Semantic representations of English verbs and their influence on psycholinguistic performance in healthy and language-impaired speakers

Christopher Plant

Doctor of Philosophy

School of Education, Communication & Language Sciences
Newcastle University

September 2011

Declaration

I declare that this thesis represents my own original work except for quotations and citations which have, to the best of my ability, been acknowledged. The material presented in this thesis has not previously been submitted for any other degree at Newcastle University or any other institution.

Christopher Plant

Abstract

Background – English verbs are linguistically more complex than nouns and this has contributed to the dearth of in-depth investigation into similarities and differences between their representations within semantic memory and subsequent implications for language processing. However, recent theoretical accounts have argued that verbs and nouns are represented within a unitary semantic system.

Aims – This thesis investigates the semantic representations of English verbs with particular attention to how verbs are inter-related as a consequence of semantic similarity. This is achieved through a series of psycholinguistic experiments with healthy adult speakers and an intervention study with adults with aphasia (i.e. acquired communication impairment). Throughout the thesis, comparisons are made to the semantic representations of nouns either directly (i.e. through parallel experimentation) or indirectly (i.e. through the existing literature).

Methods – The experiments conducted with healthy adult speakers included: (1) category listing of verbs; (2) typicality rating of verbs within categories; (3) similarity rating of verb pairs; (4) an analysis of verbs' semantic features; (5) category verification of verbs; and (6) semantically primed picture naming of actions. The intervention study carried out with adults with aphasia compared patterns of improvement in verb and noun retrieval following a semantically-based therapy task.

Results and discussion – The results of the experiments shed light on the nature of semantic representations of verbs, in particular, in relation to the similarity between the semantic representations of verbs and those of nouns and also where they differ. These insights are considered in terms of how they provide evidence for or against a unitary semantic system for verbs' and nouns' semantic representations and parallel mechanisms for accessing these representations. Two themes emerged in terms of future research potential: (1) the influence of polysemy on speaker's performance in psycholinguistic tasks; and (2) the nature and influence of typicality within categories/cluster of verbs.

Dedication

For Eunice and Reg, without whom this probably wouldn’t have happened.

Acknowledgements

I'm extremely thankful for the support of my parents, family and friends who have supported me in ways that are too numerous and varied to describe concisely and adequately. I have little doubt that without them I wouldn't have come so far.

Huge appreciation to Anne Whitworth and Janet Webster who supervised me in completing this project – I hope you can forgive me for the headaches that the readings and re-readings may have caused. I'd also like to thank the rest of the staff and students in the section of Speech and Language Sciences at Newcastle University who have given countless valuable comments and suggestions through lab meetings and encounters in the offices and corridors of the King George VI building. The same appreciation also goes to the speech and language therapists of the North East of England who assisted in various aspects of my work and who patiently listened to and acted upon repeated requests for assistance despite them not really having the time to do so.

Others who have played their own unique parts in getting me to where I am today include Jane Constantine, Lynne Murphy and Harald Clahsen, who between them inspired me and gave me the confidence to go to university, to train as a speech and language therapist, and to complete a PhD. I may not have been the most talkative in our encounters but I never stopped listening and learning.

I am grateful for the artistic talents of Sandra Lowing who produced numerous illustrations under exacting instruction that were used in various parts of this PhD project.

Finally I'd like to express my heartfelt thanks and appreciation to the participants who gave up their time to carry out my experiments. Special thanks to AB, AH, GF, JA, RH, SH, and WM who each dedicated a couple of hours a week for several months – I hope you gained as much from the experience as I did.

Chris, September 2011

Table of Contents

Appendices

List of Tables

List of Figures

Chapter 1 Verbs/Actions and Nouns/Objects in Semantic Memory and Language Processing.

1.1. Aims of Chapter

This chapter introduces the theoretical background that underpins the current thesis. This introduction begins with a description of organisation and processing within semantic memory as described by prominent theoretical accounts and the role of semantic memory in language processing. While this introduction does not attempt to provide an exhaustive critique of models of semantic memory and language processing (see Chang, 1986; Funnell, 2000, for reviews), it does aim to present a representative overview of models that have been proposed which frames the ideas discussed throughout the current thesis. Attention will be drawn to particular methods and research themes that have developed in the study of semantic memory. Similarities and differences between verbs and nouns are considered both in terms of semantic representations within semantic memory and also within other linguistic domains, including in speakers with language impairments. This chapter argues that further research is needed to understand the semantic representations of verbs and their implications for language processing. This argument lays the foundation of the current thesis. The chapter concludes by presenting the primary research questions that will be addressed in subsequent chapters.

1.2. Background

1.2.1. Semantic memory

Semantic memory is a subcomponent of long-term memory. Tulving (1972) gave one of the first formalised definitions of semantic memory when he described it as:

> The memory necessary for the use of language. It is a mental thesaurus, organised knowledge a person possesses about words and other verbal symbols their meanings, and referents, about relations among them, and about rules, formulas, and algorithms for the manipulation of these symbols, concepts and relations. (Tulving, 1972:386)

Tulving's definition identified semantic memory as a distinct component of declarative (or explicit) long-term memory which also consists of episodic (or autobiographical) memory. Where episodic memory is a store of information associated with specific events, semantic memory is a store of generalised information that has been abstracted away from specific events. For example, when a person remembers an encounter they had with a black cat that morning, this is stored within episodic memory, but a person's knowledge about cats, such as that they have whiskers and purr, is stored in semantic memory.

In spite of the large amount of research in the area, semantic memory as a whole is not clearly understood in terms of the major processing principles which are applicable to any memory subsystem: (1) encoding, i.e. how information is registered within semantic memory; (2) storage, i.e. how information is maintained over time; and (3) retrieval, i.e. how information is accessed from within semantic memory (see Baddeley, 2004).

1.2.2. Models of semantic memory

Since the 1960s, there have been numerous theoretical models developed that have attempted to describe how concepts (i.e. representational units within semantic memory) are encoded, stored and retrieved within semantic memory. This section aims to provide a brief overview of some of the different approaches that have been taken.

Hierarchical models

Collins & Quillian (1969) proposed one of the first formal theories of semantic memory. They stated that concepts within semantic memory are organised hierarchically so that general concepts are stored higher than more specific, yet related concepts. Therefore, *animal* would be stored higher than *bird* which itself would be stored higher than *canary* (see Figure 1.1). Concepts are associated with semantic features and these features are inherited from related concepts that are higher in the hierarchy. Therefore, individual features are only stored at the hierarchical level at which they become distinctive and stop being shared by all subordinate (i.e. lower level) concepts. For example, features that are distinctive of canaries, such as being yellow, are represented at the level of *canary*, whereas features that are shared between canaries and other birds, such as having wings, are represented at the level of *bird*. Such organisation was argued to provide cognitive economy as redundant information would not be represented at multiple levels (e.g. a feature such as *<has wings>* is not represented both at the level of *bird* and also at the level of *canary*). The theory also accounts for exceptions to inheritance principles by allowing the inclusion of negative features, such as *<can't fly>* for ostriches, which would otherwise inherit the feature *<can fly>* from bird.

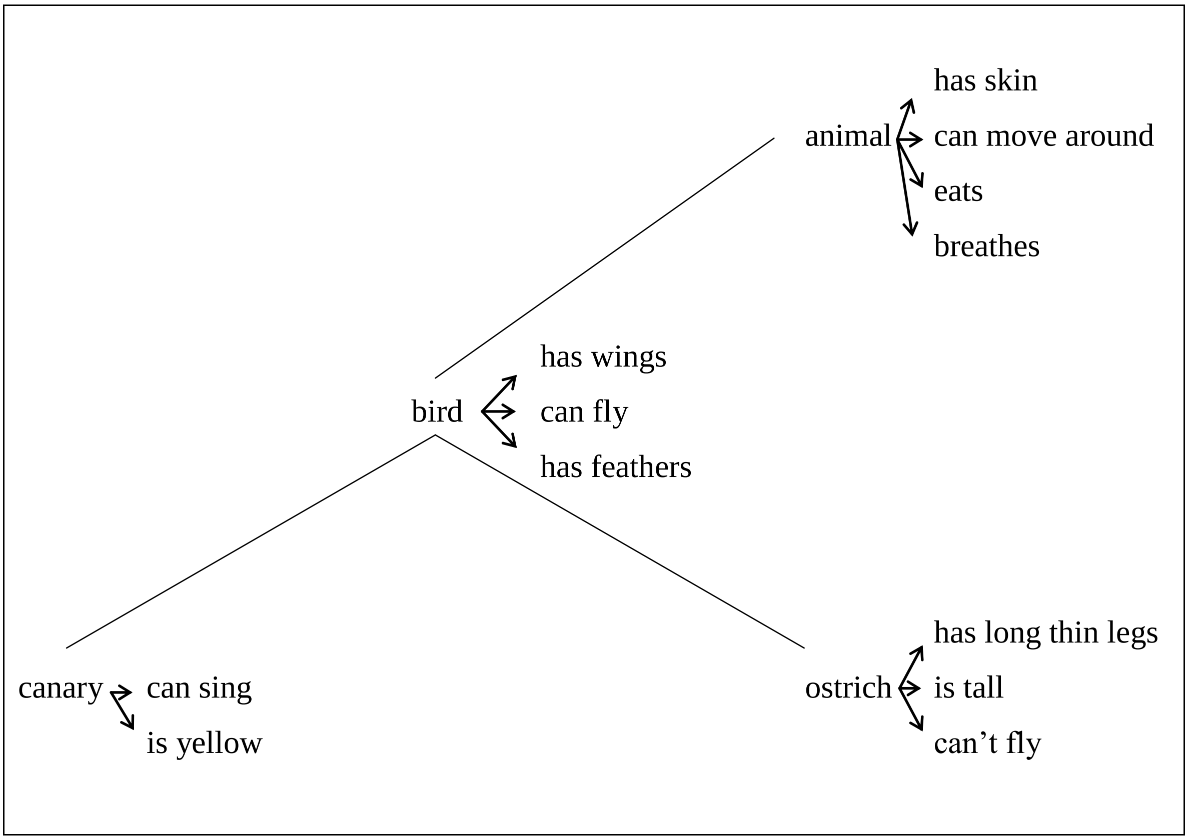

Figure 1.1 Hierarchical organisation of semantic memory (Collins & Quillian, 1969)

Collins & Quillian (1969) provided evidence for hierarchical organisation by finding that participants are slower to verify statements that involved traversing more levels of hierarchical structure. In their experiments, participants were fastest to verify property statements such as *a canary can sing* (M ≈ 1305 msec) compared to *a canary has wings* (M ≈ 1385 msec) which was itself verified faster than statements such as *a canary has skin* (M ≈ 1470 msec). The same was also true when participants were asked to verify categorical status, such as when participants verified statements such as *a canary is a canary* (M ≈ 1000 msec) compared to *a canary is a bird* (M ≈ 1165 msec) and also *a canary is an animal* (M ≈ 1240 msec). These differences in verification time were argued to reflect that search processes within semantic memory are carried out on a level-by-level approach whereby a search for relevant information begins at the lowest possible level and searching can only proceed to the next higher level when searching at the lower level has been exhausted.

Feature comparison models

Smith, Shoben & Rips (1974) proposed a feature comparison model of semantic memory to account for performance in category verification tasks (e.g. when verifying that *a robin in a bird*). The model was developed on the assumption that categories specify defining and characteristic features. In order for a concept to be considered a member of a particular category it must possess all defining features (as specified by the category). In comparison, possession of characteristic features determines a concept's typicality within the category, i.e. the more characteristic features that a concept possesses the more typical, or representative, it is of the category. Therefore, the category *bird* may specify defining features such as <*has two legs*> and <*has wings*> and may specify characteristic features including <*can fly*> (given that not all birds fly) and <*perches in trees*>. Consequently, as *robin* possesses the defining and (presumably) most of the characteristic features specified by *bird*, it is a typical member of the *bird* category.

Category verification was suggested to occur within a two-stage model whereby the first stage compares the target concept with the target category in terms of all features possessed. This initial comparison generates a similarity score, *x*, representing the number or proportion of features that are shared between the target and category concepts. If *x* exceeds a pre-specified higher critical-value then the target concept is verified as a member of the category. If *x* falls below a lower critical-value then the target concept is rejected as a member of the category. If *x* falls between the higher and lower critical values then the second stage of comparison is required. The second stage seeks to ascertain only whether defining features specified by the category are present in the target concept. If the target concept possesses all defining features specified by the category then it is verified as a category member. Alternatively, if the target concept does not possess all defining features then it is rejected as a category member.

Smith et al's (1974) model correctly predicts that typical category members are verified as category members faster than atypical category members. This is because typical category members obtain a high similarity score in the first stage of comparison as they share a greater number of features, including characteristic features, with the category. Typical category members are therefore verified following just the first stage of comparison. However, atypical category members share fewer, if any, characteristic features with the category and obtain a lower similarity score and are more likely to require the second stage of comparison. Other researchers (e.g. Hampton, 1979) have however argued that the same predictions regarding typicality and category verification speed can be supported within a single stage of comparison which derives an overall similarity score between the category and target concept. Such arguments arise from the suggestion that it is incorrect to assume that features are marked as either defining or characteristic as similarity is based according to principles of family resemblance (e.g.

Rosch & Mervis, 1975) rather than according to the classical tradition of 'necessary and sufficient' features for category membership (e.g. see Smith & Medin, 1981, for a review of classical approaches to categorisation).

Spreading activation models

Collins & Loftus (1975) proposed a structure to semantic memory which consists of a network of interconnected nodes (e.g. Figure 1.2), each of which represents a concept. By virtue of being a unit of meaning, semantic features are also represented as nodes, just as complete concepts are. These semantic networks allow activation to spread between related concepts as concepts are inter-linked within the network. For example, the conceptual nodes representing *cat* and *dog* would be interconnected via conceptual nodes representing features such as <*a domesticated animal*>, <*has a tail*>, <*has paws*>, and so on. The links between nodes vary in terms of distance and weight according to the strength of association between the concepts. Semantic features that are strongly associated with a particular concept are stored closer to the concept than features that are more weakly associated and they are also activated to a greater extent as the weight of link is stronger. Such networks therefore employ Hebbian principles of learning whereby 'cells that fire together, wire together' (Hebb, 1949) where linkage weights and distances are determined by previous experiences (e.g. frequency of co-occurrence).

Evidence for spreading activation within semantic networks was argued to come from findings within semantic priming experiments (e.g. Meyer & Schvaneveldt, 1971). These demonstrated that participants are quicker to read aloud or make lexical decisions to written words when they are preceded by semantically related words. Participants are faster to verify that *doctor* is a legal English word when it is preceded by the word *nurse*, compared to when preceded by an unrelated word such as *bread*. Therefore, subsequent word recognition of related words was argued to be facilitated due to residual activation within the semantic network making it easier to achieve threshold activation of the newly presented related target word.

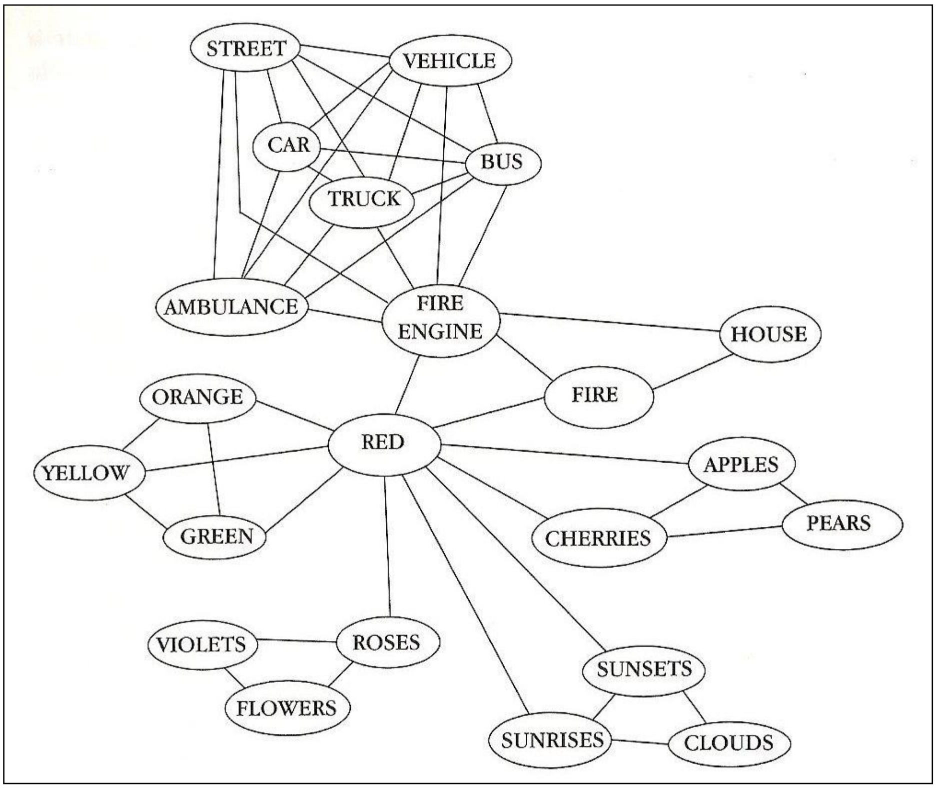

Figure 1.2 Semantic networks in semantic memory (Collins & Loftus, 1975)

Distributed models

Allport (1985) described a model of semantic memory that attempted to reflect the fact that different neuroanatomical areas show discrete activation for different sensory experiences (e.g. visual, auditory, tactile, and so on). Phonological and orthographic word-forms are activated as a result of unique and distributed patterns of activation across all sensory domains. Allport gives the example of *telephone* (i.e. Figure 1.3) which is represented by activation in visual and tactile domains which encode for shape, surface texture and size, in addition to activation in auditory and action domains as a result of physical manipulation and the use of telephones. This unique pattern of activation in sensory domains then leads to activation of the associated orthographic word-form (i.e. ‘telephone’) and/or the associated phonological word-form (i.e. /teləfəʊn/) (at least in tasks which require explicit naming).

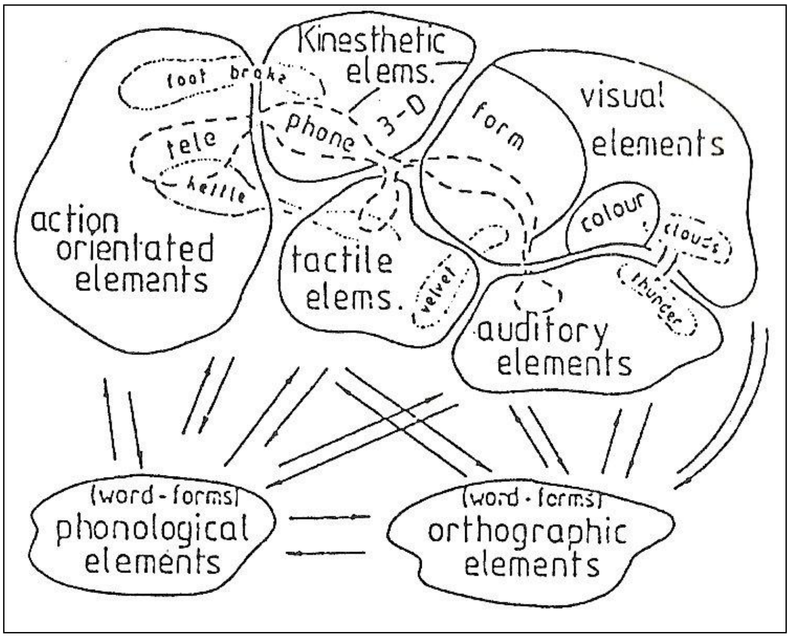

Figure 1.3 Distributed sensory semantic memory (Allport, 1985)

Allport (1985) argued that such a formulation of semantic memory accounts for the observation that object concepts are fairly resistant to localised damage as their representations are spread across various domains. In contrast, concepts represented by activation in fewer domains, are more susceptible to localised damage (e.g. colour knowledge which is represented principally by visual elements). A similar proposal has also come from Coltheart et al (1998) whereby the activation from different direct sensory experience subsystems to word forms is mediated via a more conceptual level of representation (i.e. 'knowledge stores').

Property-based models

Following speculation in reports of category specific deficits (e.g. Warrington & McCarthy, 1987; Warrington & Shallice, 1984) researchers began to model semantic memory in terms of differential representation of concepts according to different feature types (i.e. the type of information that a feature represents). Much of this research applied the division of sensory and functional features within semantic memory, especially in relation to their differential significance with living and non-living things. For instance, Farah & McClelland (1991) developed a parallel distributed processing model of semantic memory whereby living and non-living concepts were represented by patterns of activation across nodes representing either sensory or functional features. Living things were represented by a greater ratio of sensory:functional features compared to non-living things. It was found that damage (simulated by removing links between features and concepts) to sensory features (i.e. knowledge about what something looks like) within the simulated semantic system, impaired the model's ability to name living things to a greater extent than its naming of non-living things. The reverse was true when functional features were impaired (i.e. knowledge of what something does or how it is used) in that the model's ability to name non-living things was impaired relative to living things.

The sensory:functional feature division is also supported by Tyler & Moss (1997) and Moss, Tyler & Jennings (1997) although they suggested further fractionation within types of information. They suggested that functional features could be further subdivided in order to represent different aspects of function. They also proposed that sensory and functional features are not necessarily dissociable as may be implied by Farah & McClelland's (1991) model and that in fact inter-correlations between features are more influential. Through their own simulations they found that differentially changing the correlational weightings between features (i.e. the strength to which features were associated and tended to co-occur) also led to observed patterns of dissociation between living and non-living concepts. The influence of feature intercorrelation was also demonstrated by Devlin, Gonnerman, Anderson & Seidenberg (1998) who, again used a simulated model, to show that category specific deficits can arise from both widespread and focal damage. They also showed how increasing the severity of damage can reproduce patterns of category specific deficit observed in people with Alzheimer's disease, e.g. with an initial deficit for artefacts followed by a selective deficit for natural kinds (i.e. living things) as severity of damage was increased.

Computational models

Following from investigations such as those of Farah & McClelland (1991), contemporary research has sought to further employ computational and simulated models of semantic memory. O'Connor, Cree & McCrae (2009) developed a feature-based attractor network developed from semantic feature norms (i.e. McRae, Cree, Seidenberg & McNorgan, 2005).The model included superordinate (e.g. *vegetable*) and basic-level (e.g. *carrot*) concepts and each concept was associated with a pattern of activation across a set of feature nodes. No explicit hierarchical structure was incorporated into the model (hence it being described as 'flat'), however, the model was able to simulate human performance in behavioural tasks which require some degree of knowledge of hierarchical relations such as category and feature verification and semantic priming tasks.

Vigliocco, Vinson, Lewis & Garrett (2004) also reported the development of a simulation of semantic memory based on a set of semantic feature norms (Vinson & Vigliocco, 2008). This was developed according to their Featural and Unitary Semantic Space (FUSS) hypothesis and is significant in that it is the first attempt to extend modelling of semantic memory beyond the domain of nouns that refer to objects. The FUSS hypothesis proposes that object and action concepts are stored within the same unitary semantic system and that they are represented as patterns of activation across semantic features. Therefore, the retrieval of lexical forms (i.e. nouns and verbs) is triggered from activation originating from the same semantic system. The FUSS model has been applied and has been shown to replicate apparent grammatical class dissociations following simulated lesioning (Vinson & Vigliocco, 2002) and also to mirror naming latencies in picture naming of objects and actions (Vigliocco, Vinson, Damian & Levelt, 2002).

1.2.3. Semantic memory in language processing

Semantic memory, in some form or other, is an integral component of all proposed cognitive neuropsychological theories and models of language processing. For the purposes of this thesis, the adapted version of the Patterson & Shewell (1987) model of language processing presented in Whitworth, Webster & Howard (2005) will be adopted (see Figure 1.4). Three primary reasons motivate the adoption of this model in preference to other models of language processing. Firstly, the model represents the current state of knowledge with regard to language processing in that it can be seen to have developed from, and combines aspects of, previously developed models. For example, the roots of this model can be traced back to initial proposals of Morton's (1969) logogen model, which was itself revised by Morton & Patterson (1980) and further developed by Patterson & Shewell (1987). Secondly, this model attempts to present a representation of language processing which accounts for the majority of patterns of impairments to language that have been reported in the research literature investigating people with aphasia. Thirdly, this model is widely employed by clinicians (i.e. speech and language therapists) and researchers who work with people with language impairments (see Whitworth et al, 2005). As the model has potential to account for most impairments to language that occur in clinical populations, it allows hypotheses and descriptions to be made regarding which processing components and pathways have been compromised. This is therefore the most relevant model to adopt in relation to Chapter five of this thesis which presents an intervention study for participants with aphasia.

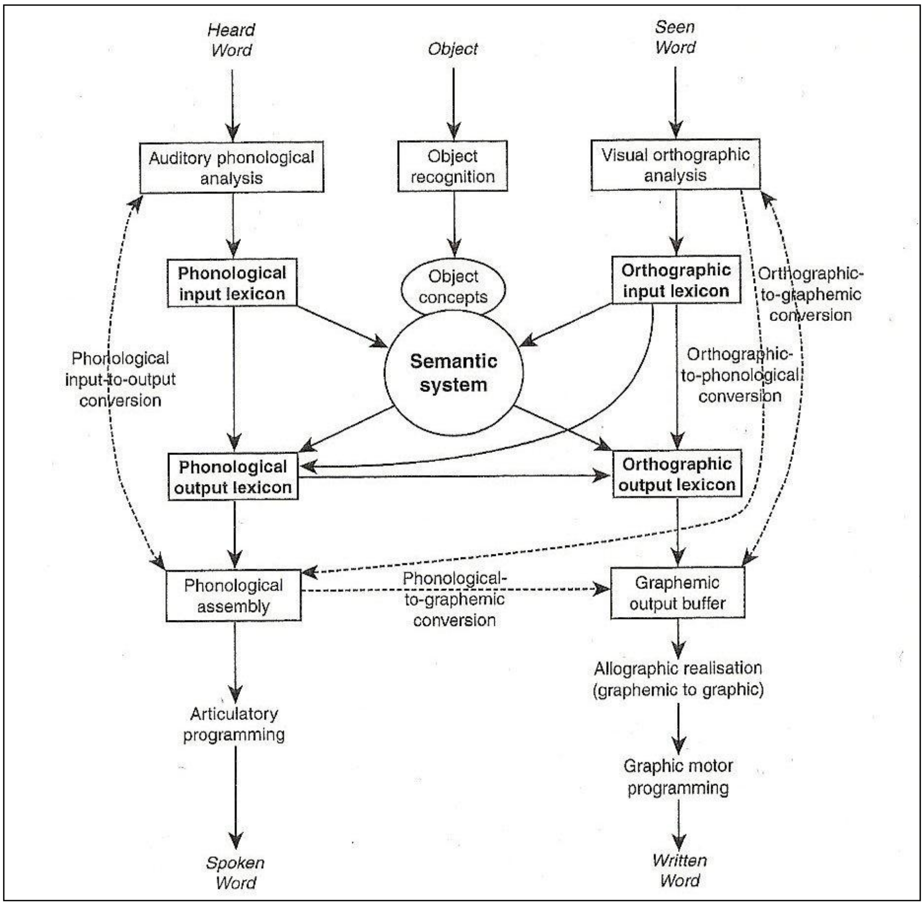

Figure 1.4 Cognitive neuropsychological model of single-word processing (Whitworth et al, 2005)

While the model is acknowledged to be underspecified in terms of the processing that occurs within each of the components (Whitworth et al, 2005), in this thesis, the semantic system will be assumed to be synonymous with, or at least include, semantic memory as has been previously discussed (see section 1.2.2). As with models of semantic memory, this model of language production has been developed with an almost exclusive focus of accounting for single-word processing associated with nouns. Therefore, at present it is not known as to whether all components and processes within this model can adequately account and represent processing of other major words classes, including verbs. The potential for such models to account for verb processing is of particular relevance as tasks utilising aspects of verb comprehension and production, at the level of single words, are frequently employed in research with healthy speakers and in research and clinical practice with speakers with language impairment (e.g. aphasia). Verb processing has been considered within other models of sentence level processing (e.g. Garrett, 1982), however, as other research into semantic memory appears to now be converging on a unitary semantic system (unitary at least in terms of word class), then it would appear logical that the validity of using such models of single-word processing in relation to verbs (i.e. Patterson & Shewell, 1987; Whitworth et al, 2005) should be considered.

1.2.4. Language impaired speakers as a window into semantic memory

Disruption to semantic memory can have a profound impact on language processing and performance (see Garrard, Perry & Hodges, 1997; Snowden, 2002, for descriptions). For models of semantic memory and language processing in general to be valid they not only need to account for normal language processing but they also need to be able to explain deficits in language processing. This necessarily implies that the investigation of language impairment can give insight into the operations of semantic memory and can validate or pose new questions for current theory. Allport (1985) has suggested that breakdown in semantic memory can have various effects on language processing, including: slower and more errorful word retrieval, incomplete and misordered word retrieval (i.e. paraphasias), misselection of words (e.g. semantic errors), impaired performance in both comprehension and production of words, and even permanent loss of specific words. A specific example of semantic memory impairment includes Alzheimer's Disease, where semantic disturbance manifests as a disorganisation of semantic memory leading to poor performance in category fluency tasks (e.g. *'name as many animals as you can think of'*) and production of semantic errors in naming (e.g. naming a picture of a *giraffe* as an *elephant*). Those with semantic dementia suffer from a loss of semantic information which makes it difficult to make fine-grained distinctions (e.g. they may be able to discriminate a picture of an *animal* from a picture of a *tree*, when asked to identify the *animal*, but they may not be able to discriminate between two animal pictures when asked to identify the *large animal* as opposed to a *small animal*; e.g. Warrington, 1975). There are then people with herpes simplex encephalitis who often show category-specific semantic deficits as indicated by reduced performance in both comprehension and production of names of items from particular semantic groups, most commonly of living things. There are also cases where semantic disruption is associated with aphasia attributable to cerebrovascular accident

(CVA) and which co-occur with disruption to other aspects of language processing. Such disorders may be characterised by poor performance in various tasks across modalities (i.e. visual recognition, written and auditory comprehension, and spoken and written production) with semantic errors or a large proportion of 'don't know' responses which may indicate damage or loss of semantic representations. Semantic disruption in aphasia can also give rise to category-specific deficits such as those seen in patients with herpes simplex encephalitis. However, in comparison to herpes simplex encephalitis, the variety of impaired categories shows greater variation in speakers with aphasia (e.g. see Capitani, Laiacona, Mahon & Caramazza, 2003; and Mahon & Caramazza, 2009, for reviews).

