language, 53*, 225-237.

Patterson, K. E. (1980). Derivational errors. In M. Coltheart, K. E. Patterson & J. C. Marshall (Eds.), *Deep Dyslexia*. London: Routledge and Kegan Paul.

Pearce, J. M. (1997). *Animal Learning and Cognition: An Introduction* (2nd ed.). Hove, UK: Psychology Press.

Pearce, J. M., & Hall, G. (1980). A model for Pavolvian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87*, 532-552.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science., 298*, 604-607.

Perfors, A., Tenenbaum, J., & Regier, T. (2006). Poverty of the Stimulus: A rational approach. In *Proceedings from the 28th Annual Conference for the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning Nonadjacent Dependencies: No Need for Algebraic-Like Computations. *Journal of Experimental Psychology: General, 133*, 573-583.

Pinker, S. (1994). *The language instinct: The new science of language and mind.* London: Penguin.

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology, 21*, 785-794.

Reali, F., Christiansen, M., & Monaghan, P. (2003). Phonological and distributional cues in syntax acquisition: Scaling up the connectionist approach to multiple-cue intergration. In *Proceedings of the 25th annual Conference of the Cognitive Science Society* (pp. 970-975). Mahwah, NJ: Lawrence Erlbaum.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855-863.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology, 81*, 115-119.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*, 425-469.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II: Current Theory and Research* (pp. 64-99). New York: Appleton-Century-Crofts.

Sachs, J. (1983). Talking about the there and then: The emergence of displaced reference in parent-child discourse. In K. E. Nelson (Ed.), *Children's Langauge* (4th ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Saffran, J. R. (2001a). The Use of Predictive Dependencies in Language Learning. *Journal of Memory and Language, 44*, 493-515.

Saffran, J. R. (2001b). Words in a sea of sounds: The output of infant statistical learning. *Cognition, 81*, 149-169.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172-196.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110-114.

Saffran, J. R., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27-52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language, 35*, 606-621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*, 101-105.

Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology, 60*, 291-309.

Shi, R., Morgan, J., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: a cross-linguistic perspective. *Journal of Child Language, 25*, 169-201.

Shi, R., Werker, J., & Morgan, J. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition, 72*, B11-21.

Smith, L. B. (2000). Learning how to learn words: An associative crane. In *Becoming a Word Learner*. New York: Oxford University Press.

St. Clair, M. C., & Monaghan, P. (2005). Categorizing grammar: Differential effects of succeeding and preceding contextual cues. In *Proceedings from the 27th Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Suppes, P. (1974). The semantics of children's language. *American Psychologists, 29*, 103-114.

Taraban, R. (2004). Drawing learners' attention to syntactic context aids gender-like category induction. *Journal of Memory and Language, 51*, 202-216.

Theakson, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language, 28*, 127-152.

Theissen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7*, 53-71.

Theissen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to work boundaries by 7- to 9-month-old infants. *Developmental Psychology, 39*, 706-716.

Thothathiri, M., & Snedeker, J. (2005). *Distributional learning of syntactic categories.* Paper presented at the Society for Research in Child Development Biennial Meeting, Atlanta, GA.

Valian, V., & Coulson, S. (1988). Anchor Points in Language Learning: The Role of Marker Frequency. *Journal of Memory and Language, 27*, 71-86.

Vasta, R., Haith, M. H., & Miller, S. A. (1999). *Child Psychology: The Modern Science* (3rd ed.). New York: John Wiley & Sons, Inc.

Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature, 427*, 352-355.

Winter, B., & Reber, A. S. (1994). Implicit learning and the acquisition of natural languages. In N. C. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp. 115-145). London: Academic Press.

Yu, C., & Smith, L. B. (2006). Statistical cross-situational learning to build word-to-world mappings. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum.

Yu, C., & Smith, L. B. (in press). Rapid word learning under uncertainly via cross-situational statistics. *Psychological Science.*

Ziori, E., & Dienes, Z. (in press). How does prior knowledge affect implicit and explicit concept learning? *Quarterly Journal of Experimental Psychology.*