1.2.5. Methods for researching semantic memory in healthy speakers

A contentious issue in research into semantic memory and semantic representations, particularly with regard to investigations using healthy participants, is the nature of the research and the generalisations that are made from it. Unlike investigations with language impaired speakers which give individual insights into semantic memory, investigations with healthy speakers rely on observations averaged across groups of speakers. Researchers have made inferences and assumptions about the nature of semantic processing based on a number of experimental techniques, the majority of which require additional levels of lexical processing or rely on participants to consciously report what they believe to be salient semantic information:

> Given that we cannot observe the content of mental representations directly, we draw inferences about their likely nature on the basis of empirical observations of conceptual processing over a range of tasks and stimuli, from both healthy and impaired language users. (Moss, Tyler & Taylor, 2007:219)

Tasks that have been used to investigate semantic memory include those which can be described as 'offline' whereby participants have a degree of conscious awareness and control over their response behaviours. One example of this is category listing (e.g. Battig & Montague, 1969) where participants are required to list things belonging within a category on the assumption that this reveals organisational principles that divide the lexicon into discrete domains. Other listing tasks have included those where participants are asked to list attributes or features that belong to particular objects and/or actions (e.g. McRae et al, 2005; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Vinson & Vigliocco, 2008) which are assumed to give insight into the composition of individual concepts. Other offline tasks include those where participants are asked to rate items along a particular dimension. Dimensions that are frequently rated are typicality (e.g. Hampton & Gardiner, 1983; Rosch & Mervis, 1975) which is taken to reveal insights into within-category organisation, and semantic similarity (e.g. Romney, Brewer & Batchelder, 1996) which attempts to give a measure as to how two concepts are perceived to be similar in terms of their meaning.

In comparison to offline tasks, 'online' tasks are assumed to reflect more automatic processing as participants have a lesser degree of awareness and opportunity to be selective in their responses. Such tasks most usually involve a measure of reaction or response time. Category verification tasks, where participants are required to verify the truth of statements or whether a categorical relationship exists between pairs of words that there exists a categorical relation between pairs of words (e.g. *a robin is a bird*; *robin-bird*), are assumed to reflect the strength of association from one concept to another. Here, faster response times are taken to indicate that the two concepts share a close association (e.g. Larochelle & Pineau, 1994; Smith et al, 1974). Another variation of the verification task is where participants are required to judge whether particular features belong to a concept (e.g. bird-wings). Such tasks have been used to provide evidence of strength of association between semantic features and concepts (e.g. Collins & Quillian, 1969; McRae, Cree, Westmacott & De Sa, 1999). A further major on-line paradigm encompassing numerous tasks, that has been used to investigate semantic memory is priming. Experiments involving priming require participants to respond to a stimulus, either verbally by naming pictures, or by pressing a button to indicate yes/no decisions (e.g. a lexical decision task). Participants make a response to the target stimulus having previously been exposed to a prime stimulus. The assumption being that if the prime and target share a sufficiently strong semantic relation then the response time to the target will be facilitated (i.e. quicker) or inhibited (i.e. slower), depending on the exact nature of the task. These facilitation or inhibition effects are ascertained when comparing response times to target stimuli primed by an unrelated stimulus (e.g. see Meyer & Schvaneveldt, 1971, McNamara, 2005 and Neely, 1991 for reviews of semantic priming).

Regardless of whether the task is offline or online, researchers make generalisations about processing associated with semantic memory based on response behaviours that are mediated via lexical processing: "words are used to stand for concepts, and it can be argued that the results reflect facts about word meanings rather than about concepts" (Hampton & Bubois, 1992:28). This therefore raises a question of validity in terms of whether such research is actually informative of pure semantic processing. There is little that can be done to resolve this situation at the present time (although recent advances in technology that allows imaging of the brain during processing tasks may offer opportunities). In addition to researchers needing to make generalisations from lexical processing to semantic processing, another issue is that such tasks may not give a full representation of the content of semantic memory as participants may give conscious responses about information that is encoded subconsciously. This is something that researchers have variously stressed: "it would be a mistake to assume that people had the ability to read and report their mental representations of concepts in a veridical manner" (Medin, 1989:1473); "because people convey their conceptual knowledge through a linguistic filter, some types of information are transmitted more clearly than others" (McRae et al, 2005:549); and "what people can tell you about their internal state is highly limited, subject to strong situational biases, and may be wildly inaccurate" (Hampton & Bubois, 1992:24). These sentiments are particularly relevant within listing tasks, and particularly in semantic feature listing where speakers tend to bias their responses towards features that are distinctive and not 'obvious' (e.g. Cree & McRae, 2003; Medin, 1989). For example, few people would list features such as <*has skin*> and <*breathes*> for a tiger and would have a preference to give features such as <*has stripes*> and <*growls*>. Despite the potential problems that it may raise, the investigation of semantic memory using such tasks as described above is insightful as they all do clearly involve aspects of semantic processing (see Cree & McRae, 2003 for discussion).

As semantic memory is investigated via lexical processing, this often leads to a conflation of terminology between terms used for different levels of processing. For example, where the terms *concept*, *object* and *action* refer to representations at a conceptual-semantic level of processing, the terms *word*, *noun* and *verb* refer to representations at a lexical level of processing. There has been a tendency for previous research to conflate these terms so that *concept* and *word*, *object* and *noun*, *action* and *verb* are treated as virtual synonyms. Therefore, much research has claimed to be investigating semantic memory by conducting investigations with nouns under the assumption that there is a one-to-one mapping between objects at a conceptual-semantic level of processing and nouns at a lexical level of processing (see Vigliocco & Vinson, 2007, for discussion). For the most part, in previous research, this has not been a serious confound as generally research has investigated semantic memory using nouns that do refer to objects. However, recent research indicates a need to be careful in using these terms as there is not necessarily a one-to-one mapping. It has been shown that nouns that refer to events often pattern, in terms of semantic composition, closer to verbs referring to events than nouns referring to objects (e.g. Vinson & Vigliocco, 2002). Therefore, this thesis will attempt to keep a clear distinction between levels of processing (i.e. between conceptual-semantic and lexical levels of processing) with its use of terminology although it is accepted that there will be occasions where the distinction can become blurred.

1.2.6. Linguistic differences between verbs and nouns

There is little doubt that verbs and nouns have a great number of linguistic and pre-linguistic (i.e. conceptual) differences which influence behaviour in many ways. In English, nouns are acquired before verbs in the normal course of language acquisition (e.g. Gentner, 1982). Children with Specific Language Impairment (SLI) use nouns more frequently and verbs less frequently and to use a smaller repertoire of verbs compared to both age-matched and language-matched non-SLI children (e.g. Conti-Ramsden & Jones, 1997). In populations with aphasia, there have been numerous reports of a so-called noun-verb dissociation where there is a relative preservation of ability with nouns compared to verbs (e.g. Caramazza & Hillis, 1991; McCarthy & Warrington, 1985; Miceli, Silveri, Villa & Caramazza, 1984), although the reverse pattern has also been observed (e.g. Bi, Han, Shu & Caramazza, 2007; Zingeser & Berndt, 1988). There is also a growing body of research which highlights the probability that noun processing and verb processing are associated with differing patterns of neural activity (e.g. Damasio & Tranel, 1993; Luzzatti, Aggujaro & Crepaldi, 2006; Shapiro & Caramazza, 2003).

There are already extensive reviews of the linguistic differences between nouns and verbs, all of which conclude that verbs are more complex than nouns (e.g. Black & Chiat, 2003; Druks, 2002; Marshall, 2003). The greater complexity of verbs compared to nouns can be observed in phonology, morphology, and syntax in addition to aspects of conceptual and semantic encoding and representation. Phonologically, verbs tend to oppose the default stress pattern of English with stress being placed on the final syllable rather than the first. Nouns also tend to be longer, both in terms of number of syllables and duration even when number of syllables is equivalent. These differences lead to nouns being more perceptually salient than verbs. Verbs, in English, are morphologically more complex as they inflect for tense (past/present) and aspect (perfective/progressive), whereas nouns only inflect for plurality (singular/plural). In relation to the conceptual-semantics of verbs, they are argued to have a looser conceptual fit with their real word referents in comparison to nouns whose real world referents tend to be objects, or are at least more concrete than the actions or states being referred to by verbs. Nouns therefore tend to be tangible and atemporal (i.e. static and unchanging as time passes) and refer to a single entity whereas verbs tend to refer to events which are temporal (i.e. dynamic and changing as time passes) and which express relations between different entities.

Perhaps the most obvious distinction between verbs and nouns concerns their role within sentences and syntactic processing. Verbs are central to sentence level language processing (e.g. Garrett, 1982) and impose constraints on the number and types of arguments, or thematic roles that can co-occur with them in a sentence. Verbs are associated with canonical patterns of transitivity whereby they may be mostly used with either a single subject argument/thematic role (i.e. intransitive), with subject and object arguments (i.e. transitive), or with subject, direct object and indirect object arguments (i.e. ditransitive). This distinction is significant as it has been observed that speakers with aphasia may have greater difficulty producing verbs (even as single verbs isolated from sentence production) as the canonical number of associated arguments increases (e.g. Kim & Thompson, 2000). This advantage for naming syntactically less complex verbs has also been observed throughout normal language acquisition (e.g. De Bleser & Kauschke, 2003). Therefore, given that, in English, nouns do not have predicate argument and thematic structure in the same manner as verbs, it is understandable that verbs are considered more complex than nouns from a syntactic perspective.

Given the linguistic differences between verbs and nouns, it might be assumed that verbs and nouns as grammatical categories are fairly distinct and separate. This was originally the view of those who observed verb and noun dissociations in speakers with language impairment (e.g. Caramazza & Hillis, 1991) whereby it was argued that grammatical class was encoded alongside a word’s semantic and phonological representations. As a consequence of differential encoding of grammatical class, this led some to argue that verbs and nouns are organised semantically with a fundamentally different architecture (e.g. Huttenlocher & Lui, 1979; Graesser, Hopkinson & Schmidt, 1987). It was argued that, whereas nouns were organised according to two-dimensional hierarchical principles which branched from general to more specific nouns, verbs were organised within a matrix-like system featuring a complex web of interconnections within a three-dimensional semantic space rather than the two-dimensions used within noun organisation.

It is now more commonly accepted that verbs and nouns are more likely to fall along a continuum of complexity which is dictated by a number of factors (e.g. see Black & Chiat, 2003). It has been suggested that the so-called noun-verb dissociation is eliminated in speakers with aphasia if nouns and verbs are matched for imageability (i.e. semantic richness; e.g. Bird, Howard & Franklin, 2003) and argued that the noun-verb dissociation is a reflection of the fact that verbs tend to be more abstract and less imageable than nouns. A similar proposition was put forward by Vinson & Vigliocco (2002) who constructed a simulated unitary model of semantic space for nouns and verbs based on featural composition. Within their model, Vinson & Vigliocco included nouns referring to both objects and also events (e.g. *the destruction*) in addition to verbs referring to events (e.g. *to destroy*). It was found that the nouns referring to events had more in common, with regards to semantic featural composition, with verbs referring to events than to nouns referring to objects. This again suggests that the noun-verb dissociation is conceptual-semantic based rather than purely grammatical-syntactic.

1.3 Rationale, Research Question and Thesis Structure

1.3.1 Rationale and central argument of the current thesis

This chapter has introduced aspects of semantic and language processing and drawn attention to typical methods that have been used to investigate these. Many of the methods employed have focused on the semantic representation of objects via the lexical processing of nouns to the neglect of parallel investigations in action representations via the processing of verbs. Despite the numerous linguistic differences between verbs and nouns, there is evidence that actions and objects may populate the same semantic space and hence, verb and noun processing may be underpinned by a single semantic system which may impose similar organisational structure and processes across the semantic and linguistic repertoire of concepts and words. However, investigation into the semantic representations underpinning verbs has mainly focused on inferences gathered from sentential level semantic properties concerning core meanings and how they dictate and express relationships and interactions between syntactic arguments and conceptual-semantic thematic roles (e.g. Jackendoff, 1983; Levin, 1993; Pinker, 1989). This approach has diverged from the methods that have been used to investigate the semantic representations that underpin noun processing (as described in section 1.2.5).

This thesis argues that further investigation is needed into the semantic representations underpinning verb processing. In particular, further investigation is needed to clarify whether the same organisational principles underpin storage of verbs and nouns thus providing evidence for or against a unitary semantic system. This thesis does not argue against the validity of sentential level approaches to investigating semantic representations underpinning verbs, as verbs are clearly crucial to sentence level processing (e.g. Garrett, 1982). However, it is currently not known as to whether this is the only semantic information that is relevant for verb processing. This thesis therefore reports the use of a variety of psycholinguistic investigations with healthy speakers and an intervention study for speakers with impaired language in order to identify similarities and differences that occur in the processing of verbs and nouns.

1.3.2 Research question and thesis structure

The principle research question that this thesis addresses is:

1) To what extent are actions/verbs and objects/nouns represented and accessed from a unitary semantic system according to similar principles?

This question is addressed through evidence derived from a combination of experimental investigations that are presented in the following four chapters (i.e. chapter two to five). These four chapters address further research questions that aim to probe more specific aspects regarding the nature of the semantic representations, and access to these, of actions/verbs. In probing these questions, the following chapters aim to make direct comparisons between the semantic representations of actions/verbs to the representations of objects/nouns. Chapter six then concludes the thesis with a general discussion of the findings from the previous chapters and their implications in addressing and informing a response to the principle research question.

Chapter 2 Exploring Categorisation and Typicality of Actions/Verbs[1]

[1] Parts of this chapter, including data, have been reported in Plant, Webster & Whitworth (2011)

2.1. Aims of Chapter

This chapter presents investigations into semantic categorisation of actions/verbs and their typicality within semantic categories. This is done through two experimental tasks with healthy adult speakers of English: (1) category listing of verbs and nouns within semantic categories; (2) typicality rating of verbs within semantic categories. Such tasks have been routinely employed to investigate semantic categorisation and organisation of objects and nouns. There has been some limited extension of category listing into the domains of events and actions but there has been no extension of typicality rating into this domain. Such tasks, when they began to be conducted with objects/nouns in the 1960s and 1970s, formed the foundations of investigations into the content and organisation of semantic memory and these tasks are still relevant within contemporary research into semantic memory.

As chapter one argued, there has been much research which has investigated the numerous differences between actions/verbs and objects/nouns and their subsequent implications for language processing. This should not imply that there are not potentially similarities between actions/verbs and objects/nouns, however, this is an area that is greatly under-represented in the research literature. This chapter therefore explores whether parallel methodologies can be used to identify any similarities as well as differences in the semantic representations of actions/verbs and objects/nouns and does so through tasks used to explore the categorical nature of semantic organisation.

This chapter begins with a discussion around theories and principles of categorisation that have developed, primarily as a consequence of using experimental tasks such as category listing and typicality rating. Much of this discussion focuses on how such theories and principles have been developed in relation to objects/nouns although there are occasions where this has extended into the domains of 'ad-hoc' categories, events, and actions and these are discussed accordingly. The discussion will also cover issues around how these methods have been applied and draws attention to possible alternative data sets containing verbs, from which, some categorical principles for actions/verbs may be inferred, although this will be accompanied by discussion of their limitations. This discussion therefore aims to justify the need for further investigation into semantic categorisation within the domain of actions/verbs using similar methods that have been used to investigate objects/nouns. The chapter then presents a description of the methods and results of the two experimental investigations carried out to investigate categorical organisation and principles of actions/verbs. The main findings of the current investigations are then discussed and potential limitations and possibilities for further research are also considered.

2.2. Background

2.2.1. Theories of categorisation

Theories of categorisation attempt to explain how and why individual concepts can be perceptually grouped together. The classical view of categorisation developed from philosophy and has been attributed as far back as Aristotle (see Murphy, 2002, for discussion). In their strictest form, classical theories claim that categories are defined by necessary and sufficient attributes, or features, and that possession of these features is enough to ensure that a concept can be considered a category member. This implies that categories have clear boundaries and that all category members hold equal status within the category.

Classical views on categorisation were questioned during the 1960's and 1970's as new probabilistic theories emerged. Smith & Medin (1981) and Medin (1989) describe how probabilistic theories can be broadly subdivided into prototype-based approaches and exemplar-based approaches. Common to these approaches is the process of comparison of concepts to a category ideal to establish whether the concept is a member of the category. Comparisons are made on the basis of shared attributes and features in order to ascertain the concept's similarity to this category ideal. In prototype-based approaches, the category ideal is a summary abstraction (i.e. the prototype) of the most typical category members. Therefore, the prototype itself is not an actual category member. In exemplar-based approaches, the category ideal is based on comparisons to confirmed category members. Probabilistic theories address some of the arguments directed against classical approaches. For example, as stated earlier, classical approaches assume category members to be equal. However, typicality effects suggest that categories have a graded structure. Participants are faster to verify categorisation statements such as '*a robin is a bird*' where a robin represents a highly typicality type of bird, compared to '*a penguin is a bird*' where a penguin is a less typical type of bird (e.g. Smith et al, 1974). Probabilistic theories also better account for unclear cases where research has shown that participants are inconsistent in their categorisation decisions of low-typicality members on different test occasions and that there is greater between participant variation in categorisation decisions to low-typicality members (e.g.

McCloskey & Glucksberg, 1978). So, a single participant may categorise a *tomato* as a *fruit* on one occasion and a *vegetable* on another, just as some participants may categorise it as *fruit* and others as a *vegetable* on a single occasion.

Probabilistic approaches are not without problems. They do not account for typicality effects which are present in goal-derived, or ad-hoc, categories (e.g. *ways to escape being killed by the mafia;* Barsalou, 1983) which presumably cannot be a consequence of comparison to a prototype or with previously encountered category members. They also do not explain how category members are initially decided in order for a prototype to develop or against which new concepts can be judged.

Medin (1989) goes on to suggest a theory-based approach to categorisation which takes elements of probabilistic approaches but embeds them within a structural framework. This was developed from an understanding of the causal and explanatory principles of why particular things/concepts display certain attributes and features. This is also consistent with what Taylor (2003) identifies as categorisation based on linguistic and encyclopaedic knowledge. To illustrate this, Medin gives the following example to illustrate how the mere presence of features does not guarantee membership of a particular (diagnostic) category and that category membership can vary according to real-world knowledge:

> A teenage boy might show many signs of the behaviours associated with an eating disorder, but the further knowledge that the teenager is on the wrestling team and trying to make a lower weight class may undermine any diagnosis of a disorder (Medin, 1989:1474).

Although there is currently no satisfactory theory of categorisation that can account for all experimental findings (see Smith & Medin, 1981, for discussion), the theories discussed above set the context for the development of various principles that underlie categorisation.

2.2.2. Principles of categorisation

During the 1970s, Rosch and colleagues identified various principles of categorisation arguing that the world is highly structured and categories form naturally in a non-arbitrary manner on the basis that attributes do not arise independently (e.g. Rosch, 1978; Rosch & Mervis, 1975; Rosch et al, 1976). Rosch et al (1976) offered an illustration of this point with:

> Creatures with feathers are also more likely to have wings than creatures with fur, and objects with the visual appearance of chairs are more likely to have functional sit-on-ableness than objects with the appearance of cats (p.383).

It was argued that systems of categorisation in mental processes are employed so as to provide cognitive economy, giving "maximum information with the least cognitive effort" (Rosch, 1978:28). Where a *category* was defined as a number of objects which are considered equivalent, categories themselves were considered to be related to one another via a *taxonomic structure* which is hierarchical and branching in design. The taxonomic hierarchy moves from the least specific and most inclusive level (i.e. the highest level), down to the most specific and least inclusive levels (i.e. the lowest levels). Each level of the hierarchy was referred to as a *level of abstraction*. An example taxonomic hierarchy is provided in Figure 2.1.

Hierarchical categorisation employs vertical and horizontal dimensions (Rosch, 1978). The vertical dimension indicates the level of inclusiveness from least to most specific (e.g. *animal* - *mammal* - *dog* - *collie*). Evidence of the vertical dimension was provided by Collins & Quillian (1969) who found that participants were faster to verify statements which required traversing fewer levels of the vertical hierarchy. Here, participants verified statements such as '*robins are robins'*, faster than '*robins are birds'*, which was itself verified faster than '*robins are animals'*. The horizontal dimension refers to the segmentation of individual categories at the same level of inclusiveness (e.g. *dog*, *cat*; *car, bus; chair*, *sofa*) and is the dimension along which individual concepts are differentiated. For example, *dogs* are differentiated from *cats,* yet one can identify a particular dog as *an animal, a mammal, a dog*, or *a collie*.

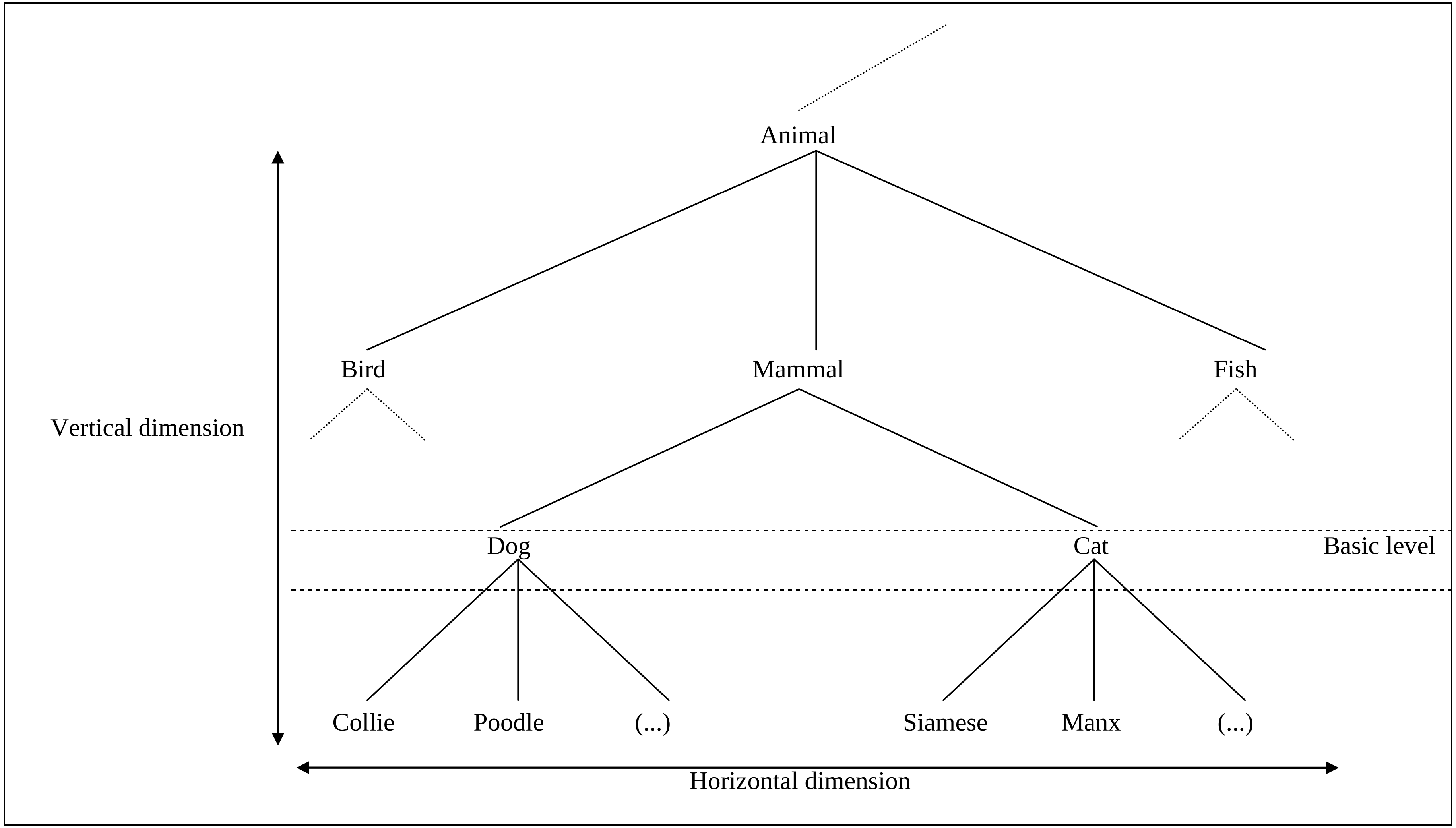

Figure 2.1 Hierarchical taxonomic organisation of objects

Within the vertical dimension of categorisation, Rosch et al (1976) identified a 'basic level' of abstraction. At this level, categories are argued to encode the most information and individual categories along the horizontal dimension at this level of abstraction are maximally distinctive. *Dogs* and *cats* both belong to the same higher level (i.e. superordinate) category *mammals*, and are differentiated to a greater extent than categories at a superordinate level (e.g. comparing *mammals* to *fish* or *birds*), and also to categories at levels subordinate (e.g. comparing *collies* and *poodles*, or *siamese* and *manx cats*). Evidence for the claim of maximal distinctiveness at the basic level was provided by Rosch & Mervis (1975) who asked participants to list attributes for concepts at superordinate, basic, and subordinate levels of abstraction. Participants listed a greater number of attributes which were common to related categories at basic level (e.g. *guitar, piano,* and *drum* which are related via their superordinate category of *musical instruments*). It was also found that these attributes shared within categories at the basic level were shared to a lesser extent with other categories within the basic level than was the case at the superordinate level.

Similar hierarchies and principles of categorisation have also been shown in the domain of events (e.g. Rifkin, 1985) where taxonomies such as *medical activity* > *surgery* > *heart surgery* have also been argued to follow the superordinate > basic > subordinate vertical hierarchical structure. Rosch (1978) had also previously suggested that speakers agreed on the level at which daily events should be described. Participants listed activities in their daily routine in terms of '*making coffee*', '*taking a shower*' and '*going to class*', rather than being more specific (i.e. expressing subordinate level events) such as '*picking up a tube of toothpaste*' followed by '*squeezing toothpaste onto a brush*', or less specific (i.e. expressing superordinate level events) such as '*getting themselves out of the house in the morning*'. Therefore, people tended to express events at the basic level where events are perceptually more salient due to their level of distinctiveness. This was argued to reflect a level of event description that is most perceptually salient and maximally distinctive and was consequently considered analogous to the basic level of abstraction for objects.

While there has been limited research into principles of categorisation with regard to actions (as lexically expressed as verbs), there have been suggestions of equivalent hierarchical structuring (e.g. Lakoff, 1987; Miller & Johnson-Laird, 1976). It has been suggested that actions such as *walking* and *drinking* are basic level actions which are hierarchically linked to superordinate actions such as *moving* and *ingesting* and subordinate actions such as *ambling* and *slurping*. Some evidence for the perceptual salience of these proposed basic level actions was provided by Hemeren (1996) who conducted a category listing task with English and Swedish speaking participants. Participants were asked to list 'actions that involve some kind of *bodily activity that can easily be recognised when seen and can be visualised as a mental image*' (p44; Hemeren's italics). It was found that those actions which had previously been hypothesised to be at a basic level of abstraction (e.g. *running*, *walking*, *jumping*, and *talking*) were listed more frequently and were listed earlier than actions which are more specific in terms of manner and also more reliant on context (e.g. *sprinting*, *strolling*, *skipping*, and *whispering* respectively).

Hemeren (1996) also reported findings with regards to consistency between participants. There were only five verbs listed by more than half of the total participants (*running*, *walking*, *jumping*, *swimming*, and *skipping*), which again showed some overlap with those actions hypothesised to be at a basic level. There was also considerable variation in terms of numbers of responses given by each participant. English participants listed a mean of 36.4 actions with a standard deviation of 10.9 and a range from 20 to 67 actions. As a result, Hemeren concluded that the participants did 'not seem to access similar semantic or categorical domains in relation to the general perceptual criteria' (1996:47). This may therefore be an indication that action taxonomies/categories have a more flexible architecture than objects which may link to previous suggestions that actions/verbs are represented within a highly interconnected three-dimensional 'matrix-like' architecture rather than a more rigid two-dimensional hierarchical architecture (e.g. Grasser et al, 1987; Huttenlocher & Lui, 1979).

One example which demonstrates that nouns and verbs may both demonstrate aspects of hierarchical organisation is the lexical database WordNet (e.g. Miller & Fellbaum, 1991, 2007). This database employs relational semantic principles for both nouns and (action) verbs. Nouns are related according to hyponymy (i.e. 'type'-relations, where *maple* is a type of *tree*) whereas verbs are related via troponymy (i.e. 'manner'-type relations, where *persuade* is a manner of *communication*). While the relations of hyponymy and troponymy are not qualitatively the same, they do establish the same entailment relations. They both establish hierarchical relations so that for objects it can be stated that '*an apple is a type of fruit*' where *fruit* is represented at a higher hierarchical level than *apple*. Conversely, one cannot state that '*a fruit is a type of apple*'. Similarly, with actions, it can be stated that '*washing is a way, or manner, of cleaning something*' where *cleaning* is represented at a higher hierarchical level than *washing.* Again, one cannot state that '*cleaning is a way, or manner, of washing something*'.

Within WordNet, hierarchies for both nouns and verbs therefore extend from least specific to most specific. The main structural difference comes in the depth of the hierarchies, i.e. the number of levels involved. All nouns extend from the same topmost level of the hierarchy which is represented by the beginner term 'entity'. To reach specific concepts, several levels of organisation need to be traversed, so, to reach the subordinate level term *roadster*, one must traverse, from least to most specific, the terms: *entity* → *artefact* → *conveyance* → *vehicle* → *wheeled vehicle* → *motor vehicle* → *car* → *roadster*. It can be seen that the hierarchies included in WordNet do not necessarily conform to the superordinate, basic, and subordinate level hierarchies previously proposed by Rosch and colleagues as there are intervening levels. The basic hierarchical principles are however the same. In comparison, verb hierarchies tend to be more 'shallow and bushy' and rarely exceed four levels of structure (Miller & Fellbaum, 1991:217). Not all verbs are derived from a single beginner term but are derived from more abstract representations, so, in reaching a term such as *march* one must only traverse the levels (from least to most specific): {*move*, *make a movement*} → *walk* → *march*. While WordNet demonstrates the possibility that verbs can adopt hierarchical taxonomies as an organising principle, Miller & Fellbaum (1991) also highlight the limitations of such principles as they may not hold for all verbs. For example, they suggest that stative verbs (e.g. *exceed*, *differ*, *match*) "do not form a semantically coherent group and share no semantic properties other than that they denote states rather than actions or events" (p216).

It should be stressed that WordNet does not necessarily claim to be a representation of how speakers actually categorise concepts. It is unlikely that speakers are ever aware of all the levels of taxonomic structure that are associated with every concept. The important aspect with WordNet is that it demonstrates that, in theory, both nouns and (at least a subset of) verbs can be hierarchically structured.

2.2.3. Typicality and categorisation

As previously discussed, concepts within object categories appear to be organised along a typicality gradient where a concept's typicality is determined by its overall similarity to other category members (e.g. through an abstracted prototype, or established category members; e.g. Rosch & Mervis, 1975). The more attributes that a concept has in common with other members of the same category, the more representative it is of the category, and the more typical it is. Rosch & Mervis (1975) also found that concepts which were highly typical within a category, shared fewer attributes with members of other categories than did atypical members from the same category. So, *orange* (a highly typical *fruit*) shares more features with *apple* and *banana* (other high-typicality *fruits*) than it does with *coconut* and *olive* (low typicality *fruits*). This reinforced the notion of greater representational coherence between typical category members than atypical category members.

Ashcraft (1978) conducted further investigations into the featural composition of categories and their relationships to typicality. After asking participants to list features for typical and atypical members of 17 different categories, it was found that there was greater between participant agreement on the features that were associated with typical category members than for atypical members. It was also found that typical category members (at the basic level of abstraction, e.g. *apple*) shared more features with the category concept itself (at the superordinate level of abstraction, e.g. *fruit*) than did atypical category members (e.g. *olive*). Ashcraft (1978) reported further findings that participants were faster to verify category membership of typical category members than they were to verify the category membership of atypical category members.

Similar typicality effects have also been reported in subsequent studies (e.g. Casey, 1992; Larochelle & Pineau, 1994) and this effect has been shown to persist in participants with impaired language (e.g. Kiran & Thompson, 2003). Typicality effects in category verification have also been demonstrated for ad-hoc, or goal-based categories, such as *things you would find in a grocery store*, or *things to take on a camping trip* (e.g. Barsalou, 1983; Sebastian & Kiran, 2007). Typicality effects are also present in other cases of language impairment, such as in semantic dementia (e.g. Lambon Ralph & Howard, 2000; Patterson, 2007). In an object naming task, typical category members were more successfully recalled in an object recall via drawing task, more typical category members were more accurately drawn. When atypical category members were drawn, distinctive features, i.e. features that mark the members as atypical, were frequently omitted and features which were indicative of more typical members of the category were often inappropriately introduced (e.g. a rhinoceros drawn without a horn or a duck drawn with four legs).

An important issue to consider when drawing conclusions about the importance of typicality in category structure is to consider the degree to which typicality is independent of other variables which may influence performance. This has been investigated in the context of category listing tasks, for example when participants are asked to list as many different *birds*, *clothes*, *fruit*, and so on, as possible. Mervis, Catlin & Rosch (1976) reported that typicality (referred to as goodness-of-example) correlated significantly with item dominance (i.e. the frequency with which a particular category member was listed within a particular category). Category members rated as highly typical were more dominant (i.e. were listed more frequently) than less typical category members. They also found that lexical frequency (i.e. the frequency that a word occurs in the language in general) only tended to correlate with item dominance and not with typicality. This suggested that typicality was independent of lexical frequency, i.e. that typicality ratings were not merely a simple reflection of frequency of occurrence. A similar conclusion was offered by Boster (1988) who investigated the internal structure of the category *birds*. He found typicality was more strongly correlated with ratings of similarity between birds than with lexical frequency and also the frequency with which participants were likely to encounter the birds themselves (inferred from local bird count statistics in the sampled area).

Hampton & Gardiner (1983) carried out further correlation analyses with regard to category structure with consideration of associative frequency (i.e. item dominance), typicality and also familiarity (i.e. the degree to which participants are familiar with the category member). This was done following suggestions that there can be two reasons for why an item may be rated as atypical of a category: (1) the member is not representative of the category as a whole; and/or (2) the category member is not well known and is rarely experienced in real-life situations for the population being sampled. Across all categories, typicality proved to be more strongly correlated with associative frequency than familiarity (r = -.76 and r = -.61 respectively) with typicality and familiarity showing a weaker correlation (r = .54). Whilst the same pattern was observed for the majority of the individual categories, for the three categories of creatures (i.e. *fish*, *birds*, and *insects*) both typicality and familiarity were equally good predictors of associative frequency. There appears therefore to be a general consensus that typicality is an (if not <u>the</u> most) influential factor which predicts association strength between a category and its members.

Independent effects of typicality have also been observed in tasks other than category listing. Holmes & Ellis (2006) found that typicality has independent effects but can also interact with age of acquisition (i.e. the age at which a word is acquired in language development) in tasks such as naming, object decision, and categorisation. Also, within category verification, typicality has been found to be a significant and unique predictor of response time when other factors such as familiarity, associative frequency, and word frequency have been either controlled or accounted for through regression analyses (e.g. Casey, 1992; Hampton, 1997; Larochelle & Pineau, 1994).

Typicality has also been suggested to have important implications for interventions for people with language impairments (i.e. aphasia). There is evidence to suggest that intervention may be more effective if targeted explicitly towards atypical category members as the greater variety of semantic features which are implicitly or explicitly activated as part of the treatment process leads to greater generalisation (i.e. improvements beyond the words that have actually been used in therapy). Such effects have been observed in animate categories such as *birds* and *vegetables*, inanimate categories such as *clothing* and *furniture*, and also well-defined categories such as *females* and *shapes* (e.g. Kiran, 2008; Kiran, Shamapant & DeLyria, 2006; Kiran & Thompson, 2003; Stanczak, Waters & Caplan, 2006).

2.2.4. The investigation of categorisation and typicality

Research into categorisation and typicality effects relies on valid and reliable normative data. In order to investigate typicality effects within categories, researchers need to know which items belong within a particular category and about each category member's typicality within that specific category. Only then can experiments be designed and hypotheses tested relating to how variables such as typicality or associative frequency affect performance in further processing-based tasks.

Battig & Montague (1969) were among the first to present an extensive set of normative data relevant to categorisation. They collected category norms for 56 object categories including *precious stones*, *countries*, *weather phenomena*, and so on. Participants were given 30 seconds to write down as many objects within each category as they could think of. A total of 442 participants completed the experiment from two different geographical areas of the United States. From these investigations, Battig & Montague reported the production frequency of each response within each category (i.e. associative frequency), the number of times a response was listed first within each category, and also the mean ordinal rank of each response within each category (e.g. where a mean rank close to 1 would suggest that the object was listed earlier in participant's lists).

Battig & Montague (1969) subsequently performed correlation analyses which provided additional insights into categorical representations in semantic memory and language processing. Firstly, it was shown that categories were generally consistent between the two geographical locations, such that participants tended to give the same objects within categories and with comparable production frequencies, and mean rank, and so on. The categories which did show geographical variation were those which were geographically focused (e.g. *states, cities, colleges, trees*, and so on). Secondly, there was between-category variation in terms of category size with the smallest category containing a mean of just 4.4 members across all participants (*buildings for religious services*) and the largest containing a mean of 11.34 members (*parts of the human body*). Finally, it was found that production frequency and mean rank were highly correlated in the majority of categories, with objects listed by most people being listed earlier.

Such has been the reliance on the Battig & Montague (1969) category norms in research in areas such as psychology, linguistics, and aphasiology, that Van Overschelde, Rawson & Dunlosky (2004) sought to provide an updated and expanded set of category norms. The updated set of norms again showed strong within-category correlations of production frequency and mean rank, although there were also a number of qualitative changes that had occurred since the original Battig & Montague (1969) norms. Some changes were observed in categories where change may be expected as a result of time progression (e.g. *dances* and *music*) whereas others were more unexpected as they would be assumed to be more stable and enduring (e.g. *time* and *colour*).

A further set of normative data was provided by Hampton & Gardiner (1983) who provided measures of associative frequency, typicality and also familiarity for items within 12 object categories which were obtained from British-English speaking participants. Typicality and familiarity measures were both obtained by asking participants to rate each item’s typicality and familiarity (i.e. how familiar they were with the meaning of the words) on scales ranging from 1 (very typical/familiar) to 5 (very atypical/unfamiliar). Associative frequency measures were gained via a category listing task as in Battig & Montague (1969).

While there is some existing research, such as Hemeren’s (1996; see section 2.2.2), that provides some normative data for categorisation principles within the domain of actions/verbs, this is fairly limited in its potential for generalisation. Hemeren only investigated a single category of actions. Therefore, when interpreting this data as a set of category norms (which was not the original intention), there is clearly less scope than if using existing object/noun category norms. Second, the category, as it was presented to participants, appears to be more abstract and perhaps more conceptual than those used in investigating object/noun categories. For example, if comparing Hemeren’s category of ‘actions that involve some kind of *bodily activity that can easily* *be recognised when seen and can be visualised as a mental image*', it appears to be more akin to goal-derived and ad-hoc categories (e.g. Barsalou, 1983) rather than discrete object/noun categories such as *birds*, *fruit*, and *tools*. It would, therefore, be problematic to place Hemeren's category at a hypothesised level of abstraction within a taxonomic hierarchy for actions/verbs.

Given the lack of normative data associated with actions and verbs, further research investigating the processing of actions and verbs must rely on other sources of information (or otherwise researcher intuition). One source may be the previously discussed WordNet database (e.g. Miller & Fellbaum, 1991, 2007) which offers a fairly comprehensive account of verbs organised within hierarchical taxonomies. However, one drawback here is that there is a lack of associative measures to indicate how category members are differentiated (e.g. measures such as production frequency/associative frequency, and typicality). Another commonly used resource is Levin's (1993) classificatory system of English verbs. This system was derived from a large-scale meta-analysis of literature within the domains of syntax and semantics and is based on the assumption that a verb's meaning influences its syntactic behaviour. Levin's system differentiates verbs, even those with superficially similar meanings, on the basis of the syntactic structures that they can and cannot appear in and this in turn is taken to reveal fine-grained semantic classes. This analysis derived a total of 49 major semantic verb classes, the majority of which also derive a number of subclasses. For example, the major class of *Verbs of Sending and Carrying* is subdivided into the classes: *Send verbs*, *Slide verbs*, *Bring and Take*, *Carry verbs*, and *Drive verbs*; whereas the major class of *Psych-verbs* is subdivided into: *Amuse verbs*, *Admire verbs*, *Marvel verbs*, and *Appeal verbs*. Whilst Levin (1993) does provide a comprehensive classification system of English verbs, the fact that it is derived from research in theoretical syntax (e.g. where experimental methods include making grammaticality judgements on part and whole sentences) means that the underlying classification principles have been derived differently from how they have been for objects/nouns. The use of Levin's system in online psychological and processing-based investigations may therefore be limited as it may not be an accurate reflection of how such information is stored and retrieved in language processing and semantic memory.

2.2.5. The current studies and research questions

The discussion so far has provided a review of how research has developed in identifying theories and principles of semantic categorisation and a justification for why further research should focus on using similar methodologies in investigating whether similar theories and principles can be applied to actions/verbs. Therefore, the following two experiments aimed to address the following general research questions:

1) Are verbs listed within verbs categories in a manner that parallels how nouns are listed within noun categories? If not, in what ways does performance differ?

2) Is participants' ability to rate the typicality of verbs within categories correlated with performance in category listing, so that the more typical a verb is within a category, the more often it is given within a category?

3) Can existing principles of semantic categorisation, previously established for objects/nouns, be generalised to principles for the semantic organisation of actions/verbs?

The chapter continues with a description of the methods and results of a category listing task and then a typicality rating task. This is followed with a general discussion which will address the above research questions.

2.3. Category Listing of Verbs and Nouns

2.3.1. Introduction

This section reports the findings from a category listing task whereby participants' performance in listing verbs within verb categories was compared to their ability to list nouns within noun categories. Participants were required to list verbs within 10 verb categories and to list nouns within 10 noun categories. After first describing the effects of a data exclusion process (which includes some qualitative analysis), reliability measures are reported along with analysis of gender effects and a pseudo-split-half analysis to check reliability of performance across two counterbalanced presentation lists. Following this, the analysis considers quantitative comparisons between performance in verb category listing and noun category listing in terms of overall performance and performance at the level of individual categories. The relationships between measures of production frequency (i.e. the frequency that a particular items is listed within a particular category) and mean rank (i.e. a measure of how early particular items are given within a particular category) are then considered.

2.3.2. Method

Participants

Thirty-five participants completed this experiment. This sample consisted of 13 males and 23 females (*M age* = 23.1 years, *SD* = 7.9, *range* = 18-57). All participants were native English speakers and were enrolled on University degree programmes. Participants were recruited via email advertisements and all gave written consent and were either paid for taking part or entered into a cash/voucher prize draw.

Stimuli selection

Ten nouns and ten verbs were selected to serve as categories for which participants would list category members. The categories were selected with reference to the WordNet lexical database (Miller and Fellbaum, 1991) so that they were immediately superordinate to the most populous layer of hierarchical organisation. The noun categories (i.e. *animals*, *birds*, *clothes*, *fruit*, *furniture*, *musical instruments*, *sports*, *tools*, *transport*, and *vegetables*) were generally consistent with those used in previous research (e.g. Hampton and Gardiner, 1983; Rosch et al, 1976) with the exception of *animals* which has tended to be treated as superordinate to categories such as *birds*, *fish*, *mammals*, and so on. The noun categories used here are also generally considered to be at the superordinate level of categorisation (i.e. the level immediately above the basic level. The verb categories selected (i.e. *breaking*, *cleaning*, *cooking*, *cutting*, *hitting*, *jumping*, *making*, *running*, *talking*, and *walking*) included some of the actions previously hypothesised to be at a basic level of abstraction for actions (e.g. Miller and Johnson-Laird, 1976; Lakoff, 1987). Therefore, noun and verb categories used here may be argued to differ in terms of their perceptual salience (i.e. with noun categories at a superordinate level and verbs at a basic level), which may impact on participants' performance in the category listing task that aims to compare participants' performance between word classes. However, this is dependent on whether the term 'basic level' can be applied equivalently across noun categories and verb categories (see section 2.5.2 for further discussion on conceptual/semantic differences between noun and verb categorisation and impact on terminology). For the current experiment, noun and verb categories were selected from the hierarchical level above the most populous in the

WordNet database. This was done to ensure that participants would be able to generate a substantial number of both nouns and verbs which were semantically diverse while still sharing a categorical-type relation.

Design and procedure

Participants were seen in groups of varying size. Each participant was given one of two workbooks to complete. Written instructions stated that participants would see a series of sentences asking them to list words that they felt belonged to various categories (e.g. *write down as many different types of game as you can think of*; or *write down as many different ways of moving as you can think of*). They were told to keep their responses to either a single word or as short as possible (i.e. not full sentences). Participants were shown two completed examples for the categories *game* and *moving*, each with eight example responses which were congruent to the target word class (e.g. *football*, *chess*, *cards*, and so on, and *pushing*, *pulling*, *rolling*, and so on). While the examples implied that participants should list nouns in noun categories and verbs in verb categories, participants were not explicitly instructed to do so.

Participants were given two minutes to list responses in each category which were each presented on a different page of the workbook. Participants were instructed not to turn the page to begin listing for a new category until the experimenter instructed them to do so every two minutes. Each page of the workbook presented 20 numbered spaces for responses to be listed. Participants were instructed that they did not need to list responses in all spaces but to complete as many as they could within the given time. If they were able to give more than 20 responses, they were instructed to do so in an orderly manner at the bottom of the page.

All participants listed responses for all 20 categories in one of two counterbalanced workbooks. The order or category presentation was randomised although the order of presentation alternated between noun and verb categories. The order of category presentation in one workbook was reversed for the second workbook to counterbalance any possible effects of fatigue that might have occurred during the 40 minute experimental session.

Data preparation

Before analysis, the raw data were subjected to a number of clean up processes. An initial phase involved standardising spelling across the data. Second, plurality and tense were standardised where no discrimination was made between different morphological realisations of the same base lemma. For example, both *cat* and *cats* were standardised and counted as examples of the same response in the same way that *drop* and *dropping* were. The only occasions when this was not the case was when it was ambiguous as to whether the participant had intended the response as a noun or a verb, e.g. where a response was written in a form ambiguous between a noun and a verb in infinitive form (e.g. *soap*) and where no other participant had written the response in a corresponding finite form (e.g. *soaping*). In such cases the ambiguous response was excluded from analysis on the basis of word class (e.g. coded as a noun given within a verb category). Third, elimination of repetition was carried out.

Some participants listed two or more responses in a single response space in workbooks. This was generally done when participants felt responses were synonymous or very similar (e.g. *sultanas*/*raisins*). All such responses were counted individually. Any responses that were repetitions of previous responses in the same category list, or which repeated the category itself, were excluded. Finally, word class was standardised. To ensure that only nouns were counted within noun categories and only verbs were counted within verb categories, responses were excluded on the basis of word class. These included responses written as: prepositional phrases (e.g. *with a hammer*, *in the oven*, *for pleasure*, and so on), adverbs and adjectives (e.g. *quickly*, *strongly*, *hard*, *fast*, and so on), nouns which were listed in verb categories and verbs which were listed in noun categories (n.b. some obvious exceptions were made, for example, in the category of *sports* which included a number of response written as verbs, e.g. *running*, *swimming*, *horse riding*, and so on).

No other criteria were applied to the data. Responses at differing levels of categorisation were included and counted as separate responses (e.g. *chair*, *rocking chair*, *dining chair*; and *frying*, *deep frying*, *stir frying*, and so on). Adaptation of the raw data was kept to a minimum even where responses appeared synonymous. For example, both *vacuum* and *hoover* were counted as separate responses in the category of *cleaning*, primarily because some participants had listed both. However, some differences in vocabulary were combined, for example, *eggplant* was subsumed and combined with *aubergine* as *eggplant* was only listed once and was presumed to be a lexical variation as *aubergine* was not also given by the same participant. A more straightforward example was with *aeroplane* and *plane* which were also subsumed.

2.3.3. Results

Effect of data preparation

Following the initial ‘clean-up’ phases (i.e. standardising spelling, plurality, tense, and so on), 3313 responses were listed across all 10 verb categories across all participants. Following application of the exclusion criteria, 738 responses were excluded leaving 2575 responses which were subjected to further analysis. In comparison, 6412 responses were listed across all 10 noun categories across all participants, 42 of which were excluded leaving 6370 responses for further analysis. Tables presenting exclusion data by word class and individual category are displayed in Appendix A.

Within verb categories, the majority of exclusions were adverbs/adjectives (53.3% of total exclusions) but there was also a substantial number of nouns (23.4%) and prepositional phrases (21.3%) excluded. The remainder were classified as repetitions (1.3%) or ‘other’ (4.1%) where the verb was applied in a different sense to that intended (e.g. *to break up with someone, to make friends*). A large proportion of exclusions were from the categories *running* (20.2% of total exclusions), *talking* (18.6%) *walking* (16.2%), and *cutting* (14.7%). The greatest number of exclusions was for adverbs/adjective responses given within the category *talking* (e.g. *loudly*, *quietly*) which accounted for 12.6% of all exclusions in verb categories.

Within noun categories, the majority of exclusions were verbs given within the category *transport* (e.g. *walking*, *jumping*) which accounted for 69% of all exclusions. Most other exclusions were repetitions (26.2%) which were spread across 6 of the 10 categories with the remaining exclusions being adverbs/adjectives (4.8%).

Effect of gender

As gender differences have been observed in previous investigations of semantic memory and category structure (e.g. Capitani, Laiacona and Barbarotto, 1999; Laiacona, Barbarotto and Capitani, 2006), a preliminary analysis was conducted to investigate whether gender affected quantitative properties of the data. This was necessary due to the over-representation of female participants in the sample which may bias the data if such gender effects were present. The data from 13 randomly chosen female participants was paired with the data from the 13 male participants. The data were then compared using independent samples t-tests (two-tailed) with the significance level set at $p < .05$.

Tables presenting gender comparison data are included in Appendix B. There were no significant differences between males and females in terms of total number of responses listed (*t* (24) = 1.146, *p* = .263, *d* = 0.65), the total number of verbs listed (*t* (24) = 1.395, *p* = 0.176, *d* = 0.55), or the total number of nouns listed (*t* (24) = 0.695, *p* = 0.442, *d* = 0.28).

Analysis was also carried out at the level of individual categories. Of the verb categories, there were significant differences between male and female participants for the categories *breaking* (*t* (24) = 2.467, *p* = .021, *d* = 0.9) and *walking* (*t* (24) = 2.137, *p* = .043, *d* = 0.84) where females listed more responses than males. There were no significant differences observed within noun categories although, *furniture* (*t* (24) = 2.042, *p* = .052, *d* = 0.81) and *vegetables* (*t* (24) = 1.917, *p* = .067, *d* = 0.76) showed trends towards significance, again with females listing more responses than males. No other comparisons showed significant differences between male and female participants. Despite there being some significant differences between males and females, these were not considered to be influencing the task as a whole and therefore, for subsequent analysis, data from all 35 participants were pooled.

Effect of presentation list

A further preliminary analysis involved quantitative comparison of the data produced from the two presentation lists with a view to this being an indicator of the reliability of the data. As two administration lists were employed primarily to counteract any possible effects of fatigue in the 40 minute experimental session, it may be possible that some differences were present, for example, between categories that are presented earlier in one list and later in the other list. This analysis was carried out with the data from 34 participants with one participant's data being randomly excluded to ensure equal number of data for each presentation list.

Tables presenting list comparisons are included in Appendix C. There were no significant differences between the two presentation lists in terms of the total number of responses listed (*t* (32) = 0.658, *p* = .515, *d* = 0.23), the number of verbs listed (*t* (32) = 0.091, *p* = .928. *d* = 0.33), or the total number of nouns listed (*t* (32) = 0.951, *p* = .349, *d* = 0.03).

Analysis was additionally conducted at the level of individual categories. The only significant difference observed was within the *walking* category (*t* (32) = 3.644, *p* = .001, *d* = 1.25) where participants completing list A (where *walking* was the 15^{th} category completed) listed significantly more verbs than participants completing list B (where *walking* was the 6^{th} category completed). No other comparisons approached significance.

Effect of word class

Analysing the data from all 35 participants, a significant difference was seen between the mean number of verbs and nouns listed (paired samples t-test; t (24) = 22.261, p < .001, d = 4.42, one-tailed) where there were more nouns (M = 182.03, SD = 29.942) listed than verbs (M = 73.54, SD = 19.137) across all respective categories.

Category analysis

Descriptive statistics for individual verb and noun categories are presented in Table 2.1and Table 2.2 respectively. These include: the mean number of responses listed (*M*) and respective standard deviations (*SD*), the range of responses listed (Range), and the number of different responses listed (Different). An independent samples t-test showed that there were significantly more different nouns given within noun categories than there were different verbs given in verb categories (t(18) = -3.58, p = .002).

The qualitative data from verb and noun categories are reported in Appendices D and E respectively. These tables show the actual responses given ordered by their production frequency (i.e. the number of participants listing the response). For responses with a production frequency of 3 or higher, the mean rank position of the response (i.e. the mean ordinal position that participants tended to give each response, e.g. the first response given within a category was assigned the rank of 1, the second response a rank of 2, and so on) is also given. Mean rank was calculated according to the number of participants who actually listed the response rather than an average across all 35 participants. The number of participants who listed each response first within a category is given. Finally, verbs with a production frequency of 2 are also listed whereas those with a production frequency of 1 are not listed.

	M	SD	Range	Different
Breaking	9.09	3.576	3-17	83
Cleaning	10.20	2.898	5-16	67
Cooking	10.37	3.379	6-20	60
Cutting	5.60	3.070	0-12	58
Hitting	7.46	3.689	0-14	76
Jumping	5.97	2.717	1-13	48
Making	8.94	3.307	3-16	90
Running	3.11	1.659	0-7	25
Talking	7.00	4.332	0-14	82
Walking	5.80	3.653	0-14	68

Table 2.1 Descriptive statistics for verb category listing

	M	SD	Range	Different
Animals	22.89	5.229	9-33	163
Birds	18.77	4.995	5-28	122
Clothes	21.26	3.744	15-31	117
Fruit	19.46	3.845	11-29	69
Furniture	14.60	3.283	9-21	99
Musical instruments	19.14	4.038	11-30	73
Sports	20.11	3.385	13-30	120
Tools	13.14	3.318	6-20	111
Transport	15.94	4.419	8-23	106
Vegetables	16.71	3.667	7-25	71

Table 2.2 Descriptive statistics for noun category listing

Relationship between production frequency and mean rank

Correlation analyses were carried out on measures of production frequency and mean rank to investigate the internal structure of the investigated categories. These were carried out for all responses, within each category, which had a production frequency of 3 or more. When calculating correlation coefficients, Spearman's coefficient is reported in favour of Pearson's coefficient as this makes fewer assumptions about the distribution and variance of the data and was considered to be more appropriate for this preliminary analysis of verb semantic categories.

Table 2.3 and Table 2.4 report the correlation statistics for individual verb and noun categories respectively in addition to correlation analysis conducted across all categories within each word class. The only categories that failed to produce significant correlations were all verb categories (i.e. *running*, *talking*, and *walking*).

	n observations	r_s =
Across categories	233	-.395**
Breaking	30	-.483**
Cleaning	30	-.559**
Cooking	28	-.629**
Cutting	18	-.513*
Hitting	26	-.459*
Jumping	18	-.785**
Making	31	-.408*
Running	6	-.754
Talking	21	-.345
Walking	25	-.338

Note.: * p < .05; ** p < .01

Table 2.3 Correlations between production frequency and mean rank (verb categories)

	n observations	r_s =
Across categories	492	-.438**
Animals	65	-.572**
Birds	63	-.274*
Clothes	51	-.534**
Fruit	45	-.382**
Furniture	41	-.543**
Musical instruments	46	-.641**
Sports	58	-.527**
Tools	36	-.614**
Transport	48	-.547**
Vegetables	39	-.734**

Note.: * p < .05; ** p < .01

Table 2.4 Correlations between production frequency and mean rank (noun categories)

2.4. Typicality Rating of Verbs within Categories

2.4.1. Introduction

This section reports the results of a typicality rating task for verbs within eight of the 10 categories used in the previous category listing task. The analysis considers the distributions of typicality ratings within categories and also the relation that typicality has with measures obtained in category listing (i.e. production frequency and mean rank) and also of lexical frequency (i.e. the frequency that a word occurs in the language).

2.4.2. Method

Participants

A total of 102 participants completed this experiment. This sample consisted of 34 males and 68 females (*M age* = 20.9 years, *SD* = 4.4, *range* = 18-58). All participants were native English speakers and were enrolled on University degree programmes. All participants were recruited via email advertisements, gave written consent and were entered into a cash/voucher prize draw.

Stimuli selection

Responses from eight verb categories were used. These were drawn from the categories: *breaking*, *cleaning*, *cooking*, *cutting*, *hitting*, *making*, *talking*, and *walking*. The categories of *jumping* and *running* were excluded for this experiment due to the limited number of responses given within them in the category listing experiment (see section 2.3).

Within the eight categories, all verbs with a production frequency of 3 or greater in Experiment 1 were included in this experiment. This totalled 209 verbs for which typicality ratings were obtained. The number of verbs drawn from each category ranged from 18 (*cutting*) to 31 (*making*).

Design and procedure

Typicality ratings were collected via a web-based survey tool (www.surveymonkey.com). Participants were instructed that they were to judge how typical certain actions were in relation to a more general category. To elucidate this point, participants were given the following standard passage to consider of how typicality is utilised within the domain of objects:

> How typical is a penguin as a type of bird? You may also think of this as asking: How 'bird-like' is a penguin? You can also compare this to the question: How typical is a robin as a type of bird? Most people would probably say that a robin is more bird-like, or typical, than a penguin.

Participants were instructed to indicate their judgements on a scale ranging from 1 (very typical) to 7 (not very typical). They were also instructed to indicate in a separate field if they felt that a particular action was not part of the mentioned category.

All participants rated the typicality of each of the 209 verbs within their respective categories. All verbs for each individual category were presented on a single page and participants had to rate all verbs before moving on to the next category. All participants completed the same survey. The eight categories were presented in a random order and the verbs to be rated in each category were also presented in a random order. Before rating the verbs, participants were encouraged to look over all the verbs appearing in the list so that they would (be encouraged to) use the whole scale.

2.4.3. Results

Typicality ratings

Mean typicality ratings (and standard deviations) for verbs within the 8 investigated semantic categories are presented within Appendix F. The number of participants who rated the response as not belonging to the category is also given. Finally, lexical frequency values obtained from the British National Corpus (BNC) which were used in the following correlation analyses are included. Appendix G presents comparable data for noun categories. For these tables, typicality ratings were taken from Hampton & Gardiner (1983) and, as with verbs, lexical frequency values were obtained from the BNC. Table 2.5 and Table 2.6 present the mean typicality ratings for each category in additional to distributional statistics for verbs and nouns respectively. The typicality distribution statistics for nouns represent only a subset of the nouns included in the Hampton & Gardiner typicality ratings. They should not therefore be considered an accurate reflection of the category as a whole but are merely included for comparative purposes. For example, category members that received a relatively high typicality rating in Hampton & Gardiner's (1983) ratings, such as *slacks* in the category of *clothes*, did not appear in the current data which may therefore misrepresent the internal structure of the category.

	M	SD	Min	Max	Skew	Kurtosis
Across categories (209)	2.68	0.87	1.10	5.47	0.457	-0.077
Breaking (30)	3.01	0.78	1.47	4.37	-0.088	-0.937
Cleaning (30)	2.34	0.78	1.17	4.24	0.467	-0.234
Cooking (28)	2.38	0.87	1.26	4.96	1.074	1.404
Cutting (18)	2.68	0.92	1.24	4.11	0.121	-1.189
Hitting (26)	2.86	0.89	1.21	4.68	0.334	-0.447
Making (31)	2.59	0.66	1.56	3.69	-0.132	-1.129
Talking (21)	2.45	0.97	1.10	4.55	0.707	-0.087
Walking (25)	3.11	0.91	1.56	5.47	1.069	1.368

Table 2.5 Typicality distribution statistics of verb categories

	M	SD	Min	Max	Skew	Kurtosis
Across categories (345)	2.06	0.83	1.00	5.16	1.034	1.235
Birds (63)	2.02	0.62	1.00	3.51	0.374	-0.265
Clothes (51)	1.88	0.72	1.00	3.29	0.436	-1.038
Fruit (45)	1.98	0.73	1.02	3.58	0.619	-0.774
Furniture (41)	1.66	0.67	1.00	3.59	1.558	2.876
Sports (58)	2.13	0.83	1.00	5.16	1.287	3.813
Transport (48)	2.78	1.19	1.00	4.87	0.173	-1.029
Vegetables (39)	1.82	0.61	1.00	3.29	0.786	0.260

Note. † Based on Hampton & Gardiner (1983) typicality rating scale of 1 (very typical) to 5 (very atypical)

Table 2.6 Typicality† distribution statistics of noun categories

Relationship between typicality and production frequency and mean rank

Mean typicality ratings were correlated with measures obtained in Experiment 1, namely production frequency and mean rank. This was done to investigate the validity of typicality as an indicator of category structure for verb categories. A similar analysis was done with nouns by obtaining typicality ratings from Hampton and Gardiner's (1983) ratings. However, not all nouns with a production frequency of 3 or more in the current study were present in the Hampton and Gardiner data, therefore these were excluded from the correlation analyses. The number of observations upon which correlation analyses were based in indicated in parentheses. Table 2.7 and Table 2.8 present the Spearman's correlation coefficients for verb and noun categories respectively. As in the category listing experiment, correlations are reported across all categories (for which typicality ratings were available) and for individual categories.

Typicality significantly correlated with production frequency in 7 of the 8 individual verb categories and in all 7 of the noun categories where is correlated significantly with mean rank in 4 of the 8 verb categories and again all 7 of the noun categories.

	typicality - production freq	typicality - mean rank
Across categories (209)	-.519**	.260**
Breaking (30)	-.647**	.580**
Cleaning (30)	-.487**	.312
Cooking (28)	-.751**	.619**
Cutting (18)	-.497*	.329
Hitting (26)	-.617**	.430*
Making (31)	-.487**	.402*
Talking (21)	-.164	.104
Walking (25)	-.545**	.280

Note.: * p < .05; ** p < .01

Table 2.7 Correlations between typicality, production frequency and mean rank (verbs)

	typicality - production freq	typicality - mean rank
Across categories (215)	-.607**	.583**
Birds (44)	-.362*	.617**
Clothes (28)	-.534**	.579**
Fruit (34)	-.727**	.597**
Furniture (18)	-.647**	.543*
Sports (35)	-.878**	.538**
Transport (28)	-.533**	.786**
Vegetables (28)	-.641**	.535**

Note.: * p < .05; ** p < .01

Table 2.8 Correlations between typicality, production frequency and mean rank (nouns)

Relationships with lexical frequency

A final set of correlation analyses were conducted where lexical frequency (i.e. the frequency with which each target word occurs in the language in general) was included. Values for lexical frequency were obtained from the BNC. Where a value for lexical frequency was not available due to it not being present in the BNC, this target item was excluded from the analyses. As a number of targets were now excluded from correlational analyses, Table 2.9 and Table 2.10 present results of all correlation pairs which were recalculated based only on those target words included in all analyses. As such, the number of items upon which correlations for each category were based is indicated in parentheses.

	pro - rank	pro - typ	pro - lex	rank - typ	rank - lex	typ - lex
Across (196)	-.363**	-.523**	.040	.261**	-.122	.113
Breaking (29)	-.503**	-.691**	.180	.587**	-.217	-.039
Cleaning (29)	-.618**	-.487**	.534**	.310	-.205	-.072
Cooking (22)	-.837**	-.788**	-.018	.701**	.345	.231
Cutting (18)	-.513*	-.497*	.027	.329	-.209	.469*
Hitting (25)	-.458*	-.626**	-.250	.441*	.166	.185
Making (31)	-.408*	-.487**	.108	.402*	-.125	-.076
Talking (21)	-.345	-.164	.200	.104	-.079	-.160
Walking (21)	-.634**	-.560**	.035	.273	-.214	-.092

Note.: pro – production frequency; rank – mean rank; typ – typicality; lex – lexical frequency

* $p < .05$; ** $p < .01$

Table 2.9 Correlations between lexical frequency, typicality, production frequency and mean rank (verbs)

	pro - rank	prod - typ	prod - lex	rank - typ	rank - lex	typ - lex
All (207)	-.505**	-.594**	.368**	.578**	-.354**	-.284**
Birds (44)	-.279	-.632*	.380*	.617**	-.004	-.261
Clothes (27)	-.645**	-.524**	.623**	.561**	-.259	-.317
Fruit (32)	-.453**	-.727**	.327	.567**	-.337	-.371*
Furniture(18)	-.502*	-.647**	.599**	.543*	-.503*	-.163
Sports (32)	-.613**	-.880**	.215	.572**	-.450**	-.339
Transport (24)	-.704**	-.484*	.404	.736**	-.652**	-.540**
Veg' (28)	-.800**	-.641**	.204	.535**	-.398*	-.235

Note.: pro – production frequency; rank – mean rank; typ – typicality; lex – lexical frequency * $p < .05$; ** $p < .01$

Table 2.10 Correlations between lexical frequency, typicality, production frequency and mean rank (nouns)

Lexical frequency failed to produce consistent significant correlations with other measures, especially within individual verb categories where it correlated significantly just once with typicality and once with production frequency. While there were more significant correlations involving lexical frequency with noun categories, these still only involved 3 with production frequency, 4 with mean rank, and 2 with typicality.

2.5. Discussion

2.5.1. Summary of main findings

The results of the category listing and typicality rating experiments can be summarised in terms of the differences and similarities that were observed between verbs and nouns. In category listing (following exclusions), participants listed fewer verbs in verb categories than they did nouns in noun categories both in terms of total responses and the number of different responses that were produced. A greater number of participants' responses were excluded from the verb categories than from the noun categories on the basis that responses were not verbs or nouns respectively. Responses within verb categories showed a greater tendency to appear in multiple categories (e.g. with *ripping* and *tearing* both listed as ways of *breaking* and *cutting* something), whereas noun responses tended to be discrete and only appeared in one category (with the categories used here at least).

In terms of similarities, participants appeared to use similar principles for completing both the category listing task and typicality rating task for both verbs and nouns. In category listing (again, after data exclusions), those responses listed by most participants were also listed earlier for both verb categories and noun categories. Typicality was consistently correlated with production frequency in the category listing task and also (albeit to a lesser extent) the order that responses were given for both verb categories and noun categories (although less consistent in verb categories). There was limited evidence that lexical frequency significantly correlated with typicality, or production frequency and mean rank (i.e. order of responses).

2.5.2. Discussion of main findings

Greater number of verbs in verb categories than nouns in noun categories

The finding that participants (following exclusions) listed greater numbers of nouns than verbs and a greater number of different nouns than different verbs is not surprising given that there are a greater number of nouns in the English language than verbs. There was also no overlap between mean number of responses in verb and noun categories with the verb category with the highest mean number of verb members (i.e. cooking; 10.37 verbs) falling below the noun category with the lowest mean number of members (i.e. tools; 13.14 nouns). The same is not true when considering number of different responses across all participants where the noun category with the lowest number of members (i.e. fruit; 69 nouns) is surpassed by four of the verb categories totals. Despite this, noun categories were still associated with significantly greater numbers of different category members. It could be argued that the noun and verb categories used here may not have been equivalent in terms of level of abstraction within a hierarchical organisation from least to most specific. Category headings here selected from the level just above the most populated level in the WordNet database (i.e. Miller & Fellbaum, 1991) as it was anticipated that this category selection would offer the fairest quantitative comparison of performance in category listing of nouns and verbs. Despite this, quantitative differences were seen and are probably a reflection of the greater number of nouns available compared to verbs.

Greater number of responses excluded in verb categories

The finding that more data was excluded within verb categories is insightful and is one of the significant differences in participants' performance in listing verbs and nouns within categories. Aside from the large numerical and proportional amounts of responses excluded within verb categories compared to noun categories, the qualitative differences are also striking. The majority of noun responses that were excluded were simply repetitions, either of responses already given by a participant within a particular category, or repetitious of the category itself (e.g. listing *hit* within the *hitting* category). It may be speculated therefore that responses were excluded from noun categories due to participants listing so many responses that they could not mentally keep track of responses they had previously given (which would assume they were also not visually keeping track by reviewing their responses). In comparison, the responses that were excluded from verb categories could be viewed as modifiers which expressed subtle semantic distinctions from the category verb. For example, a number of responses excluded from the *talking* category offered modifications of the manner of talking (e.g. *loudly*, *softly*, *quietly*). Similarly, within the categories *walking* and *running*, excluded responses included those expressing directionality (e.g. *backwards*, *forwards*, *sideways*) in addition to manner (e.g. *fast*, *slow*). The justification for excluding these responses in the current investigation was that they do not in themselves express idiosyncratic actions which could be expressed as a single verb and which could not therefore differentiate one particular action from another (e.g. most actions involving motion or movement could be modified with *slow*, *fast*, and so on). However, it should also be recognised that this justification may only be applicable for English verbs as verbs in languages other than English can often be lexicalised to differentiate directionality, manner, and so on.

From the participant's point of view, for nouns, the type of information required within the category listing task is clearly defined and is limited to other nouns which are in a subordinate relation to the category prompts (i.e. at a lower. or more specific, level of hierarchal structure). For verbs however, there are fewer verbs to choose from which are subordinate and more specific than the category prompt and so participants may look to exploit semantic representations about the action itself in order to make more fine-grained distinctions which may not be applied as readily in more rigid yet noun representations. The reason why participants gave such modifying responses for verbs can only be speculated at the present time. It may have been an attempt to provide a roughly equivalent number of responses in each category (as noun and verb categories were presented alternately). Equally, it may have been a way of filling time after exhausting all subordinate/specific verbs before the two minute period expired for that particular category – time which was not available for noun categories as there were more potential responses select from. For instance, Laws (2004) has demonstrated that participants are still able to give large numbers of responses within noun categories into a third minute for most object categories (that are traditionally used in such research) and even into a fourth minute for some categories (e.g. animals). Therefore, it would be difficult to anticipate if and when participants would use equivalent ‘modifying’ strategies in object category listing (e.g. *red apple*, *green apple*; *sharp knife*, *blunt knife*; and so on).

Greater polysemy of verbs compared to nouns

It was found that more verbs had a greater tendency to be listed within multiple categories than nouns which may be interpreted as demonstrating the greater polysemy (i.e. the tendency for the same lexical form to be used to express conceptually/ semantically/ thematically related senses which radiate from a core, or central sense, e.g. Fillmore & Atkins, 2000) of verbs where verbs tend to be associated with a greater number of meaning senses than nouns. For example, the verbs *tearing* and *ripping* both appear within the *breaking* and *cutting* categories just as *baking* appears in both the *cooking* and *making* categories. Also, the category verbs themselves appear within other categories, e.g. *cooking* is listed within the *making* category, and *cutting* and *hitting* are both listed in the *breaking* category. The greater polysemy of verbs used here is potentially demonstrated by the fact that the 10 nouns used as categories within the category listing task were associated with a total mean number of nouns senses of 3 (SD = 2.16) within the WordNet lexical database (Miller & Fellbaum, 1991) compared to the 10 verbs used as categories which had a total mean number of verb senses of 25.3 (SD = 20.05). It should be noted that the number of meaning senses in WordNet cannot be used as a direct measure to equate to polysemy as it represents meaning senses that are both related and unrelated. Nevertheless, the WordNet measure does give a tentative indication. The greater number of meaning senses for verbs over nouns here, in itself may give strong indications that a hierarchical category structure cannot be applied to the representation and organisation of verbs in the same way as it can with nouns. Such an argument may only be valid if it is considered that each meaning sense is represented as a discrete unit of semantic representation. Otherwise, the fact that verbs tend to be more polysemous may be a strong indication that verbs are organised more according to a three-dimensional matrix-like network (e.g. Grasser et al, 1987; Huttenlocher & Lui, 1979).

Greater polysemy of verbs may have been expected to cause problems for participants in completing a task such as category listing as this may cause their responses to diverge from the meaning intended by the researcher. However, this is not supported in the exclusion data where less than 5% of the raw responses were excluded from verb categories on the basis that the verb was being used in an 'inappropriate' sense (e.g. *to break up with someone* [*breaking* category]; *to make friends* [making *category*]). These findings therefore suggest that verbs are polysemous enough to allow flexibility in participants' responses within verb category listing, however, participants are generally consistent in accessing the core (i.e. the intended) meaning from the category prompt. The same is not the case for noun categories which had a tendency to be more discrete with little overlap between categories, apart from some examples which might traditionally be labelled as 'unclear cases' (e.g. *tomato* listed as both a *fruit* and a *vegetable*). There is naturally a sense that the degree of discreteness will be reflected in the choice of categories in terms of the number of categories investigated and also the level of categorisation (i.e. superordinate, basic level, or subordinate). Therefore, although these results in relation to polysemy may be expected, it should also be remembered that the categories used here represent only a small portion of the total verb and noun lexicon of English, and indeed of the conceptual domains of actions and objects, and so generalisations should be made with caution.

Production frequency, mean rank, typicality and lexical frequency

The finding that production frequency and mean rank of responses show a tendency to correlate significantly for both verb categories and noun categories is consistent with previous studies involving category listing (e.g. Battig & Montague, 1969, with nouns; and Hemeren, 1996, with verbs). The finding that typicality showed significant correlations with measures of production frequency and mean rank is also consistent with previous investigations into noun categories (i.e. Hampton & Gardiner, 1983; Mervis et al, 1976). The current findings are also consistent with previous research which has found that lexical frequency is not a consistent predictor of these other variables. These findings therefore extend the conclusion that typicality is independent of lexical frequency and can predict performance in category listing for both nouns and verbs. However, it cannot be known at this stage whether typicality influences performance in the category listing task, or indeed whether strength of semantic association between a category and a category member influences typicality.

A further indication that participants complete the typicality rating task with verb categories in a similar manner to noun categories comes from inspecting the distribution of typicality ranks within categories. For nouns and verbs, within individual categories, typicality ratings were skewed towards typical ratings as opposed to atypical ratings. This is, in some senses, to be expected given that all verbs that were rated for typicality were assumed to be definite members of the category (i.e. they were listed by at least three participants within category listing). The same patterns of skewness were also evident for noun and verb categories for both production frequency (i.e. more high-production frequency responses listed than low-production frequency responses) and mean rank (i.e. more responses listed earlier than later). One difference that did arise was in the shapes of the distribution for typicality ratings between noun and verb categories as indicated by the kurtosis values. An across-category analysis showed that verb typicality ratings showed an almost normal bell-shaped distribution (despite its skewness) whereas nouns showed a more peaked distribution implying a high concentration of nouns around a mean typicality value. These typicality distributions differ slightly to those found by Hampton & Gardiner (1983) who calculated the distributions over their 12 investigated noun categories. They found typicality to again be positively skewed towards typical ratings (as opposed to atypical ratings) although they found that kurtosis indicated an almost perfect bell-shaped distribution (kurtosis = 0.04) as opposed to a highly peaked distribution.

While the general patterns seen in the current investigation did only emerge in the across-category analysis, this may imply that noun categories, or at least those members that are most frequently listed within the categories, are more semantically coherent and similar thus making it more difficult to differentiate them in terms of typicality. In comparison, verb categories may be broader with differentiation easier to recognise which may then be reflected in the distribution of typicality ratings.

Conceptual/semantic differences

There are possible differences that are relevant between some of the verb categories and which may impact on some aspects of participants' performance, or make it problematic to compare quantitative results. Categories such as those associated with movement (e.g. *jumping*, *running*, *walking*) showed results which suggested the categories are relatively small, even in comparison to most other verb categories. It may have been that if a more general category was included instead, such as *moving* (as in Hemeren, 1996), rather than being split into more specific ways of moving, the quantitative patterns may have been different. This may also indicate that *jumping*, *running*, and *walking* are potentially at a lower level of hierarchical organisation than the other verbs used as categories. It should also be observed that these categories along with *talking*, differ from the other verb categories both syntactically and thematically in that they are intransitive (requiring only a single argument and thematic role) as opposed to the other verbs used as categories which are canonically transitive (requiring two syntactic arguments in two thematic roles). As such, the lower number of category members may be reflective of the possibility that it is more straightforward to list category members for verbs which are transitive as they may evoke a wider variety of situations. For example, it may be easier to imagine situations where you might *break* different objects in a variety of different ways, compared to thinking of situations where you might *run* in different ways.

A further distinction that may be made between the verbs used as categories and the nouns used as categories may be in relation to their location within the vertical dimension of categorisation (see section 2.2.2). As stated previously, these verbs and nouns were selected from the level above the most densely populated level of hierarchical organisation within the WordNet database. The nouns used here as categories are consistent with those employed in previous research (e.g. Hampton & Gardiner, 1983; Rosch et al 1976) and are usually considered to be at a superordinate level of abstraction within the vertical dimension of categorical organisation. In comparison, some have suggested that at least some of the verbs used here would be located at a basic level of abstraction within the vertical dimension (e.g. Lakoff, 1987; Miller & Johnson-Laird, 1976). It may therefore be argued that this difference may lead to some of the differences observed between performance in listing within verb and noun categories, or at least make it problematic to make fair comparisons. However, it is not clear whether such terminology can be applied equivalently between conceptual domains of actions and objects and lexical domains of verbs and nouns. If it were assumed that the basic level is conceptually equivalent across objects/nouns and actions/verbs then it would be assumed that *apple*, *chair*, and *car* are equivalent in terms of semantic content (e.g. comparable numbers of semantic features) to *breaking*, *talking*, and *walking*. From a purely intuitive sense, in terms of the specificity and distinction between and within organisation levels, it may appear that *apple*, *chair*, and *car* are more conceptually equivalent to *smashing*, *chatting*, and *strolling* and that concepts *breaking*, *talking*, and *walking* are more conceptually equivalent to *fruit*, *furniture*, and *transport*. This is an issue worth further investigation as it is not clear that terminology of categorisation principles can be applied equivalently across the conceptual and lexical domains of objects and actions and nouns and verbs respectively. This is especially salient, if it is assumed that nouns and verbs are organised according to fundamentally different architectures, where verbs are organised within a three-dimensional matrix rather than a two-dimensional hierarchy (e.g. Grasser et al, 1987; Huttenlocher & Lui, 1979).

2.5.3. Limitations and further research

One limitation associated with the category listing task concerns the number of participants involved. Large scale category norms, such as Battig & Montague's (1969) set and the revised and expanded set by Van Overschelde et al (2004), have used much larger samples of participants (442 and in excess of 600 respectively). Even Hemeren's (1996) study involved 119 English speakers in the investigation of production frequency and order of listing despite not attempting to offer a set of 'category norms' as such. Larger samples naturally afford greater ability to generalise the patterns found. Given that this research has suggested that category listing tasks are a valid method of investigating semantic associations and organisation, similar research could now be conducted on a larger scale to involve more participants and possibly an extended range of categories.

A further possible adaptation of the category listing task would be to have participants respond orally as in Van Overschelde et al (2004). This may be insightful in that it may aid understanding as to why there was more data excluded within verb categories than there was in noun categories. Response time data may give indications as to the degree to which participants were finding it difficult to give verb responses. Even inspection of raw audio recording may be insightful if it indicates that participants were giving numerous filled pauses (e.g. 'ums' and 'ers') as opposed to verbal responses (whether they would be excluded or not from final analysis). Further insight may also be gained if participants were systematically probed with questions following the main task to assess whether they felt the task was difficult or whether they found any particular aspects or categories difficult. This was not done within the current investigation although a number of participants informally commented that they did indeed find the verb categories more challenging.

In further research it would be necessary to ensure that typicality ratings for verbs were not merely reflections of familiarity and other non-semantic variables. Hampton & Gardiner (1983) collected familiarity ratings alongside their typicality ratings for noun categories and showed the degree to which typicality and familiarity were independent and this has also been demonstrated in other work (e.g. Boster, 1988). In hindsight, it would have been desirable to do so within the current investigations, especially since retrospective analyses using external ratings of familiarity (e.g. via the CELEX database, Baayen, Piepenbrock & Gulikers, 1995) only allow for extraction of familiarity ratings for a small subset of the verbs used in the current investigations.

2.6. Conclusion

The investigations reported in this chapter demonstrate the usefulness of applying category listing and typicality rating methods in the domain of actions and verbs. The results do not, and were not ever going to, argue against the proposition that actions/verbs are organised according to fundamentally different principles to objects/nouns (e.g. Grasser et al, 1987; Huttenlocher & Lui, 1979). However, what the results do show is that participants are able to complete these tasks by applying comparable principles of semantic organisation. In other words, participants can list verbs in a hierarchical manner if asked to do so (which is the restriction that a category list task imposes). There do appear to be numerous factors which may make category listing a more difficult task for verbs than for nouns, however, some of these apparent difficulties (e.g. more responses excluded) may simply be a reflection of fewer verbs available to select from and a broader network of semantic associations and modifications which allow non-verb responses. Participants are also able to rate verbs' typicality within categories in an apparently similar manner to rating typicality of nouns within categories.

What needs to be investigated further is why certain actions are grouped within a category and what underpins typicality within categories. These themes will be explored in the following chapter.

Chapter 3 Investigating Semantic Similarity between Verbs

3.1. Aims of Chapter

This chapter investigates the notion of semantic similarity between verbs. Chapter two demonstrated that healthy adult speakers are able to list verbs that are categorically related to a general verb denoting a related action. It was also shown that participants can rate verbs according to their typicality in relation to other members of the same category. In previous research, explanations of how speakers give such responses and judgements are based upon semantic similarity between the concepts that the words denote (e.g. Rosch & Mervis, 1975). Semantic similarity is generally considered to arise as a consequence of semantic featural representations that are encoded as part of concepts within semantic memory.

This chapter reports the outcomes of two series of experimental procedures which aimed to consider semantic similarity from both a perceptual perspective (i.e. the extent to which speakers perceive and rate actions to be similar) and a decompositional perspective (i.e. as a function of semantic feature composition). Firstly, semantic similarity was investigated using a similarity judgement task where participants were asked to rate how semantically similar pairs of verbs were (e.g. *baking – frying*). Secondly, semantic similarity was investigated through a series of analyses of actual semantic feature composition using an existing set of semantic feature norms gathered for verbs (i.e. Vinson & Vigliocco, 2008).

This chapter begins with a discussion of semantic similarity and how most models of semantic processing consider this to be a function of semantic feature composition. This will include discussion of how previous research has investigated semantic similarity between concepts by visually representing this within a simulated semantic space. The chapter then clarifies how the term 'semantic feature' is being used in the current investigations and explores research themes which have developed in relation to semantic features and their role in semantic similarity. This includes discussion of analyses using: (1) feature types; (2) feature correlations; and (3) feature distinctiveness. This will also consider how these analyses have been applied, for example, in investigating similarity between and within categories in order to explain category specific deficits and typicality effects within behavioural tasks. This will also consider how features have been investigated in relation to concepts at differing levels of specificity (i.e. categorisation; e.g. superordinate/general levels to subordinate/specific levels). A brief discussion will also follow on how semantic features have been argued to differentiate the grammatical classes of nouns and verbs. Although this is not directly investigated further in the analyses presented within this chapter, as the current chapter focuses exclusively on semantic similarity between verbs, this is a significant issue to introduce within the context of this thesis. These discussions will lead into the reports of the series of experimental investigations and their subsequent discussion of their respective outcomes.

While hypotheses regarding semantic similarity between verbs are made on the basis of previous research which has been conducted to investigate semantic similarity between nouns, the analyses presented here are primarily exploratory. For reasons which will be discussed later (see section 3.7.3), the current analyses are limited in their scope and so the results may be difficult to generalise. However, the results do provide interesting insights into the semantic featural composition of verbs, especially when considered in the context of the parallel investigations into nouns.

3.2. Background

3.2.1. Semantic memory, semantic features and semantic similarity

Most theories of semantic memory specify a role for semantic features in representation of concepts. However, these theories differ in terms of how semantic features are represented. Semantic network models (e.g. Collins & Loftus, 1975) specify that semantic features are themselves represented as units, or concepts, within semantic memory and that these feature units are activated through associative links when the central concept node is activated. Other models assume that featural representations are distributed across different neural systems which specialise in processing different types of information. For example, there are models which differentiate different types of knowledge (e.g. Coltheart et al, 1998) and those that go further and attempt to map featural knowledge onto specific specialised neural areas (e.g. Allport, 1985). These latter theories consider that concepts are decomposable into distributed patterns of activation across semantic features.

Semantic features dictate semantic similarity between concepts to the extent that features differentiate domains or categories of knowledge (although see Medin, 1989 who also stresses the importance of situational context in conjunction with featural composition in determining similarity between concepts). Rosch and colleagues conducted much of the research which prompted the move away from classical approaches to categorisation whereby category membership was considered ‘all-or-nothing’ and all category members were considered equal in terms of their status within the category (e.g. Mervis et al, 1976; Rosch, 1978; Rosch & Mervis, 1975; Rosch, Mervis, et al, 1976; Rosch, Simpson & Miller, 1976). This worked helped to formalise and provide empirical support for the ideas that were discussed by Wittgenstein (1953) who described categories of knowledge as having 'fuzzy' boundaries and claimed that category membership operated on the basis of 'family resemblance' in that there would be few if any identifiable features that would be common to all members of a particular category despite there being clear featural overlap within subsets of members:

> Consider for example the proceedings that we call "games". I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all? ... if you look at them you will not see something that is common to all, but similarities, relationships, and a whole series of them at that ... Look for example at board-games, with their multifarious relationships. Now pass to card-games; here you find many correspondences with the first group, but many common features drop out, and others appear. When we pass next to ball-games, much that is common is retained, but much is lost. – Are they all 'amusing'? Compare chess with noughts and crosses. Or is there always winning and losing, or competition between players? Think of patience .... The result of this examination is: we see a complicated network of similarities overlapping and criss-crossing; sometimes overall similarities, sometimes similarities of detail (p31-32)

Before describing semantic features in more detail, the following section will present an overview into how semantic similarity between concepts has been visually represented. This is important to consider as it often provides a more tangible representation of the effect of featural representation and semantic similarity in general which can add validity to further in-depth investigations.

3.2.2. Modelling concepts/words within a dimensional semantic space

Various methods have been used to map concepts/words within dimensional space in order to provide a visual representation of semantic similarity between concepts/words. Hemeren (1996) employed multidimensional scaling techniques following a category listing task where participants listed actions within the category of 'actions that involve some kind of *bodily activity that can easily be recognised when seen and can be visualised as a mental image*' (p44; Hemeren's italics). Hemeren transformed mean rank positions between verbal responses into proximal distances on the assumption that responses that were listed sequentially were represented closer in semantic memory than those where other responses intervened. Although Hemeren did not attempt to provide any speculation on the dimensions along which the verbs were distributed, the resulting three-dimensional analysis (*stress* = 0.209, R^2 = 0.64; see Figure 3.1) showed that within the general category, two clusters of actions emerged based around motion to or from a location (e.g. *run*, *jump*, *swim*) and vocal or mouth actions (e.g. *talk*, *laugh*, *sing;* i.e. between dimensions 2 and 3: lower portion of Figure 3.1).

Vinson & Vigliocco (2002) presented two-dimensional representations of semantic space for both object and action concepts. These were produced from averaging a series of computationally derived self-organising maps based on production frequencies across participants of semantic features that were listed for individual concepts within a semantic feature listing task (see Figure 3.2). Comparing the two subsets of concepts, it was demonstrated that object concepts clustered within discrete categories (e.g. fruits and vegetables in Figure 3.2) whereas actions were represented within a smoother space where there were gradual changes between semantic fields (e.g. verbs of sound emission, communication, and exchange within Figure 3.2). This was argued to provide an explanation as to why category specific deficits are frequently observed in cases of language impairment that affect object concepts (i.e. the production of nouns) on the basis that a focal lesion can effectively damage category members that are represented within a close proximal distance. In comparison, a focal lesion would be unlikely to lead to such category specific effects in action concepts (i.e. the production of verbs) as categories, or semantic fields, are less discrete and show a greater degree of overlap. Subsequent lesioning of the simulated maps did reproduce such category specific observations although it would still be an assumption to consider that such computational representations can be neatly mapped onto discrete neural areas, especially given Vinson, Vigliocco and colleagues' central claim that objects and actions are represented within a unitary semantic space (e.g. Vigliocco et al, 2004).

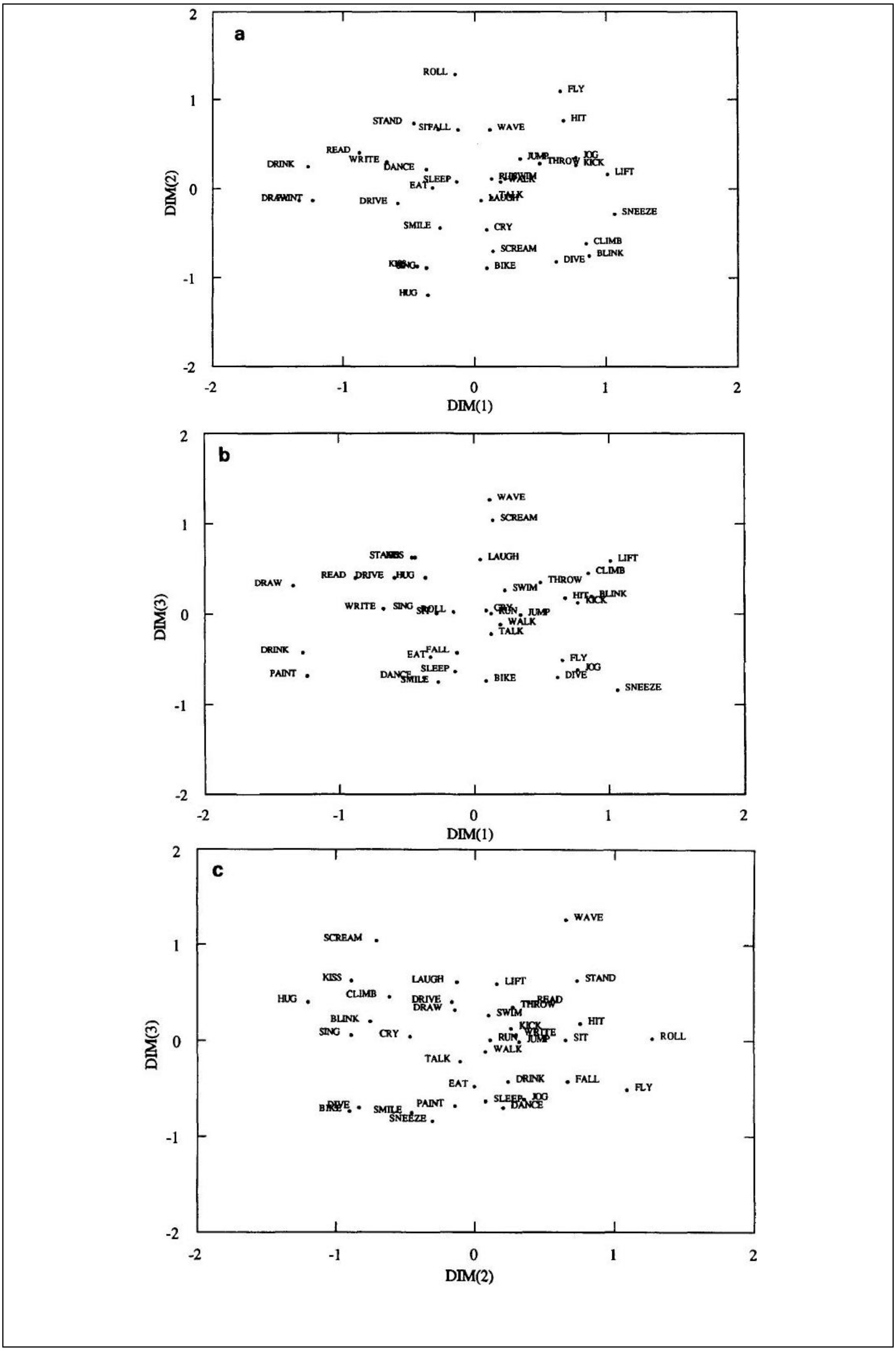

Figure 3.1 Three-dimensional scaling of actions (from Hemeren, 1996)

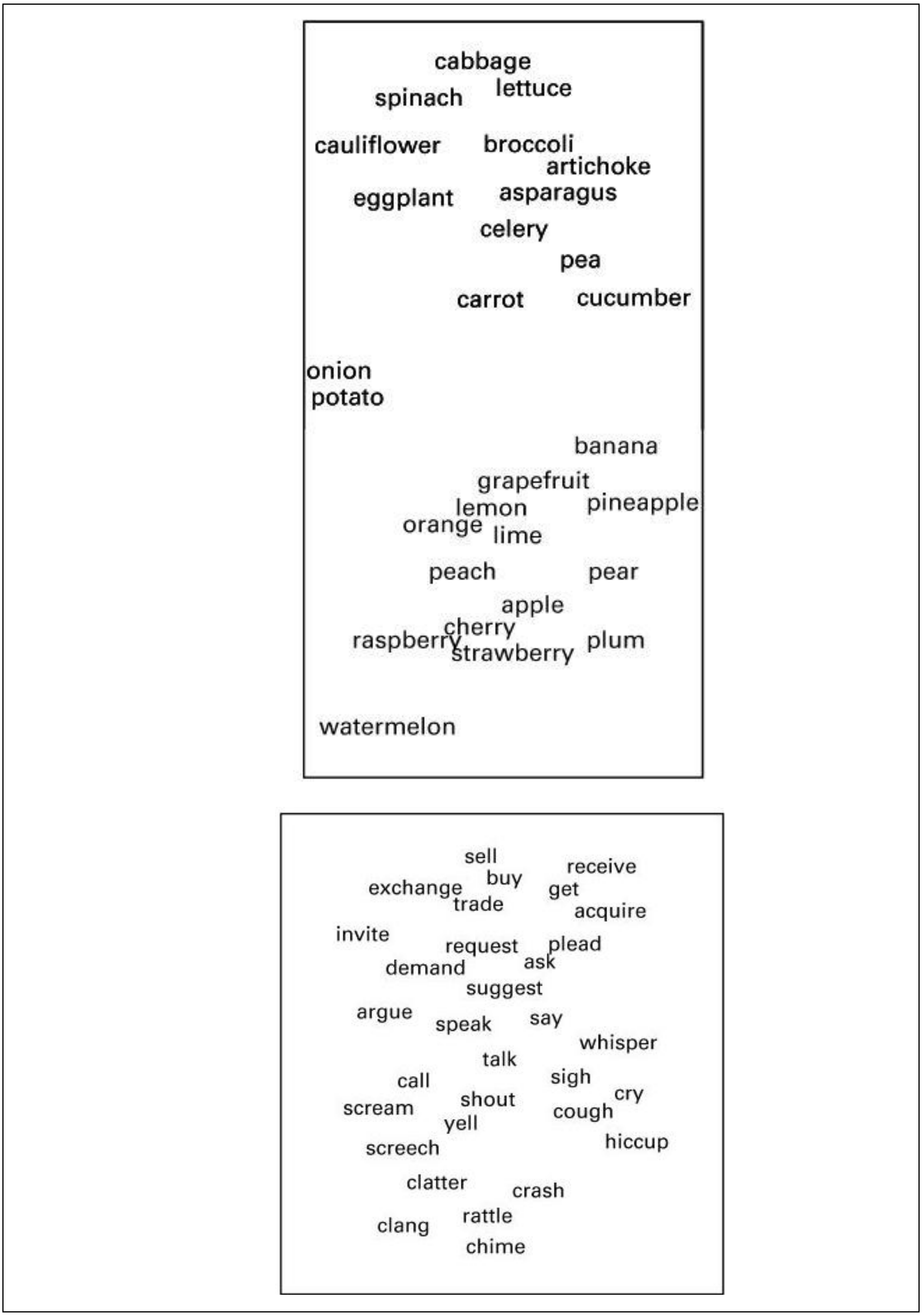

Figure 3.2 Self-organising maps of subset of objects (above) and actions (below) (Vinson & Vigliocco, 2002)

While acknowledging that Hemeren (1996) and Vinson & Vigliocco (2002) derived their representations of semantic space using different methods, they nevertheless present an insightful contrast with regard to the derived representation of actions. Hemeren's representation demonstrates that categories, or semantic fields of actions can be disassociated within semantic space (i.e. there is a division, or gap between them). This was achieved within the broader category of 'actions that involve some kind of *bodily activity that can easily be recognised when seen and can be visualised as a mental image*' which acted to define the scope of responses that participants gave within the experiment. This is in contrast to Vinson & Vigliocco (2002) whose representations were derived from features gained from a wide selection of actions cutting across several semantic fields. This consequently led to a lack of discreteness in the semantic representations although there were some exceptions with some fields described as showing relative clear boundaries (e.g. actions of light emission). This may therefore demonstrate that semantic representations of verbs can yield distinctions in categories until the scope of representations becomes too broad or until the measure on which similarity is judged becomes too insensitive.

Romney et al (1996) provide a further investigation using multidimensional scaling of semantic similarity with the additional consideration of typicality within semantic categories of objects (e.g. *vehicles*, *vegetables*). Under assumptions of family resemblance theory (e.g. Rosch & Mervis, 1975), it was hypothesised that typical category members should cluster in relatively close proximity to one another and to the category prototype with less typical category members being more widely distributed from the category prototype (i.e. the centre of the semantic space). Across four semantic categories, evidence for this was most convincing for *vehicles* and least convincing for *vegetables* (see Figure 3.3, where typicality is represented according to the size of circles representing individual category members). Given the variability in typicality reflecting degree of semantic similarity, Romney et al (1996) concluded that caution needs to be taken when inferring the precise principles that participants apply when making typicality rating judgements as it may not always be a straightforward production of semantic similarity in terms of semantic features as is assumed in some models of categorisation.

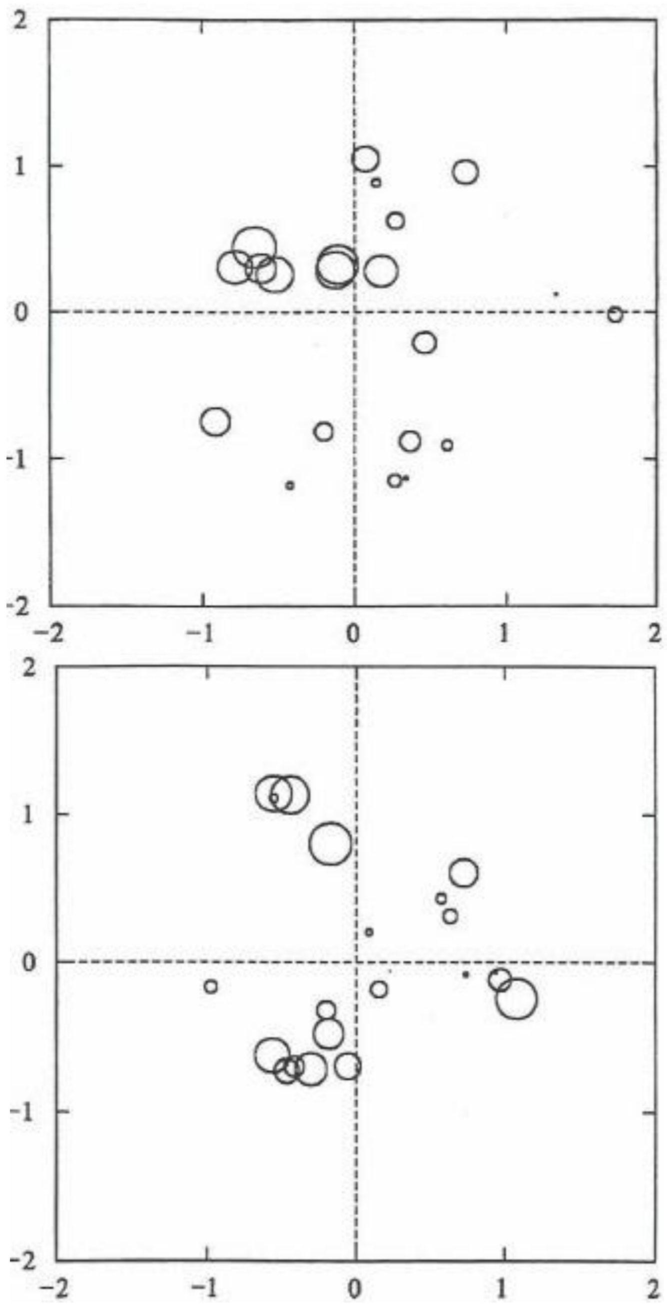

Figure 3.3 Two-dimensional scaling within vehicles (above) and vegetables (below) (Romney et al, 1996)

3.2.3 Semantic features

There are two principle approaches to identifying semantic features belonging to concepts. Some consider semantic features to be abstract primitive concepts which are inherent to meaning (e.g. Jackendoff, 1983; Pinker, 1989). For example, a concept such as tiger would encode abstract primitive features such as THING and ANIMATE (among others) which would be differentiated from a concept such as hammer which would encode features THING and INANIMATE (among others). The difficulty with such positions is that the features are claimed to be highly abstract and as such difficult to empirically identify.

Other approaches to identifying semantic features have involved asking speakers to list the features they believe are relevant for the concept. Most studies which collect speaker-generated semantic features are generally in agreement with what they regard to constitute a semantic feature. For example, Rosch et al (1976) merely stated in their task instructions that an attribute is something that is ‘common to and characteristic’ of the concept in question and most studies display a similar sentiment in their instructions in that the feature given should contribute some aspect to the meaning of the concept. Therefore, this approach may lead to identifying features for tiger including <*has stripes*>, <*lives in the jungle*>, <*a type of feline*> which are generally more concrete than those primitive features used in other approaches. Most studies also agree on what is not a semantic feature. Studies generally discourage (either by stating in the instructions, or through exclusion of data) features which are based on word associations, such as *tea-coffee*, or *hop*-*skip-jump*. Knowing that *tea* and *coffee* are associated is generally not informative in terms of understanding what the concept actually means. Rather, the purpose of collecting speaker-generated features is to understand why these concepts are associated on the basis of similarity and what features they have in common, such as both being drinks containing caffeine which are usually served hot and which can be accompanied with milk and/or sugar.

The primary limitation of this second approach is that it attempts to investigate conceptual semantic representations mediated through lexical semantic representations. In other words, in order for participants to generate features, they are consciously thinking and lexicalising representations which for day-to-day purposes are supposedly subconscious. This has implications in that it has been argued that when speakers are asked to generate features, there is a bias towards generation of features that are distinctive as opposed to features which are ‘obvious’ (e.g. Cree & McRae, 2003; Medin, 1989). For example, when listing features for *tiger*, few people will generate *<has skin>*, *<breathes>* despite these being fundamental attributes of tigers. Despite the reservations, the use of semantic feature listing tasks has proved to be a popular method of researching semantic memory and of representations stored within. Ashcraft (1978) collected features to investigate the featural distinction between typicality and atypical category members, whereas Garrard, Lambon Ralph, Hodges & Patterson (2001) collected features to investigate the living-non-living things dissociation and

Marques (2007) collected features to investigate featural differences in superordinate and basic level concepts. Such studies have however tended to draw conclusions on semantic features generated for a relatively small sample of concepts which are sufficient for the purpose at hand but which may be limited in terms of making generalised claims beyond the concepts under investigation.

There are currently at least two large collections of semantic norms which have been developed for English which may be useful in overcoming the potentially limited ability to generalise previous findings. The first, collected by McRae et al (2005), comprises speaker-generated semantic features for a total of 541 basic level concepts which include living (e.g. *dog*) or non-living (e.g. *chair*) things. A second set of norms by Vinson & Vigliocco (2008) is, however, of greater significance to the current thesis as it claims to be the first to have collected speaker-generated semantic features not only for objects (i.e. nouns) but also for events and actions (i.e. either as nouns, e.g. the construction, or verbs, e.g. to construct).

The discussion will now consider some of the main themes and findings in relation to semantic features and their role in dictating semantic similarity. As should be clear from discussion so far, most research has been conducted in relation to object concepts and has used nouns as stimuli (i.e. the stimuli used to elicit semantic features). As the experimental procedures presented later in this chapter focus on exploring semantic similarity between verbs, the following sections aims to highlight findings which may allow some speculative hypotheses to be formed regarding the featural composition of verbs. Despite this, some discussion will be presented with regard to featural differences that have been observed between objects and actions through the use of the Vinson & Vigliocco (2008) set of feature norms.

3.2.4. Research using semantic features

McRae & Cree (2002) highlight that there have been three principal approaches to the investigation of semantic features in semantic memory representations and semantic similarity. These approaches have been the investigation of: (1) feature types; (2) feature correlations; and (3) distinguishing features.

Feature types

The investigation of feature types arose out of the literature surrounding category specific deficits in language impairment (e.g. see Capitani et al, 2003; and Mahon & Caramazza, 2009, for reviews of category specific deficits). Warrington &

McCarthy (1983) were among the first to suggest that features could be differentiated according to the type of information they encoded. They made this suggestion following investigations with patient V.E.R who had aphasia following a left hemisphere infarct. V.E.R showed a selective impairment of general objects compared to a selective preservation of categories of food, animals and flowers. It was suggested that such a pattern could have arisen through damage to certain types of features, specifically functional features (i.e. semantic knowledge of how something moves of how something is used) which were argued to be more influential in differentiating objects. This was in comparison to preserved sensory features (e.g. semantic knowledge of something's appearance, e.g. size, shape, and colour) which were argued to be more influential in differentiating living things.

Support for this 'sensory-functional' feature hypothesis was provided by Farah & McClelland (1991) who constructed a computer simulated model of semantic memory. In the model, living and non-living things were represented by different proportions of either sensory or functional features. The differential proportions were determined according to data that was obtained in a pre-test where participants identified sensory and functional features that made up dictionary definitions under the names for various living and non-living things. It was found that living things had a sensory-functional feature ratio of 7.7:1 compared to non-living things whose ratio was 1.4:1. When the model was developed with similar weightings for features, it was found that damage to sensory features led to poorer activation of living things whereas damage to functional features led to poorer activation of non-living things. Further support is provided in behavioural tasks where it has been found that participants generate features which are considered to be important for the category that an item comes from. For example, for living things, participants have a preference for listing sensory features (i.e. there are more listed and they are listed earlier) whereas they tend to prefer listing functional features for non-living things (e.g. Vanoverberghe & Storms, 2003). In the same study, it was also found that the presence of sensory features was more predictive of participants' typicality ratings of items in categories of living things whereas presence of functional features was more predictive of typicality ratings of non-living things.

Whilst the distinction between sensory and functional features may be adequate for distinguishing the broad categories of living and non-living, it has been argued that such a two-way distinction of feature types is both unrealistic and implausible in terms of capturing all that speakers know about something and also in terms of explaining all of the various patterns of category specific deficits that have been observed (e.g. Cree & McRae, 2003). Therefore, the investigation of feature types has extended beyond simply a sensory and functional distinction and this is reflected in large-scale sets of features norms. For example, McRae et al (2005) classify feature types along a nine-way distinction which was derived from Wu & Barsalou's (2002, cited in Cree & McRae, 2003) Knowledge-Type Taxonomy which originally identified 28 different feature types. These nine feature types were: visual-colour, visual-parts and surface properties, visual-motion, smell, sound, tactile, taste, function, and encyclopaedic. In another large-scale set of feature norms, Vinson & Vigliocco (2008) classified features as one of five types (i.e. visual perceptual, other perceptual, functional, motoric, and other features which included taxonomic and encyclopaedic knowledge). Analyses of feature types which cover more distinct types of features have been useful in further explaining category dissociations and distinctions (see section 3.2.5) and also conceptual/grammatical class distinctions between objects/nouns and actions/verbs (see section 3.2.8).

Feature distinctiveness

Distinctive features are those that allow fine-grained discrimination between concepts. For example, within the category of living things the feature <*moos*> would be highly distinctive (i.e. in identifying a cow) as it is associated with relatively few living things in comparison to the feature <*eats*> which would be associated with all living things (in one form or another). Distinctiveness therefore complements the notions of shared features and family resemblance (e.g. Rosch & Mervis, 1975). Distinctiveness is typically presented as a proportion (e.g. from zero to one, where zero indicates highly distinctive and one indicates not distinctive, i.e. shared amongst all members) of concepts possessing a particular feature. This has variously been calculated within categories (i.e. as a proportion across a relatively small set of related concepts, e.g. Garrard et al, 2001) or across categories (i.e. as a proportion across a large set of related and unrelated concepts, e.g. McRae & Cree, 2002).

McRae & Cree (2002) found that a total of 22% of semantic features associated with living things were distinctive (i.e. possessed by only 1 or 2 concepts across a large set concepts) and that each concept for a living thing possessed a mean of 3.2 distinctive features. In comparison, a total of 42% of features associated with non-living things were distinctive and each concept possessed a mean of 5.2 distinctive features. Garrard et al (2001) provided complimentary evidence on distinctiveness but also considered feature type. It was found that feature distinctiveness was distributed in a 'U' shape across most feature types with more features at the extreme ends of the distinctive continuum (i.e. either highly distinctive or highly shared) and fewer features falling in the mid-ranges of distinctiveness. In comparison, non-living things demonstrated a strong bias towards possession of distinctive features which was again across all feature types.

Distinctive features have been shown to receive preferential processing in behavioural tasks. Cree, McNorgan & McRae (2006) showed that participants are faster to verify that <*purrs*> is a property of cat than they were to verify that <*eats*> is. All animals will eat whereas very few will purr, hence <*purr*> is a distinctive feature. Therefore, this was argued to demonstrate that distinctive features are accessed more readily when retrieving conceptual-semantic representations. Despite distinctiveness being identified as a significant dimension in conceptual-semantic organisation, there has so far been little research into the role of distinctiveness in the representation of actions.

Feature correlations

Features are said to be correlated when the presence of one feature predicts the presence of another feature. For example, within the category of birds, the presence of the feature <*small*> predicts the presence of the feature <*sings*>, whereas the presence of the feature <*large*> predicts the presence of the feature <*talons*> (e.g. in eagles, vultures, etc; see Malt & Smith, 1984). Features may be also be negatively correlated where presence of a feature predicts the absence of another feature, or vice versa. The idea of feature correlation arose out of the work of Rosch et al (1976) who observed that features tended to cluster within and across concepts that belonged to particular categories, for example, features such as <*has feathers*> and <*flies*> tend to be possessed by most, but not all, birds. Malt & Smith (1984) also found that feature correlations were generally a better predictor of a within category typicality compared to Rosch & Mervis' (1975) measure of family resemblance which was based on presence of individual features.

Feature correlations have been shown to vary within different domains of knowledge (i.e. living vs. non-living). Devlin et al (1998) constructed a simulated model of semantic memory on the basis that natural kind (i.e. living thing) category members possessed a greater number of inter-correlated features than artefact (i.e. non-living) category members. When the model was lesioned, concepts with greater numbers of correlated features were more resistant to mild damage as features with damaged direct activation links could still be activated via their correlated features. This meant that in the early stages of lesioning, the model performed worse in activating artefact concepts than natural kind concepts. However, as severity of lesioning increased, the number of intact feature correlations decreased and performance in natural kind concepts quickly declined to a level below that of artefact concepts. Further work by Tyler, Moss and colleagues (e.g. Moss, Tyler, Durrant-Peatfiled & Bunn, 1998; Moss et al, 1997) suggests that the claim that living things possess more correlated features than non-living things is an over-simplification. They suggest that living things possess a greater number of feature correlations within shared features whereas non-living things have more feature correlations within distinctive features. It was also noted however, that overall, non-living things still possessed fewer correlated features with distinctive features than did living things.

McRae, de Sa & Seidenberg (1997) have found that overall the number of feature correlations that are significant tends to be relatively low. McRae & Cree (2002) have also argued that feature correlation as an organisational principle on its own cannot account for all trends observed in category specific impairments. Malt & Smith (1984) were also doubtful as to whether speakers are consciously aware of correlations between features to the same level that they are aware enough to list features individually.

There has been little investigation in terms of feature correlations within action concepts. Vinson, Vigliocco, Cappa & Siri (2003) did however report that the average correlation coefficient for feature correlations associated with actions was relatively low (0.081) and was significantly lower than feature correlations associated with animals (0.146) and tools (0.119). Therefore, while it is still an under-researched aspect of semantic representation of action concepts (and verbs), it could be speculated that the likely influence of feature correlations would be small, especially given the fact that verbs are more polysemous than nouns and may have a looser and/or broader conceptual representation.

3.2.5. Features as a basis for categorisation

Both Garrard et al (2001) and McRae & Cree (2002) report the findings of hierarchical cluster analyses whereby clusters of categories were derived from the presence or absence of semantic features. Both analyses demonstrate that semantic feature composition is sufficient to dictate semantic similarity between objects so as to identify categories that are consistent with patterns reported in speakers with category specific deficits.

Garrard et al (2001) report a cluster analysis performed on basic level concepts (e.g. *apple*, *mouse*, *candle*, and so on) that was derived on the basis of presence or absence of individual features. The resulting cluster analysis (see Figure 3.4) demonstrated discrete clusters of concepts that were consistent with membership of superordinate categories. For example, *fruits* (e.g. *apple*, *cherry*, and *orange*) clustered together, *mammals* (e.g. *mouse*, *tiger*, *cow*) clustered together, as did *birds*, *vehicles*, and a larger group of other objects (e.g. *barrel*, *scissors*, *toothbrush*).

McRae & Cree (2002) report a cluster analysis based on the weightings of feature types in 37 categories at a superordinate level of categorisation (e.g. *bird*, *vegetable*, *tool*, and so on). The resulting cluster analysis (see Figure 3.5) identified two broad clusters, or categories consisting of either living or non-living things with the exception of musical instruments which clustered with living things (although this may not be unexpected given that musical instruments, along with living things, tend to be differentiated according to visual features, as opposed to functional features which tend to differentiate non-living things).

Cree & McRae (2003) report a further cluster analysis of superordinate level concepts which was based on weightings of feature types and also included information about several susceptibility factors associated with the superordinate concepts (i.e. distinguishing features possessed by the concept, feature distinctiveness, visual similarity, semantic similarity, visual complexity, familiarity, lexical frequency, and percentage of correlated features possessed by the concept). This was argued to produce the most satisfactory hierarchical cluster analysis (see Figure 3.6) as the resulting clusters could best account for seven trends that had been identified in relation to category specific deficits in language impairment (e.g. that creature categories tend to pattern together; that fruits and vegetables can be impaired with either living or non-living things; that musical instruments can pattern with living things despite them being non-living; and that deficits of living things are more frequently observed than deficits of non-living things).

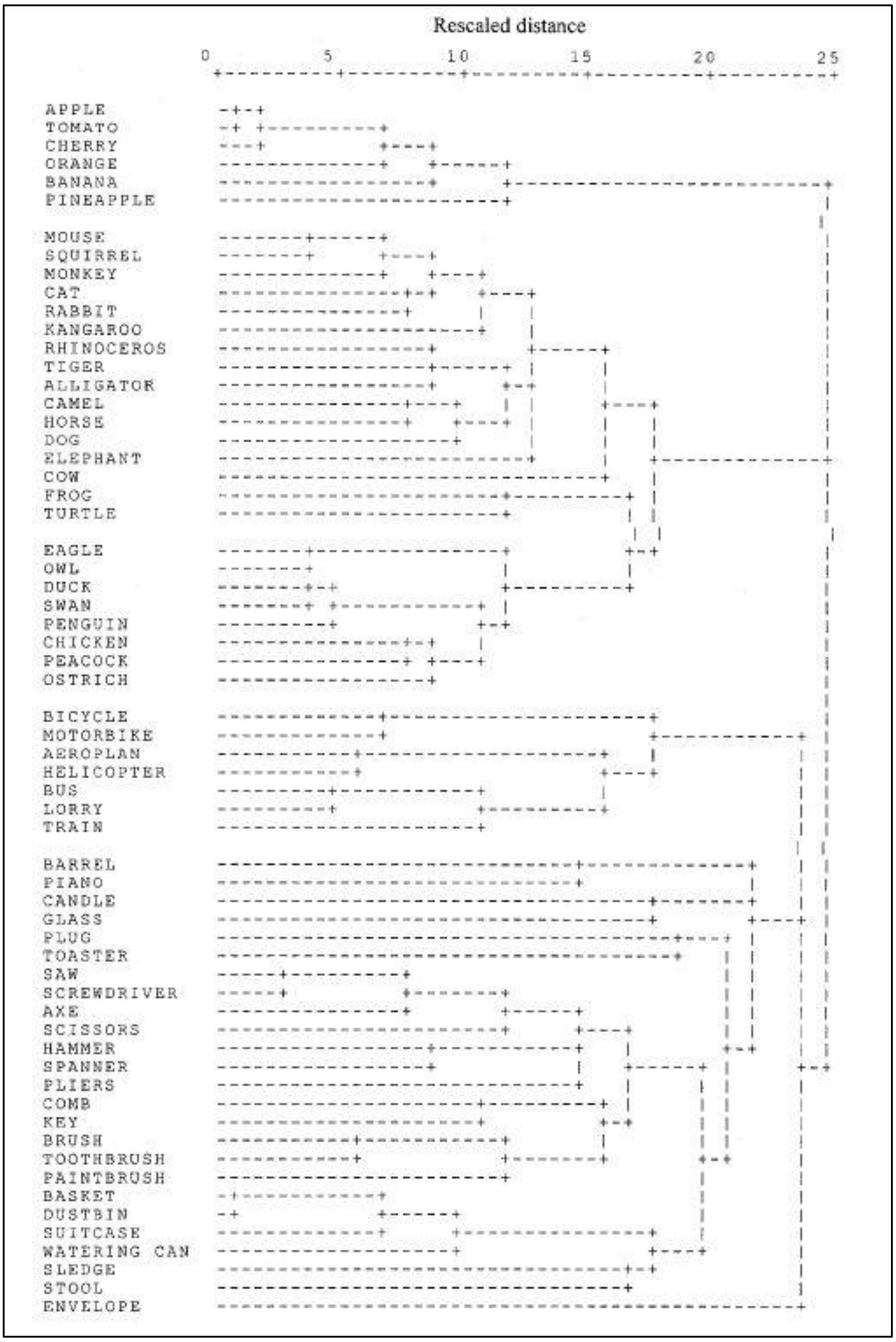

Figure 3.4 Cluster analysis of basic level concepts (Garrard et al, 2001:134)

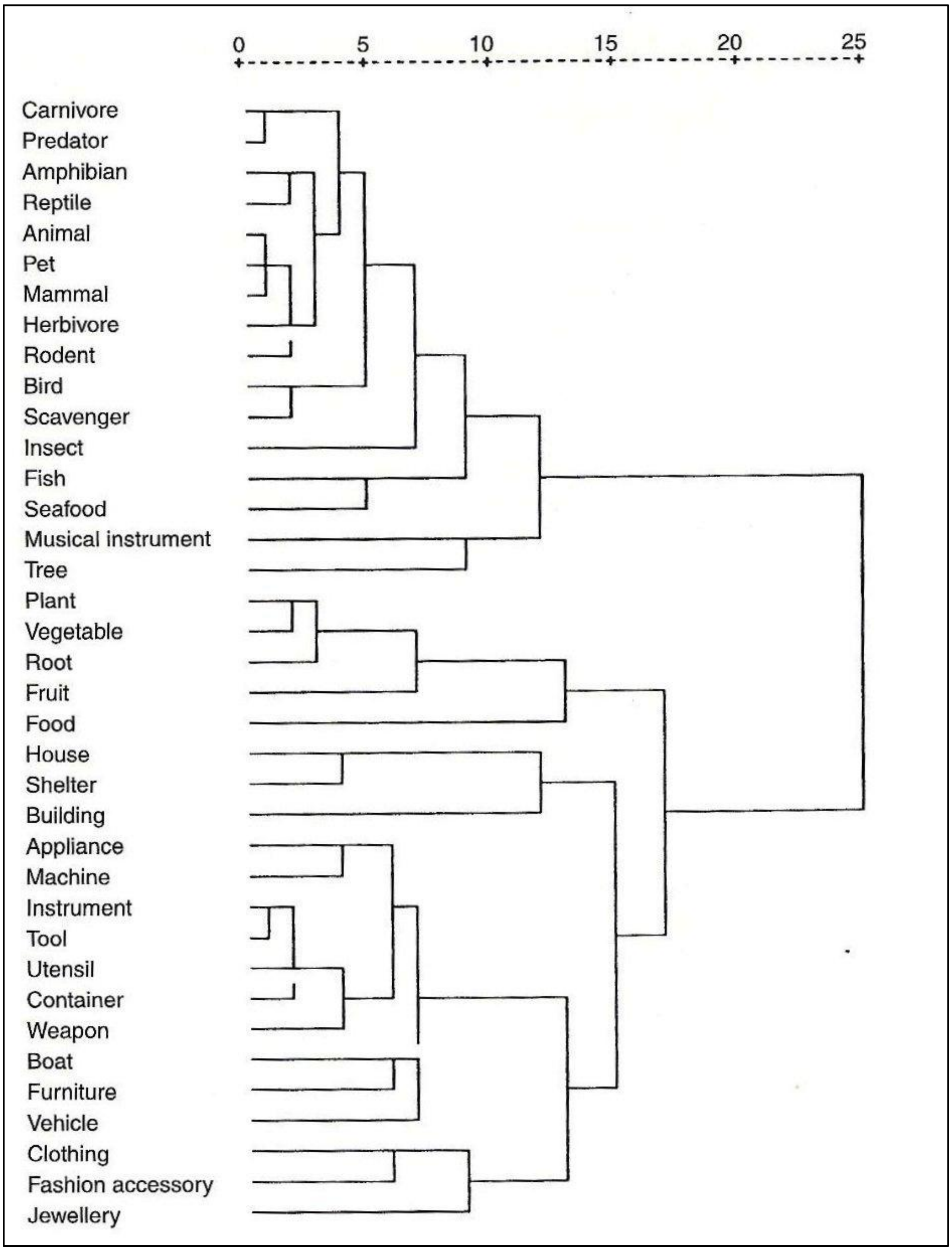

Figure 3.5 Cluster analysis of superordinate level object concepts (McRae & Cree, 2002:231)

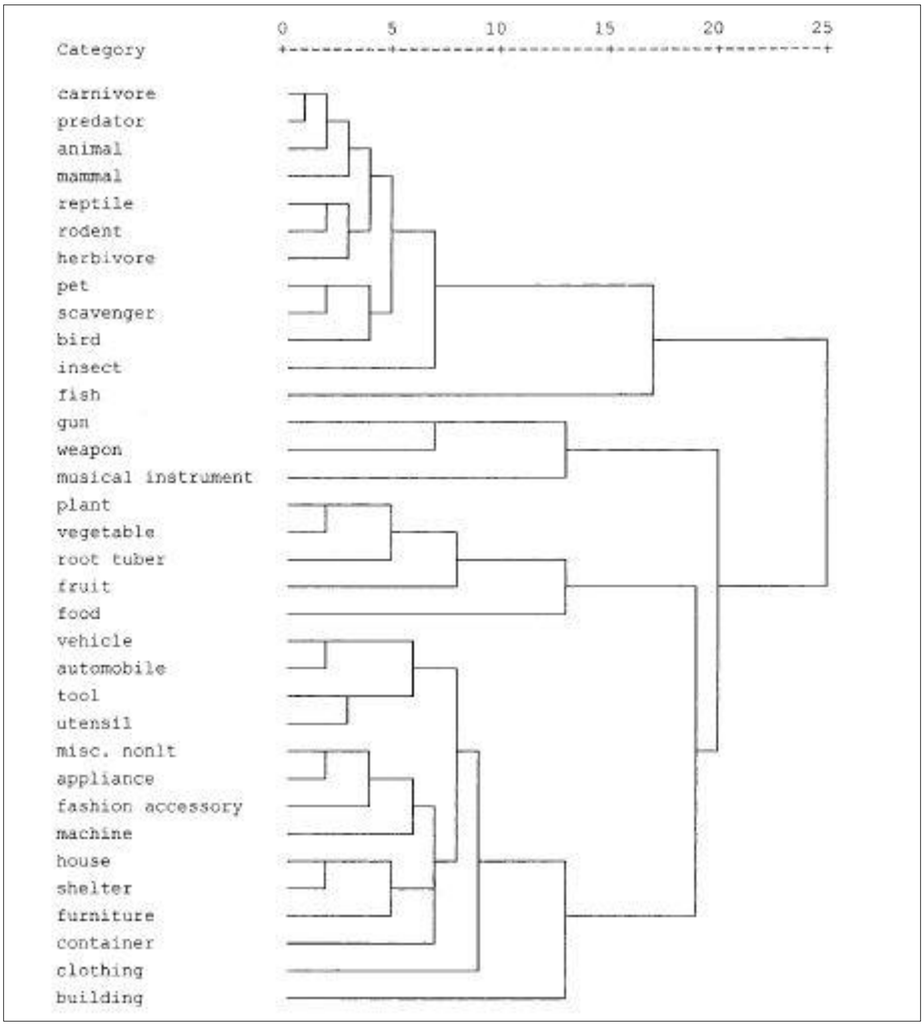

Figure 3.6 Cluster analysis of superordinate level object concepts (Cree & McRae, 2003:191)

What these cluster analyses show is that categories can potentially arise in semantic memory as a consequence of semantic similarity based on featural composition. Furthermore, these categories also cluster consistently with patterns observed in cases of category specific deficits (e.g. the often observed dissociation between living and non-living things). Such analyses have so far only been conducted in relation to object concepts as this domain is where category specific deficits have traditionally been observed. While actions may possibly fall into categories, there have been no such reports of category specific deficits affecting action concepts and verb retrieval. Therefore, if loss of featural knowledge is a characteristic of language impairment, it may be predicted that such clear-cut cluster analyses could be derived for actions on the basis of their featural composition.

3.2.6. Features as a basis for typicality effects

Overlap of semantic features between concepts within the same category has been argued to dictate a concept's typicality, or representativeness of the category as a whole. High-typicality concepts share a large number of features with other category members, whereas low-typicality concepts share fewer features. Rosch & Mervis (1975) investigated this in relation to the *family resemblance* theory of categorisation. Family resemblance theory provided a more robust explanation of typicality effects in behavioural tasks than more traditional 'classical' theories of categorisation where all category members were considered equal. Rosch & Mervis showed that highly typical category members shared more semantic features with each other than they did with atypical category members (from the same category). Atypical category members were also shown to share more features with members of other categories than did highly typical category members. They also demonstrated that atypical category members possessed a higher proportion of distinctive features than did highly typical category members. Ashcraft (1978) also found that typical category members shared a greater proportion of features with their respective category concept than did atypical category members (e.g. the highly typical *apple* shared more features with its category concept *fruit* compared to the atypical member *raisin*). It was also found that participants generated a greater number of features for typical category members compared to atypical category members. Finally, it was shown that there was greater between-participant variation in terms of the features that were generated for atypical category members compared to typical category members where there was a high level of consistency.

As yet there have been no such investigations into the featural composition along a dimension of typicality with verbs as previous research has not investigated the notion of typicality in relation to categories of verbs. However, given that Chapter two demonstrated that speakers are able to rate verbs' typicality within categories in a similar manner to how they complete such a task with nouns, then investigating whether there are featural distinctions along this dimension with verbs appears both appropriate and worthwhile.

3.2.7. Features as a basis for specificity effects

In addition to reports of category specific deficits in speakers with language impairment, there are reports of differential performance at differing levels of conceptual and lexical specificity. As with the investigation of category specific deficits, this has led researchers to investigate the possibility of a semantic featural basis for such patterns of impairment. Warrington (1975) was among the first to report such dissociations in patients with visual agnosia (i.e. impairment in recognising visually presented objects, e.g. using actual objects or pictures). These patients were poorer at identifying objects at a subordinate (i.e. specific) level of categorisation compared to superordinate (i.e. general) levels of categorisation. For example, patients were able to verify that pictures were either animals or not animals (pictures were animals, birds, and objects) but had more difficulty in verifying whether animals were English or not English, or large or small. Warrington (1975) therefore suggested that such patients had deficient semantic representations of specific featural information which were stored at subordinate levels of categorisation that would usually allow speakers to make fine-grained semantic distinctions between concepts in the same category. Such a conclusion was at the time consistent with hierarchical models of semantic memory (e.g. Collins & Quillian, 1969) which have since fallen out of favour. However, the hypothesis that there was a featural distinction between superordinate/general and subordinate/specific concepts is still valid.

Crutch & Warrington (2008) provided further evidence of dissociation between levels of specificity within the context of a naming task with patients with semantic dementia and patients with aphasia. The patients with semantic dementia exhibited better performance at naming pictures using a superordinate level term (e.g. *bird*, *insect*) compared to its specific name (e.g. *goose*, *beetle)*. However, patients with aphasia showed the reverse pattern of better naming using specific terms compared to superordinate terms. Patients with aphasia were further assessed in terms of their ability to comprehend basic level names (e.g. *bird*, *dog*) compared to subordinate level names (e.g. *robin*, *greyhound*). The patients were also more accurate when comprehending at a more specific level (i.e. subordinate) than at a comparably more general level (i.e. basic level) of categorisation. These results therefore demonstrated that the general-specific dichotomy can be doubly dissociated and also that patients with language impairment may not have a preference, or at least preserved ability, to identify objects at a basic level, as is the case with healthy speakers (e.g. Rosch, Mervis, et al , 1976). Such findings are most frequently attributed to impaired semantic representations specifically affecting semantic features, particularly with regard to distinctive features which differentiate concepts within categories.

Differential performance at different levels of specificity has also been reported in speakers with aphasia in their use of verbs. Breedin, Saffran & Schwartz (1998) found that six out of eight participants with a selective impairment in verb production were poorer at retrieving semantically 'light' (e.g. *to go*) and 'general' (e.g. *clean*) verbs than they were at retrieving semantically 'heavy' (e.g. *hurry*) and 'specific' counterpart verbs (e.g. *wipe*) in the context of a story completion task. This was taken as evidence to support the hypothesis that 'heavier' and more specific verbs have richer semantic representations in terms of number of semantic features (in an abstract feature sense; e.g. Jackendoff, 1990; Pinker, 1989) which make them more resistant to damage despite that fact that these semantically richer verbs tend to occur with lower frequency than semantically simple verbs.

These findings of differential performance with verbs at differing levels of specificity were extended by Gordon & Dell (2003) and Barde, Schwartz & Boronat (2003). In addition to assuming that more specific verbs had a greater number of semantic features, these reports also claimed that semantically simpler verbs possess a greater number of syntactic features as they tend to appear in a wider variety of syntactic contexts and can be used with a wider variety of complements (i.e. nouns phrases) than more specific verbs. This assumption was supported by Gordon & Dell (2003) who showed that, within a computer simulated model, lesioning semantic units (to simulate an anomic aphasia) led to poorer retrieval of semantically complex verbs, whereas lesioning syntactic units (to simulate Broca's aphasia) led to poorer retrieval of semantically simpler verbs. Barde et al (2006) subsequently reported that speakers with an agrammatic pattern of aphasia followed the pattern of poorer retrieval of semantically simpler verbs, whereas speakers with a non-agrammatic pattern of aphasia showed no preference for semantically simple or complex verbs. A pattern of preference for semantically simple (i.e. light verbs) has also been reported in the context of narrative story recall (e.g. Berndt, Haendiges, Mitchum & Berndt, 1997) where it was found that some speakers with aphasia overuse light verbs in comparison to more complex verbs when compared to speakers without language impairment.

Given that there is an assumption that general-specific dissociations in language impairment are attributable to differential representation in terms of semantic features between these levels of specificity, Marques (2007) conducted an analysis of speaker-generated semantic features comparing these different levels. Contrary to previous assumptions, Marques showed that object concepts at a superordinate level were not less informative (as indicated by number of unique features associated with concepts) than basic level concepts, and the features associated at concepts at the differing levels showed similar distributions in terms of distinctiveness. One difference found was that basic level concepts shared more features with other basic level concepts than superordinate concepts did with other superordinate concepts. Therefore, given that there was little to differentiate concepts at differing levels of specificity in terms of featural representation, Marques (2007) argued that dissociations in performance with naming and understanding concepts at differing levels must be attributable to differential weightings and connection strengths between features and concepts within semantic memory, rather than qualitative differences in featural composition per se.

Given that there is an apparent disparity between the assumptions of the featural composition of superordinate/general and subordinate/specific nouns and empirical investigations, it will be insightful to see if there are similar discrepancies with verbs. For example, researchers have argued for an apparent dissociation between general and specific verbs as specific verbs have a richer semantic representation in terms of number of semantic features. This has been argued on the basis of abstract semantic features but has yet to be investigated empirically using speaker-generated semantic features.

3.2.8. Features as a basis for conceptual and grammatical class distinctions

Semantic features have also been investigated in relation to the supposed dissociations observed between nouns and verbs. Such dissociations in speakers with language impairment are generally observed whereby nouns are better preserved than verbs (e.g. Caramazza & Hillis, 1991; Miceli et al, 1984; McCarthy & Warrington, 1985) although the reverse pattern has also been observed where verbs are better preserved compared to nouns (e.g. Bi et al, 2007; Zingeser & Berndt, 1988). Explanations of such dissociations have variously been attributed to grammatical encoding within lexical storage (e.g. Caramazza & Hillis, 1991) to differential weighting of semantic representations, for example, with respect to verbs generally being less imageable than nouns and thus being more susceptible in cases of language impairments, especially when semantic deficit is present (e.g. Bird et al, 2003).

Vinson & Vigliocco (2008) collected semantic feature norms for object nouns, action nouns (e.g. *the destruction*), and action verbs (e.g. *to destroy*) and used these norms to represent both nouns and verbs within the same semantic space within their Featural and Unitary Semantic Space (FUSS) hypothesis model (Vigliocco et al, 2004). In a series of reports, the FUSS model has been demonstrated to predict performance in a number of behavioural tasks and also patterns of language impairment (e.g. Vinson & Vigliocco, 2002; Vinson et al, 2003).

Within the FUSS model it was found that, based on featural representations, object nouns were spread over a wide semantic space in clusters that mirrored natural categories whereas action nouns and action verbs were spread over a narrower semantic space and did not form identifiable clusters as nouns did. Therefore, there was greater variation of semantic featural composition between categories of objects compared to action nouns and verbs where there was a lesser degree of diversity. This was further discussed by Vinson & Vigliocco (2002) who ran simulations which confirmed that the model was more likely to suffer impairment in isolated domains of knowledge to object nouns following damage to featural representations as compared to action nouns and verbs. Therefore, it was argued that speakers were unlikely to show category specific deficits with actions in a similar manner to objects as there was little to clearly differentiate categories of verbs in terms of featural composition.

Vinson & Vigliocco (2002) also reported the patterns of activations within the FUSS model following simulated lesioning of the various feature types that made up the semantic representations. Following lesioning of visual features, activation of object nouns was reduced in comparison to action nouns and action verbs. Lesioning of other perceptual features led to reduced activation in both object nouns and action nouns with less impact on action verbs. Lesioning of non-perceptual features (i.e. functional and motoric) led to decreased activation in action verbs which in comparison to object nouns whereas action nouns fell between the two other sets. These simulations were therefore argued to provide evidence for a semantic underpinning of grammatical class deficits as a result of featural damage which could differentially affect grammatical classes which had differing semantic referent (e.g. the dissociation observed between object nouns and action nouns). A similar simulation study by Bird, Howard & Franklin (2000) which selectively lesioned feature types also found that lesioning of sensory features led to selective deficit of animate objects with sparing of inanimate objects and verbs. In contrast, lesioning of functional features led to the reverse patter of deficit of inanimate objects and verbs and sparing of animate nouns. Therefore, studies such as these provide evidence that the distinction between objects and actions in conceptual-semantic levels of representation is not as clear cut as the distinction between noun and verbs in lexical level representation.

3.2.9. The current studies and research questions

The discussions so far have summarised the role of semantic features in theories of semantic memory and processing and also considered the main findings from analyses of semantic feature composition across several dimensions. The key points to take from this discussion are firstly that semantic features appear to play a significant role in how concepts are judged to be similar and dissimilar, both in terms of between categories and within categories (i.e. in terms of typicality), and secondly, that while this has been investigated and there is evidence for this first claim in relation to object concepts (i.e. nouns) there has been little research into action concepts (i.e. verbs). Therefore the following two series of analyses attempt to address the following general questions:

1) Are speakers' perceptions of semantic similarity between verbs consistent with performance in category listing?

2) Does typicality influence speakers' perceptions of semantic similarity between verbs?

3) Are speakers' perceptions of semantic similarity between verbs based on featural composition?

4) Is there a featural basis for a general-specific dichotomy in verbs?

5) Is there a featural basis for typicality within verb categories?

The chapter will continue with the following structure: Firstly, a description of the method and results of the similarity rating task. Secondly, a description of the method and results of a series of analyses regarding semantic feature composition of verbs along various dimensions: (a) within and across semantic categories; (b) between general and specific verbs; and (c) between high- and low-typicality verbs. The chapter concludes with a general discussion.

3.3. Rating the Similarity of Verbs

3.3.1. Introduction and specific questions

This section reports the use of a pairwise similarity rating task in which 32 verbs were selected across four semantic categories (i.e. *breaking*, *cooking*, *cutting*, and *making*) and participants were required to rate pairwise comparisons of verbs in terms of perceived semantic similarity. Pairwise similarity rating has been shown to be a time-efficient (in comparison to collecting semantic feature norms) and reliable method of gaining a measure of semantic similarity in terms of semantic feature overlap in object concepts (e.g. Maki, Krimsky & Muñoz, 2006), however, this doesn't appear to have been investigated within the investigation of actions. Here, half the verbs selected from each category had previously been rated as high-typicality category members and the remaining verbs were rated as low-typicality category members (see Chapter two). Data were analysed using multidimensional scaling techniques where mean similarity ratings were transformed into distances to allow verbs to be mapped within a simulated semantic space. This section aimed to address the following questions:

1) Are participants' perceptions of semantic similarity between verbs consistent with performance in category listing?

2) Do verbs drawn from semantic categories cluster within semantic space on the basis of participants' perceptions of semantic similarity?

3) Does typicality of verbs within categories influence participants' perception of similarity to other verbs within the same category?

3.3.2. Method

Participants

Similarity ratings were obtained from a total of 69 native English speaking participants. All participants were enrolled as students at Newcastle University and the sample included 38 males and 31 females (*M age* = 21.84, *SD* = 6.57 years, range = 18-59). Participants were recruited via email advertisements which contained electronic links to the web-based surveys which presented the similarity rating task. Participants indicated consent to take part by clicking a checkbox on the opening pages of the survey. In return for taking part, participants were entered into a cash/voucher prize lottery

Stimuli selection

Thirty-two verbs were selected as stimuli for this experiment (see Appendix H). These were selected from four semantic categories for which category and typicality data were previously obtained (i.e. *breaking*, *cooking*, *cutting*, and *making*). The stimuli included four high-typicality and four low-typicality verbs drawn from each category. Mean typicalities of stimuli used are presented in Table 3.1.

	Breaking		Cooking		Cutting		Making	
Typicality	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
High-	1.615	0.044	1.843	0.327	1.355	0.080	1.580	0.365
Low-	3.328	0.245	3.925	0.228	3.140	0.246	3.578	0.373

Table 3.1 Typicality-split data for similarity rating

Design and procedure

Similarity ratings were obtained via an online survey which was created and distributed via www.surveymonkey.com. Each verb was paired with all other verbs creating 496 pairwise comparisons for which similarity ratings were obtained. Verb pairs were only presented in one order here to minimise the number of ratings required. For example, ratings were only obtained for *'assembling-constructing'* and not the reverse *'constructing-assembling'*. However, individual verbs appeared equally often (as could be) as the first verb in a pair and as the second verb in a pair.

The 496 verb pairs were randomly allocated and ordered within four presentation lists each containing 124 verb-verb pairwise comparisons for which similarity ratings were obtained. No attention was given to ensure an equivalent number of each verb in each list, or an equivalent number of typicality based pairs in each list (e.g. equal numbers of high-high typicality, low-low typicality, and low-high and high-low typicality pairs). Given the relatively large number of comparisons to be made it was felt that random allocation would be sufficient to achieve a broad measure of equality across presentation lists.

Each participant only completed one presentation list and they were guided to a particular survey link in the invitation email based on the first letter of the first name (e.g. 'If your first name begins with a letter from A-M click the first link, if it begins with a letter from N-Z click the second link'). Participants were instructed that they were to rate pairs of action words on the basis of how similar they were. When thinking about similarity the specific instructions were:

> Think about all the aspects of the action. This can include: the purpose of the actions (e.g. why you would do them); any tools or objects (including body parts) you may use to carry out the actions; where you might carry out these actions; what kind of movement is involved in the actions; and so on.

Participants were also given the following example to consider:

> The actions running and walking may be fairly similar because they both involve movement done by humans primarily using the legs. Now compare running and sprinting, these may again be similar but they may be more similar because they both involve quicker movement than does walking. Now compare, running and sleeping. These actions may not be very similar; only that they are both actions that humans would do.

Participants were instructed to rate similarity of verb pairs on a scale ranging from 1 (very similar) to 9 (nothing in common).

Data analysis

The dependent variable under investigation was the mean similarity ratings for each verb-verb pair. These were transformed into ordinal ranks for the purposes of subjecting the data to multi-dimensional scaling analysis. The verb-verb pair which had the lowest mean similarity (i.e. the pair rated as being most similar) was assigned a rank of 1 and the pair with the highest mean similarity rating (i.e. the pair rated least similar) was assigned a rank of 496. Where there were ties in mean similarity, standard deviations from the mean were used to establish rank order.

There were unequal numbers of participants completing each of the four presentation lists, ranging from a minimum of 12 participants to 25 participants. However, as the dependent variable was based on mean ratings, all data was included within analysis.

3.3.3. Results

MDS analysis of verbs across categories

A semantic representation was simulated using the rank ordering of verb pairs using multidimensional scaling (MDS) techniques which were computed using SPSS version 17.0. Models were simulated containing from 2 to six dimensions as this was allowable given the number of items entering into pairwise comparisons (n = 32). Table 3.2 presents the *stress, s-stress*, and r^2 values derived from each simulation where a stress value closer to zero indicates a better fit of the model to the raw data. For data comparing the similarity of two items, it has been proposed that s-stress is a more reliable measure than the conventional stress value. However, both are reported here as there are few guidelines for the interpretation of either value, particularly s-stress values. The data fulfilled a square symmetric shape and was assumed to be matrix conditional meaning that ratings were not generalisable beyond the simulated models themselves. Models were computed according to Euclidian distances derived from the rank ordering of pairwise comparisons of similarity.

	Dimensions				
	2	3	4	5	6
stress	.355	.207	.145	.112	.096
s-stress	.278	.260	.208	.171	.144
r^2	.636	.709	.802	.845	.864

Table 3.2 Stress and r^2 values MDS simulation solutions

Whilst, a two-dimensional simulation appears to be the least reliable in terms of fitting the raw data, visual inspection is insightful and Figure 3.7 presents this two–dimensional solution. Visual inspection reveals a relatively compact cluster of verbs in the upper-right quadrant composed exclusively of *cooking* verbs. *Making* verbs appear predominantly in the lower-right quadrant and are more dispersed than *cooking* verbs but still appear to form a coherent cluster with defined boundaries. *Breaking* and *cutting* verbs appear predominantly in the upper-left quadrant and share an area of semantic space with no clear boundaries between the two categories. It is noticeable however, that three *cutting* verbs (i.e. *chopping*, *dicing*, and *grating*) which are perhaps more associated with *cooking* preparation show a tendency to be more similar to *cooking* verbs than other *cutting* and *breaking* verbs. A similar comment could also be made for *making* which appears to extending towards the cluster of *cooking* verbs.

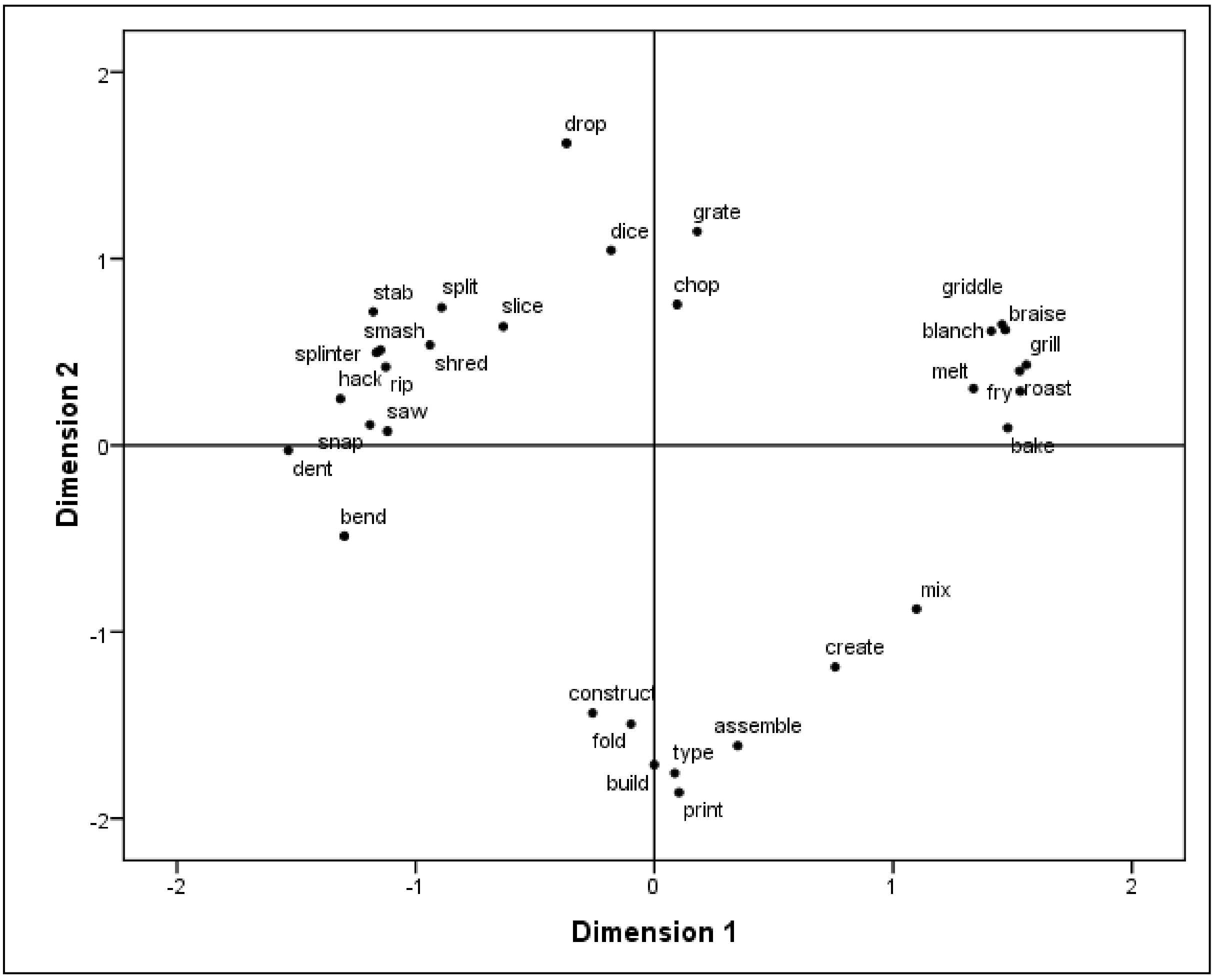

Figure 3.7 Two-dimensional solution of verb similarity

MDS analysis of typicality within categories

The similarity ratings comparing only items within each of the four categories were extracted from the complete data set (i.e. four sets of 28 similarity ratings) and subjected to further multidimensional scaling analyses. The basic procedure was the same as before whereby the verb pairs were ordered and ranked from most similar to least similar with the ordinal ranks being used as the basis for MDS analysis. The purpose of this was to investigate the distribution of high- and low-typicality items in relation to a simulated category 'core' (i.e. coordinate (0,0) in a two-dimensionally scaled solution). Figure 3.8, Figure 3.9, Figure 3.10 and Figure 3.11 present the two-dimensional simulations for each of the four categories.

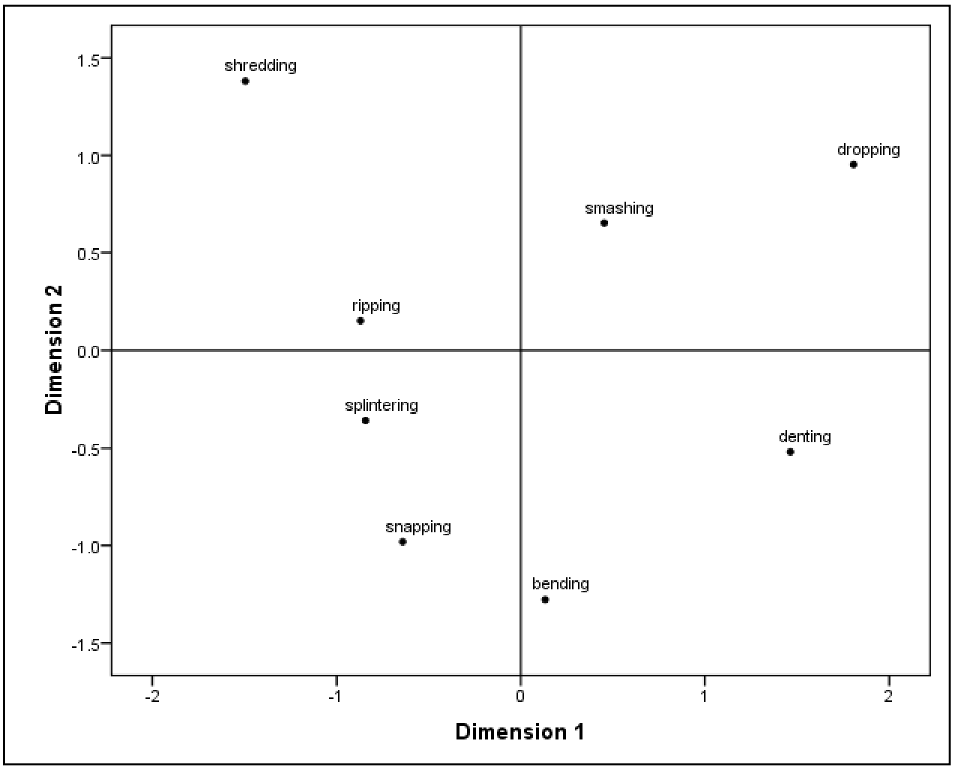

Figure 3.8 Two-dimensional solution of *breaking* verbs similarity

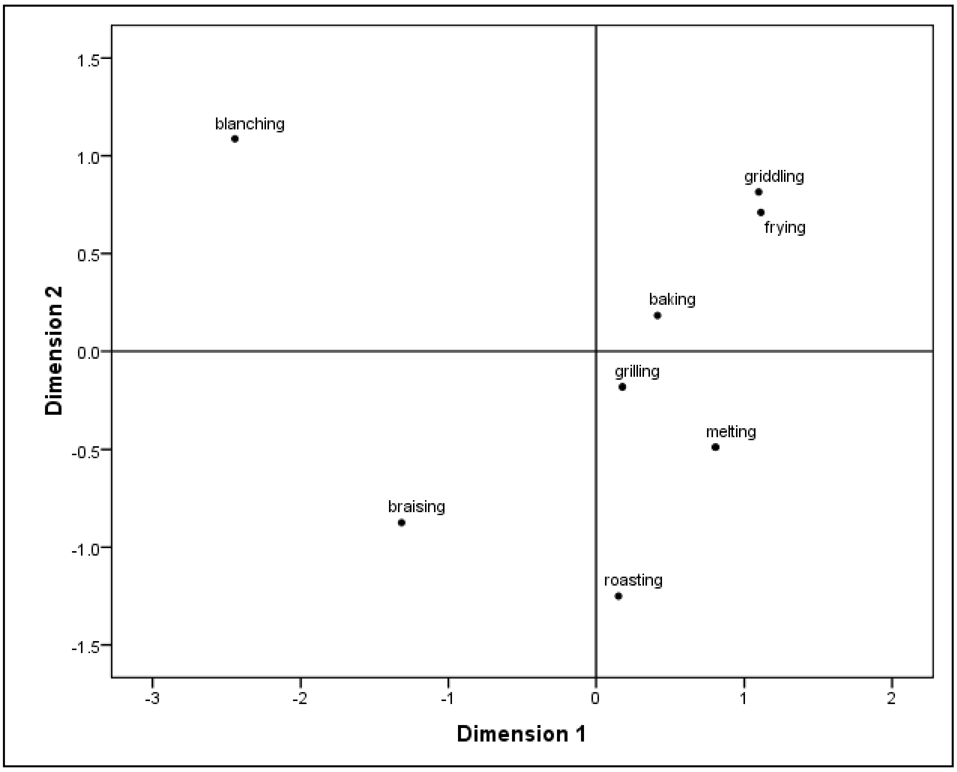

Figure 3.9 Two-dimensional solution of *cooking* verbs similarity

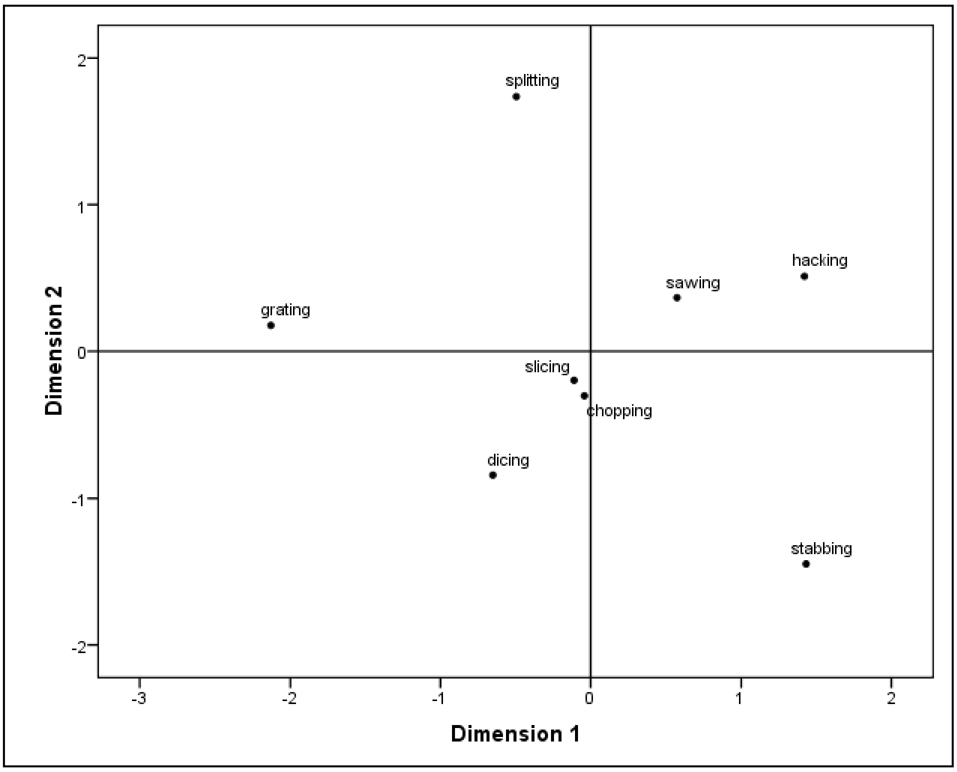

Figure 3.10 Two-dimensional solution of *cutting* verbs similarity

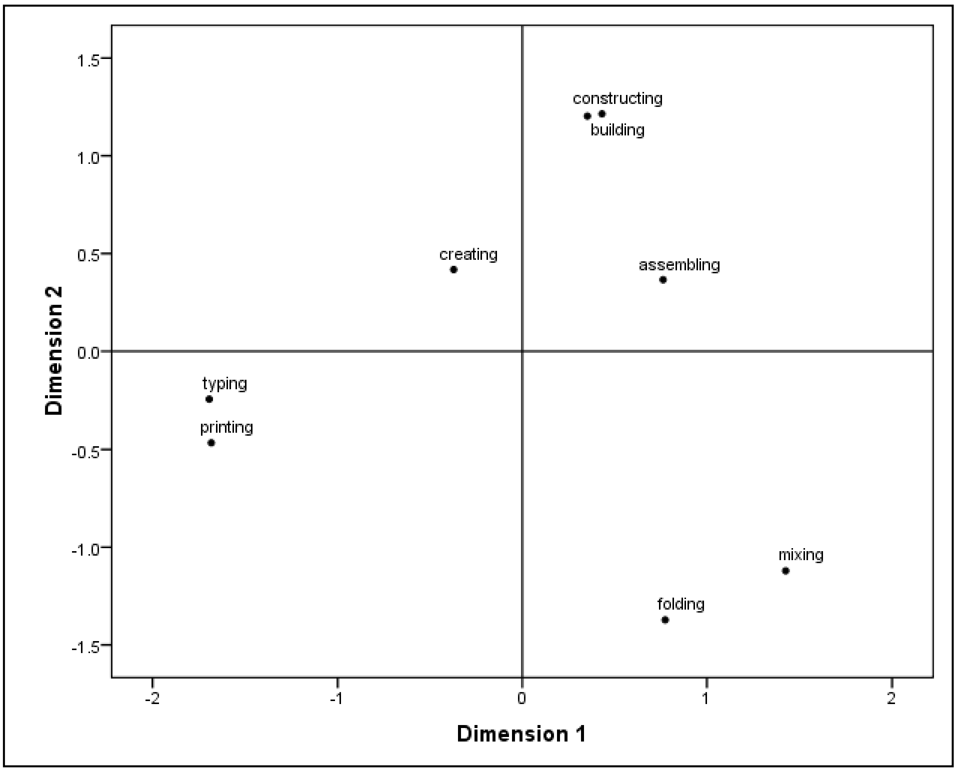

Figure 3.11 Two-dimensional solution of *making* verbs similarity

Table 3.3 presents the stress, s-stress and r^2 values for the MDS solutions for the four semantic categories.

	Breaking	Cooking	Cutting	Making
Stress	.120	.098	.132	.118
S-stress	.104	.089	.141	.075
r^2	.898	.941	.926	.919

Table 3.3 Stress and r^2 values for two-dimensional MDS solutions by category

To analyse the distribution of the category members, each member's distance from the category core was derived from their respective Euclidian coordinates. Mean distances of the high- and low-typicality members were then compared using independent sample t-tests (two-tailed) to investigate whether mean distances were significantly different. Table 3.4 presents the results of these analyses. As can be seen, two of the four categories demonstrated significant differences in distance from the category core between high- and low-typicality members.

	High-typicality		Low-typicality		t =	p =	d =
	M	SD	M	SD			
Breaking	1.223	0.569	1.447	0.471	-0.608	.566	0.432
Cooking	0.822	0.547	1.640	0.736	-1.785	.124	1.276
Cutting	0.621	0.350	1.873	0.278	-5.597	.001	3.983
Making	0.987	0.349	1.711	0.100	-3.991	.007	3.227

Table 3.4 High- and low-typicality distances from category centre (i.e. coordinate 0,0)

ANOVA analysis of typicality across categories

The mean distances from category cores on high-typicality and low-typicality category members were entered into a two-way within-subjects ANOVA with the variables category (4 levels) and typicality (2 levels). The ANOVA is represented in Figure 3.12. There was a significant effect of typicality ($F(1,3) = 35.89$, $p = .009$) but no significant effect of category ($F(3,9) = 0.29$, $p = .830$), nor a significant interaction between typicality and category ($F(3,9) = 0.96$, $p = .454$).

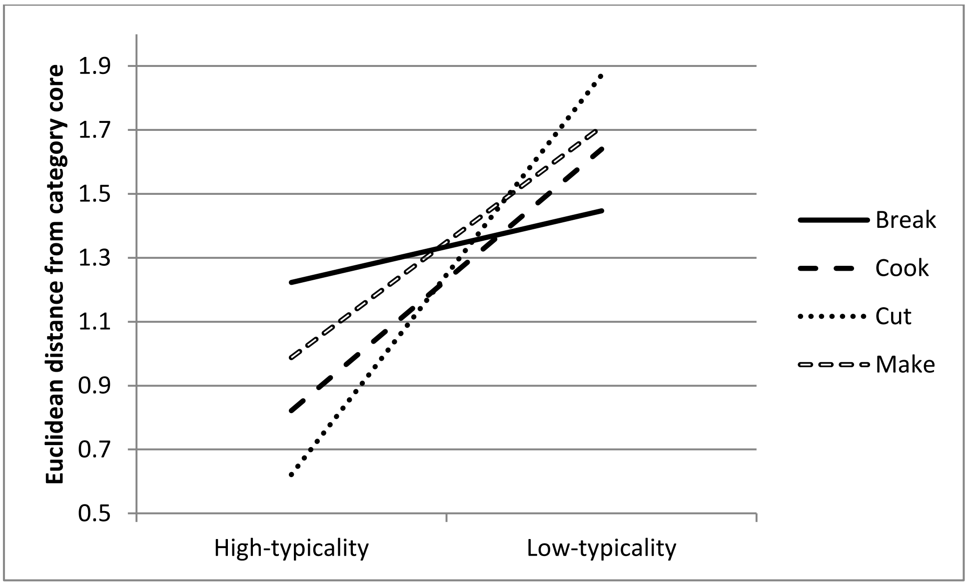

Figure 3.12 ANOVA analysis of typicality (high- vs. low) and category

3.4. Feature Composition of Verbs Across and Within Semantic Categories

3.4.1. Introduction and specific questions

This section presents the methods and analysis of the featural composition of 55 verbs which are distributed across eight semantic categories. This analysis considers a variety of different factors on which featural composition may differentiate verbs within and across the categories. The verbs and associated semantic features were extracted from the Vinson & Vigliocco (2008) set of semantic feature norms (described below) and the issues considered include:

1) Are categories of verbs differentiated by number of features possessed?

2) Are categories of verbs differentiated by types of features possessed?

3) Are categories of verbs differentiated in terms of proportions of distinctive features?

4) Is featural similarity correlated with typicality within categories?

5) Is feature possession alone sufficient to identify categories or clusters of actions/verbs?

The Vinson & Vigliocco (2008) feature norms

Vinson & Vigliocco (2008) report the collection of semantic feature norms for a total of 456 words including 169 object nouns, 71 event nouns (e.g. *the destruction*), and 216 event verbs (e.g. *to destroy*). Features were listed by undergraduate psychology students at the University of Wisconsin, United States. Each word in the total set had its semantic features listed by 20 participants who were instructed to define and describe the word using features taking as much time as was needed to do so comprehensively. Following collection of the speaker-generated features, features were classified according to feature type: visual features ('information gained through sensory input'), other perceptual features (input from sensory modalities other than vision), functional features ('features referring to the purpose of the things ... or the purpose or goal of an action'), motoric features ('how a thing is used, or how it moves'), or other features (e.g. encyclopaedic knowledge and category/taxonomic relations). The full procedure for participants to list features and the subsequent feature analysis and classification of features according to feature type is described in Vinson & Vigliocco (2008).

3.4.2. Method

Stimuli selection

A total of 55 verbs and their associated semantic features were extracted from the Vinson & Vigliocco (2008) feature norms. These 55 verbs were associated with eight of the previously investigated semantic categories (i.e. *breaking*, *cleaning*, *cooking*, *cutting*, *hitting*, *making*, *talking*, and *walking*) with a minimum of four verbs being associated with the category (not including the category verb itself). This analysis did not include category verbs themselves with the exception of *cooking*, *cutting*, and *hitting, which* were only included by virtue of being members of other categories (e.g. *cooking* was a member of the *making* category). A total of 13 of the 55 verbs were associated with two semantic categories. A complete list of verbs and their associated categories are given in Appendix I.

3.4.3. Results

Overall characteristics

The 55 different verbs were associated with a total of 532 unique semantic features which led to a total of 1635 unique verb-feature pairs. The distribution of verb-feature dominance (i.e. production frequency) is given in Figure 3.13. A large proportion of verb –feature pairs were only given by a single participant (i.e. dominance value of 0.05), and relatively few verb-feature pairs were given by 10 or more participants (i.e. >50% participants; n = 107 verb-feature pairs).

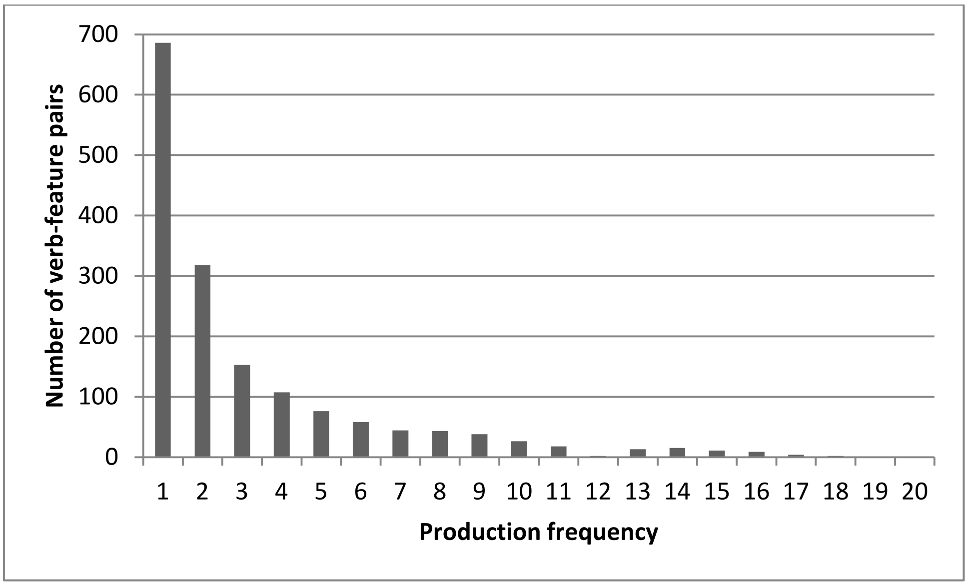

Figure 3.13 Distribution of production frequency in original 1635 verb-feature pairs

As with comparable research (e.g. Garrard et al, 2001), further analysis was based only on those verb-feature pairs given by more than a single participant in order to eliminate potential erroneous pairings. This reduced sample led to the 55 verbs being associated with 365 unique semantic features with 949 verb-feature pairs. Descriptive statistics for features associated with each category of verb are given in Table 3.5.

	Verbs (n)	Unique features	Verb feature pair	M features per verb	SD	Range
All	55	365	949	17.26	3.82	10-27
Breaking	13	131	239	18.38	2.50	15-23
Cleaning	5	57	75	15.00	3.39	12-20
Cooking	9	80	127	14.11	3.44	11-22
Cutting	4	60	82	20.50	1.91	18-22
Hitting	9	98	183	20.33	2.29	17-24
Making	12	124	216	18.00	4.57	12-27
Talking	8	65	136	17.00	3.49	10-21
Walking	8	80	129	16.13	4.42	10-24

Table 3.5 Distribution of features across verb categories

Distribution of feature types

The feature classification system of Vinson & Vigliocco (2008) was retained in order to conduct an analysis of feature types. Figure 3.14 presents the proportion of features, according to feature type, associated with the whole sample of 55 verbs and Figure 3.15 presents the feature types associated with each category of verbs. As some features were coded as more than one feature type, where this occurred, each feature type associated with a single verb was counted as a separate feature and overall proportions were calculated according to these adjusted feature totals for the category.

Chi-squared analyses were conducted to compare the raw number of features possessed for each feature type for each category to the overall number of features per feature type across all 55 verbs. These analyses showed that four categories showed significantly different feature type patterns compared to the overall pattern: *clean* ($\chi^2(4) = 10.04$, $p = .040$); *make* ($\chi^2(4) = 18.46$, $p = .001$); *talk* ($\chi^2(4) = 22.38$, $p < .001$); and *walk* ($\chi^2(4) = 21.50$, $p < .001$) with the category *hit* also approaching significance ($\chi^2(4) = 19.15$, $p = .057$).

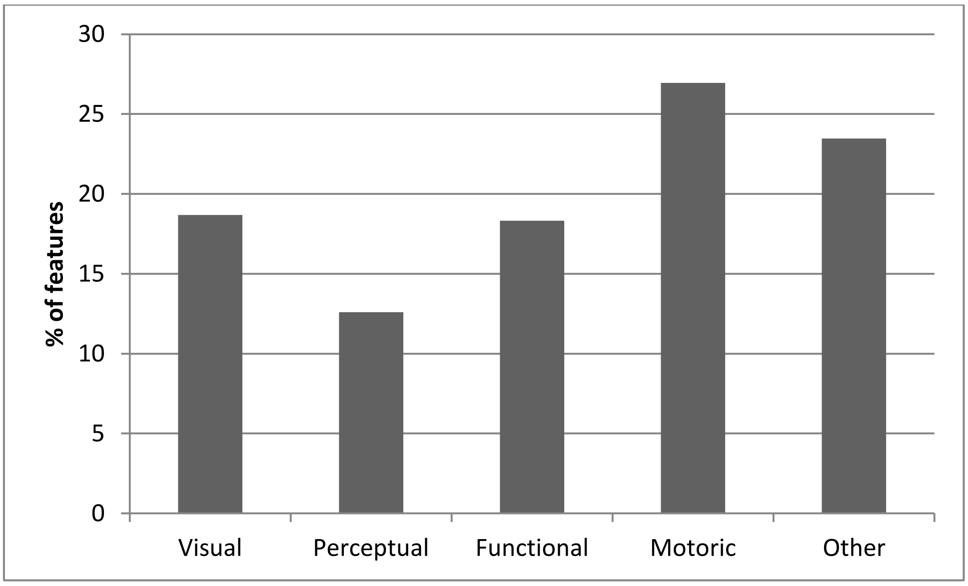

Figure 3.14 Proportion of feature types across all categories

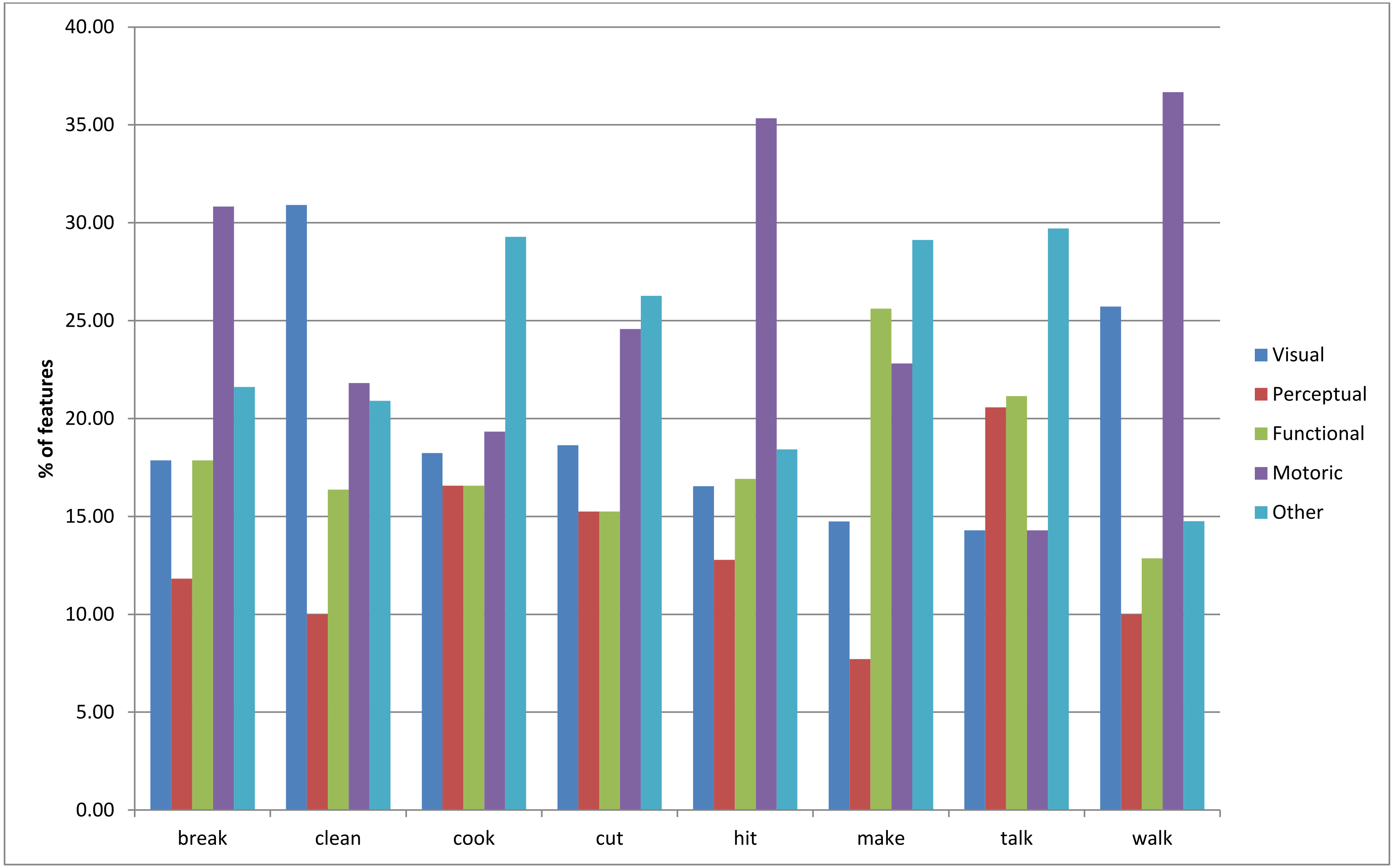

Figure 3.15 Proportion of feature types within individual categories

Feature distinctiveness

Two measures of feature distinctiveness were calculated based on whether a verb possessed a feature or not – a verb was considered to possess a feature if it was given by more than one participant in the feature norms. Otherwise, the production frequency of individual features was not represented within this analysis. One measure of distinctiveness considered the feature's distinctiveness across the set of 55 verbs and the second calculated a feature's distinctiveness within its respective semantic category. Distinctiveness was calculated as the proportion of verbs possessing the features across the respective set; therefore, a distinctiveness value of 1 indicated that the feature was shared by all verbs within the respective set (i.e. low-distinctive value), whereas a value closer to 0 indicated that it was possessed by few verbs within the set (i.e. highly distinctive). Figure 3.16 shows the distribution of feature distinctiveness across the entire set of 55 verbs.

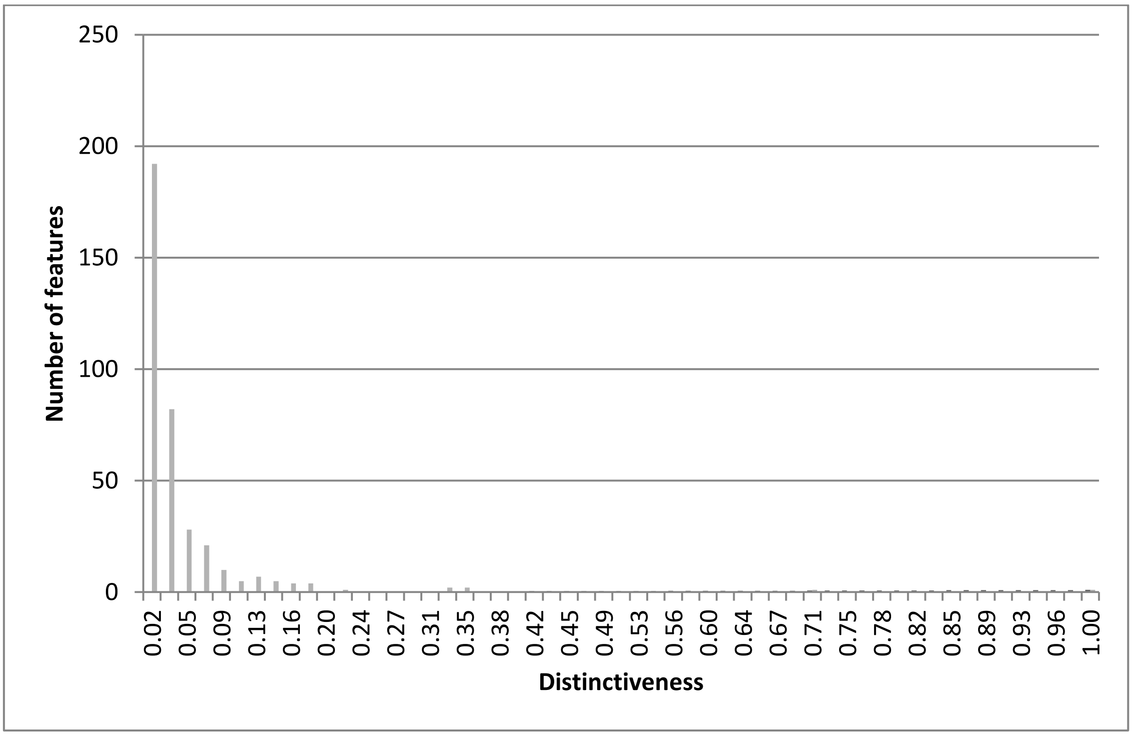

Figure 3.16 Feature distinctiveness distribution across 55 verb set

A similar pattern of feature distinctiveness was observed within each of the eight categories whereby the majority of features were possessed by only a single verb (see Appendix J).

Distribution of feature types according to distinctiveness

Further analysis was conducted to investigate whether different feature types have differential impact at differing levels of feature distinctiveness. Features were categorised into whether they were highly-distinctive (i.e. not shared by many members) within a category (i.e. held a distinctiveness value between 0 and 0.4) or low-distinctive within a category (i.e. held a distinctiveness value between 0.4 and 1). The division of high- and low- distinctiveness at 0.4 was somewhat arbitrary but was guided by the fact that the majority of features were fairly high in terms of distinctiveness and so a division at 0.5 (i.e. the hypothetical mid-point) would see few features being categorised as low-distinctiveness (i.e. shared by many members). This may have meant that the subsequent calculation of percentage proportion of feature types may have been more influenced by exceptional cases. Appendix K presents the percentage proportions of features types at the two levels of feature distinctiveness for each category. Appendix L also presents the features that were low-distinctive within each category in order to illustrate the specific features that tended to be shared between category members.

Chi-squared analyses were then conducted using the raw number of features for each feature type (i.e. as opposed to percentage proportions as chi-squared would not be sensitive to changes in differences in the total number of features) to investigate whether the feature type composition differed between the two levels of distinctiveness. Chi-squared tests showed that feature type composition differed significant between distinctiveness levels in five of the eight categories: break ($\chi^2(4) = 15.97$, $p = .003$); cook ($\chi^2(4) = 11.97$, $p = .018$); make ($\chi^2(4) = 14.81$, $p = .005$); talk ($\chi^2(4) = 12.02$, $p = .017$); and, walk ($\chi^2(4) = 17.24$, $p = .002$). Although the size of change was variable between categories, all five of the categories with significant changes were associated with a decrease in visual features and increases in perceptual and functional features as features became more distinctive (i.e. less shared). Four of the five categories were also associated with decreases in motoric features as features became more distinctive.

Feature distinctiveness and feature dominance

In order to investigate whether there is a relation between a feature's dominance (i.e. the frequency within which it is associated with a verb across participants) and its distinctiveness (i.e. the frequency within which it is associated across verbs within a particular category), Spearman's correlation coefficients were calculated between these measures. This was calculated within each of the eight categories. Table 3.6 presents the correlation coefficients and associated p-values. As previous results would suggest, most correlations were based on a high number of low-dominance and low-distinctive features and all correlations were highly significant.

Category	n features	r_s =	p =
Break	131	.697	< .001
Clean	57	.486	< .001
Cook	80	.640	< .001
Cut	60	.635	< .001
Hit	98	.763	< .001
Make	124	.706	< .001
Talk	65	.764	< .001
Walk	65	.625	< .001

Table 3.6 Correlation between feature distinctiveness and feature dominance

Family resemblance and typicality

Previous research into noun categories has suggested there is a relationship between a category member's *family resemblance* (i.e. the degree to which it is similar to all other category members, or, the degree to which it is similar to the hypothesised category prototype) and its rated typicality within the category (e.g. Garrard et al, 2001; Rosch & Mervis, 1975). Therefore, this was investigated within the verb categories currently being investigated using the typicality measures obtained in Experiment 2 (see Chapter two) and a measure of family resemblance.

Family resemblance was calculated according to the method proposed by Rosch & Mervis (1975). A mean distinctiveness rating was calculated from the sum of weighted values of attributes possessed by each category member where the weighting represented the distinctiveness proportion of the feature within the category (i.e. the proportion of category members possessing the feature). Therefore, family resemblance was represented by a value between 0 and 1 where a value of 0 showed no featural overlap with other category members and a value of 1 would show a total featural overlap with other category members.

Spearman's correlation coefficients were calculated for the relations between family resemblance and rated typicality within each category. The coefficients and associated p-values are presented in Table 3.7. Correlations in two of the eight categories (i.e. *cook* and *make*) were significant whereby a higher resemblance value indicated a lower typicality rating (i.e. more typical within the category). Given the relatively small number of observations on which correlations were based, the lack of significant correlations is not surprising. However, one further category (i.e. *cut*) showed a relationship in the same direction, while a further category (i.e. *clean*) showed a relationship in the opposite direction (i.e. as family resemblance increased, then typicality increased to become less typical). Furthermore, when correlation was conducted across all categories (n observations = 68), there was a trend towards significance whereby as family resemblance increased, typicality decreased, i.e. as verbs were more similar to other members of the category as a whole then they tended to be rated as more typical of that category (r_s = -.224, p = .067).

Category	n	r_s =	p =
Break	13	.333	.266
Clean	5	.600	.285
Cook	9	-.750	.020**
Cut	4	-.600	.400
Hit	9	-.667	.050*
Make	12	.007	.983
Talk	8	-.048	.911
Walk	8	.072	.866

Table 3.7 Correlation between family resemblance and rated typicality

Cluster analysis

As with Garrard et al (2001), a hierarchical cluster analysis was performed whereby presence and absence of features for each verb was coded as a binary variable. All features given by two or more participants were considered to be present for a particular verb therefore cluster analysis was based on the 949 verb-feature pairs following exclusion of low-dominance features (i.e. those given by only one participant). The resulting dendrogram is presented in Figure 3.17.

It is noticeable within the cluster analysis that there are few verbs which appear closely related, as indicated by most verbs joining clusters towards the right of the figure rather than the left. There is also little indication of discrete clusters forming on the basis of the categories that the stimuli verbs were selected from. The exceptions to this would be verbs within the categories *talk* and *walk* although with *talk* the verb *scream* does not cluster with the other verbs, and the same is true with the verb *step* for the category *walk* which also has the verb *mix* appearing within the main cluster of walk verbs. Verbs within the *break* and *make* categories are distributed widely across the hierarchy. *Clean* verbs are within the lower third although they do not appear as a discrete cluster. There is a concentration of *cook* verbs in the lower third, although (perhaps understandably) *chop* and *mix* do not cluster with the others. *Cut* and *hit* verbs are mostly in the upper third of the hierarchy but again do not form discrete clusters.

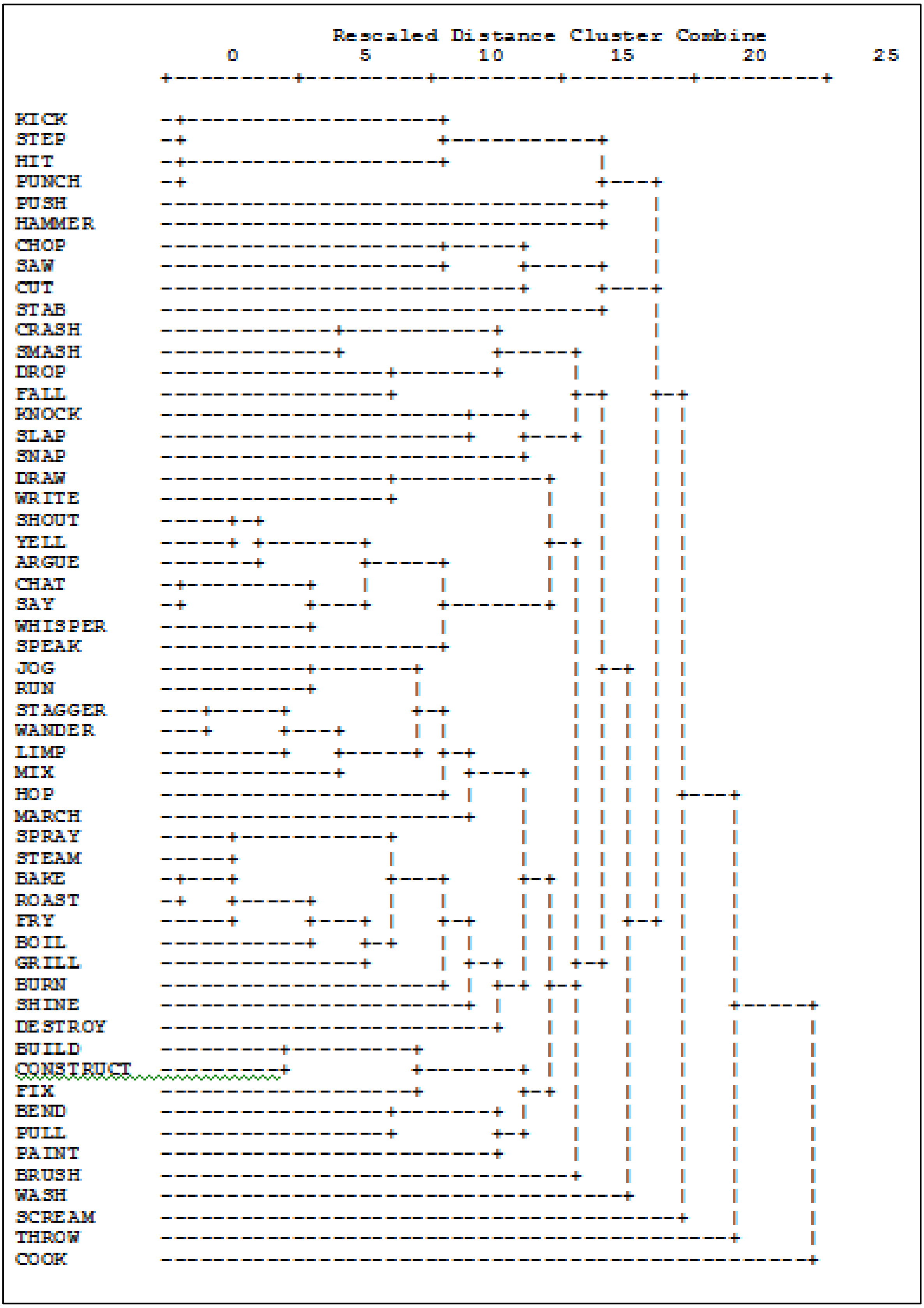

Figure 3.17 Dendrogram of hierarchical cluster analysis of 949 verb-feature pairs

3.5. Feature Composition of General and Specific Verbs

3.5.1. Introduction and specific questions

This section reports analysis of a subsection of the original 55 verbs in order to investigate the Featural composition of general (i.e. superordinate) verbs in comparison to more specific (i.e. subordinate) verbs. This smaller subsection of verbs was used to allow an equal number of verbs to be included across the categories being investigated. Based on previous research it might be expected that specific verbs consist of 'richer' semantic representations, therefore this was investigated in terms of numbers of features, feature types, and feature distinctiveness.

3.5.2. Method

Stimuli selection

Seven of the previously investigated superordinate category (i.e. general) verbs and their associated features were extracted from the Vinson & Vigliocco (2008) feature norms: *break*, *cook*, *cut*, *hit*, *make*, *talk*, and *walk*. In addition, four category member (i.e. specific) verbs were selected within each category. Two of these category members had previously been rated as highly typical category members and two had been rated as low typicality category members (see Chapter two). An Independent samples t-test confirmed that across all categories, there was a statistically significant difference in mean typicality ratings between verbs selected as high- and low- typicality category members ($t(26) = 11.18$, $p < .001$). A complete list of stimuli is presented in Appendix M.

3.5.3. Results

Characteristics

The total 35 verb set was associated with 433 unique semantic features leading to 1055 verb-feature pairings. A total of 429 of these verb-feature pairings were given by only a single participant and were subsequently excluded. Therefore all further analysis was based on the remaining set which was associated with 278 different features and 626 verb-feature pairs.

The seven general verbs were associated with 104 features and 130 verb-feature pairs whereas the 28 specific verbs were associated with 243 features and 496 verb-feature pairs. In total, 33.01% of the semantic features associated with the 35 verb set were shared by at least one general and at least one specific verb. Figure 3.18 presents the percentage of verb-feature pairings at each production frequency for both general and specific verbs (including those with a production frequency of 1 which were subsequently excluded).

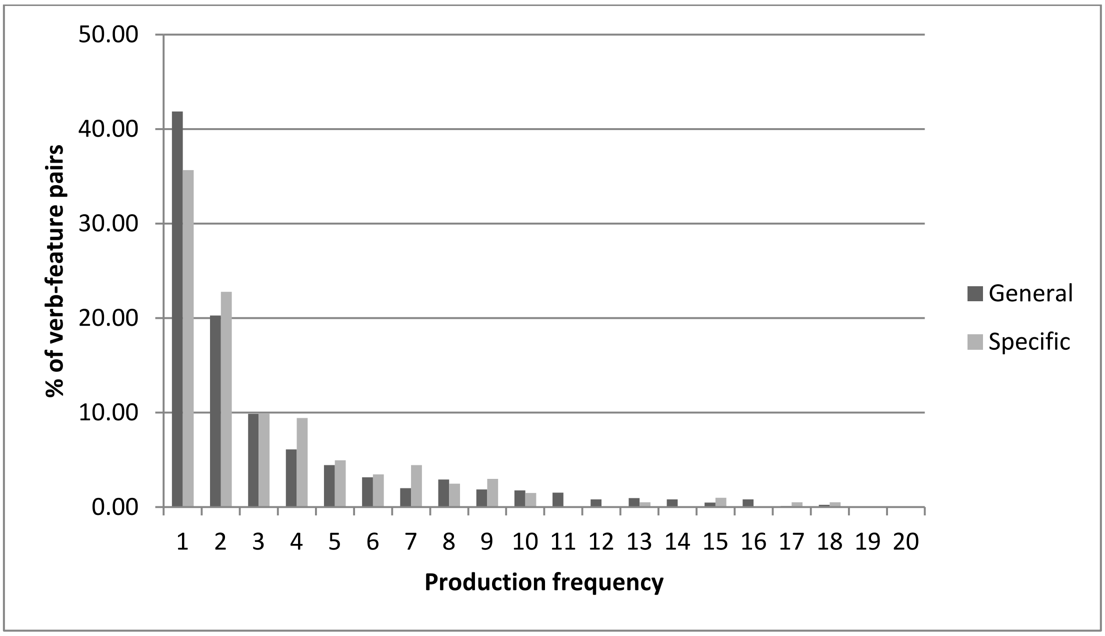

Figure 3.18 Percentage of verb-feature pairings for production frequencies (general vs. specific verbs)

A paired samples t-test was conducted between the number of features associated with each general verb and the mean number of features for its four related specific verbs. There was no significant difference in the number of features associated with general verbs (M = 18.6, SD = 4.5) and specific verbs (M = 17.7, SD = 2.5) ($t(6)$ = .405, p = .699, two-tailed).

Feature types

The percentage proportion of feature type composition of general and specific verbs is presented in Figure 3.19. A chi-squared test on the raw number of features showed there was no significant overall difference in the feature type composition between general and specific verbs (χ^2 (4) = 5.49, p = .241).

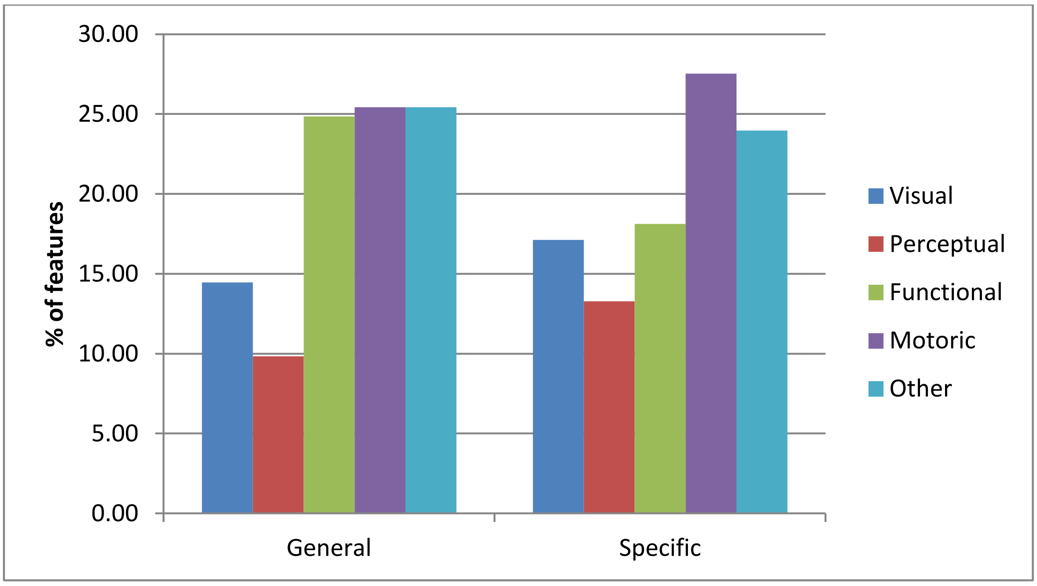

Figure 3.19 Percentage proportion of feature types for general and specific verbs

Feature distinctiveness

Each features' distinctiveness was calculated according to the number of verbs it was possessed by across all 35 verbs. Therefore, a highly distinctive feature, possessed by only one verb out of the total 35 would hold a value of 0.029, whereas a feature shared by all verbs would have a value of 1. Figure 3.20 presents the proportion of features at each level of distinctiveness for general and specific verbs.

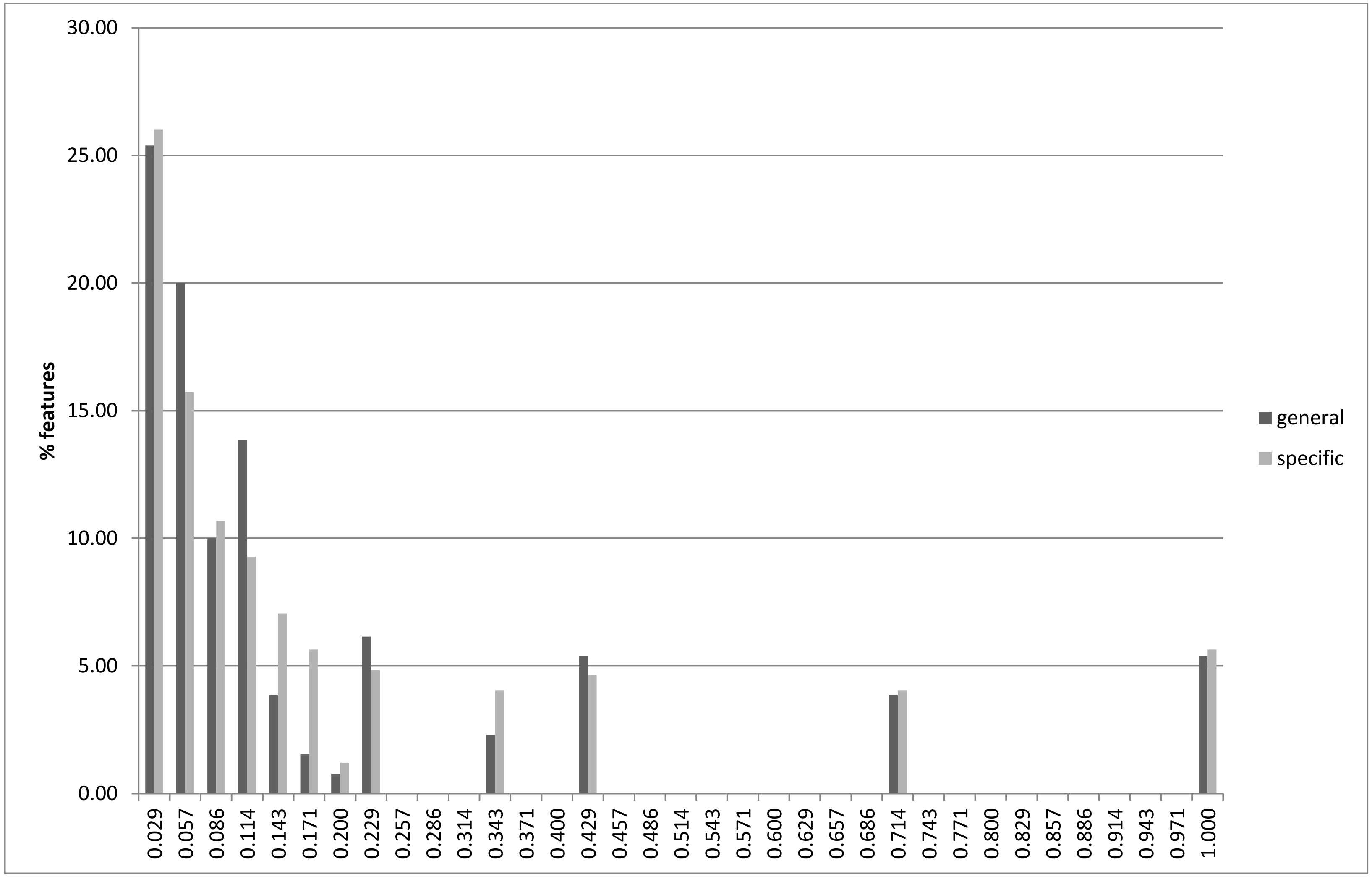

Figure 3.20 Feature distinctiveness and percentage proportions for general and specific verbs

3.6. Feature Composition of High- and Low-Typicality Verbs

3.6.1. Introduction and specific questions

In the following analysis, specific verbs from the previous analysis were further subdivided into the groups of high- and low-typicality verbs. This would allow for comparison between the featural compositions of high- versus low-typicality verbs in addition to comparing how these sets' featural composition compared with that of the general verbs. The analysis will only report featural characteristics following the exclusion of verb-feature pairings which had only been given by one participant (i.e. analysis based on the remaining 278 different features and 626 verb-feature pairs).

3.6.2. Method

Stimuli selection

See section 3.5.2 (and Appendix M)

3.6.3. Results

Characteristics

The characteristics for general verbs were the same as per the previous general vs. specific analysis. The 14 high-typicality verbs were associated with 146 different features and 249 verb-feature pairs. The 14 low-typicality verbs were associated with 163 different features and 247 verb-feature pairs. In total, 27.16% of the semantic features associated with the 35 verb set were shared by at least one high-typicality and at least one low-typicality verb Figure 3.21 presents the percentage of verb-feature pairings at each production frequency for general, high- and low-typicality verbs (including those with a production frequency of 1 which were subsequently excluded

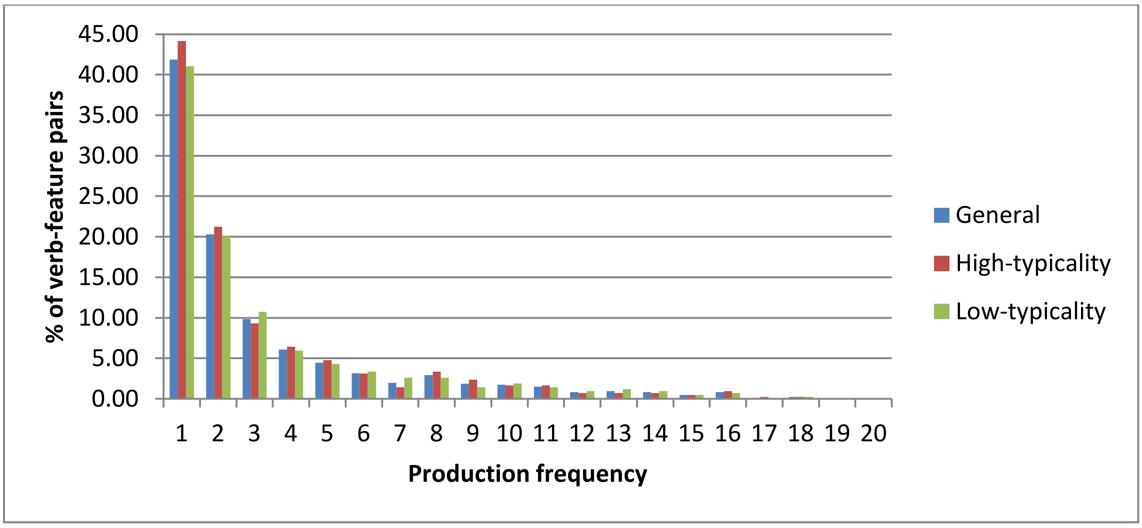

Figure 3.21 Percentage of verb-feature pairings for production frequencies (general vs. high- vs. low-typicality verbs)

Paired samples t-tests were conducted between the number of features possessed by general verbs and the mean number of features between their respective high-typicality and low-typicality specific verbs. There were no significant differences in the mean number of features in any comparison: general vs. high-typicality verbs ($t(6) = 0.358$, $p = .733$); general vs. low-typicality verbs ($t(6) = 0.451$, $p = .668$); and high- vs. low-typicality verbs ($t(6) = 0.295$, $p = .778$).

Feature types

The percentage proportion of feature type composition of general, high- and low-typicality verbs is presented in Figure 3.22.

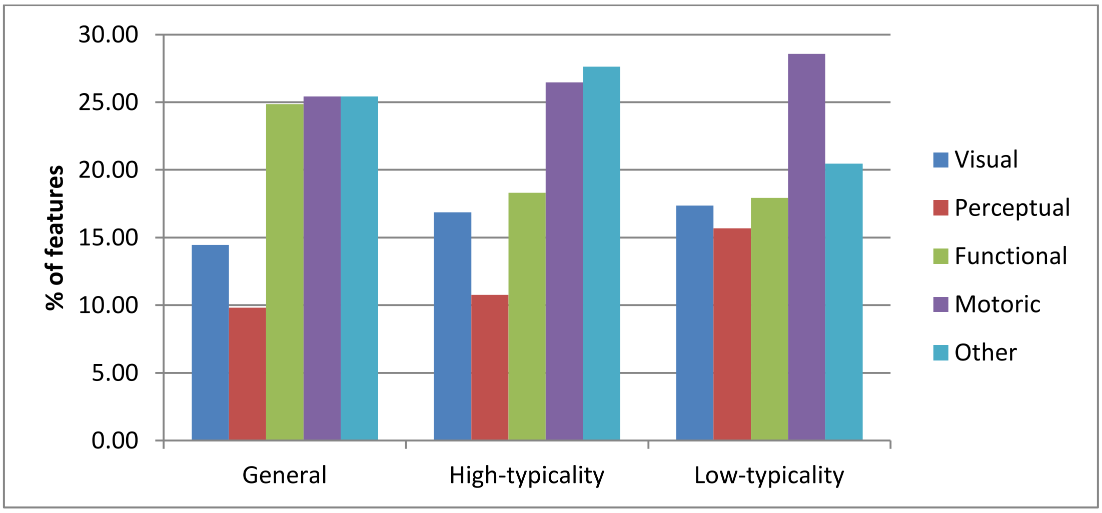

Figure 3.22 Percentage proportion of feature types for general, high- and low-typicality verbs

A chi-squared test on the raw number of features showed there was no significant overall difference in the feature type composition between general, high- and low-typicality verbs (χ^2 (8) = 12.91, p = .115). There were also no differences in the distribution of feature types between: general verbs and high-typicality verbs (χ^2 (4) = 3.16, p = .531); general verbs and low-typicality verbs (χ^2 (4) = 18.01, p = .091); and high-typicality verbs and low-typicality verbs (χ^2 (4) = 7.29, p = .121).

Feature distinctiveness

A feature's distinctiveness was calculated in the same way as the previous analysis (i.e. across all 35 verbs). Figure 3.23 presents the proportion of features at each level of distinctiveness for general, high-, and low-typicality verbs.

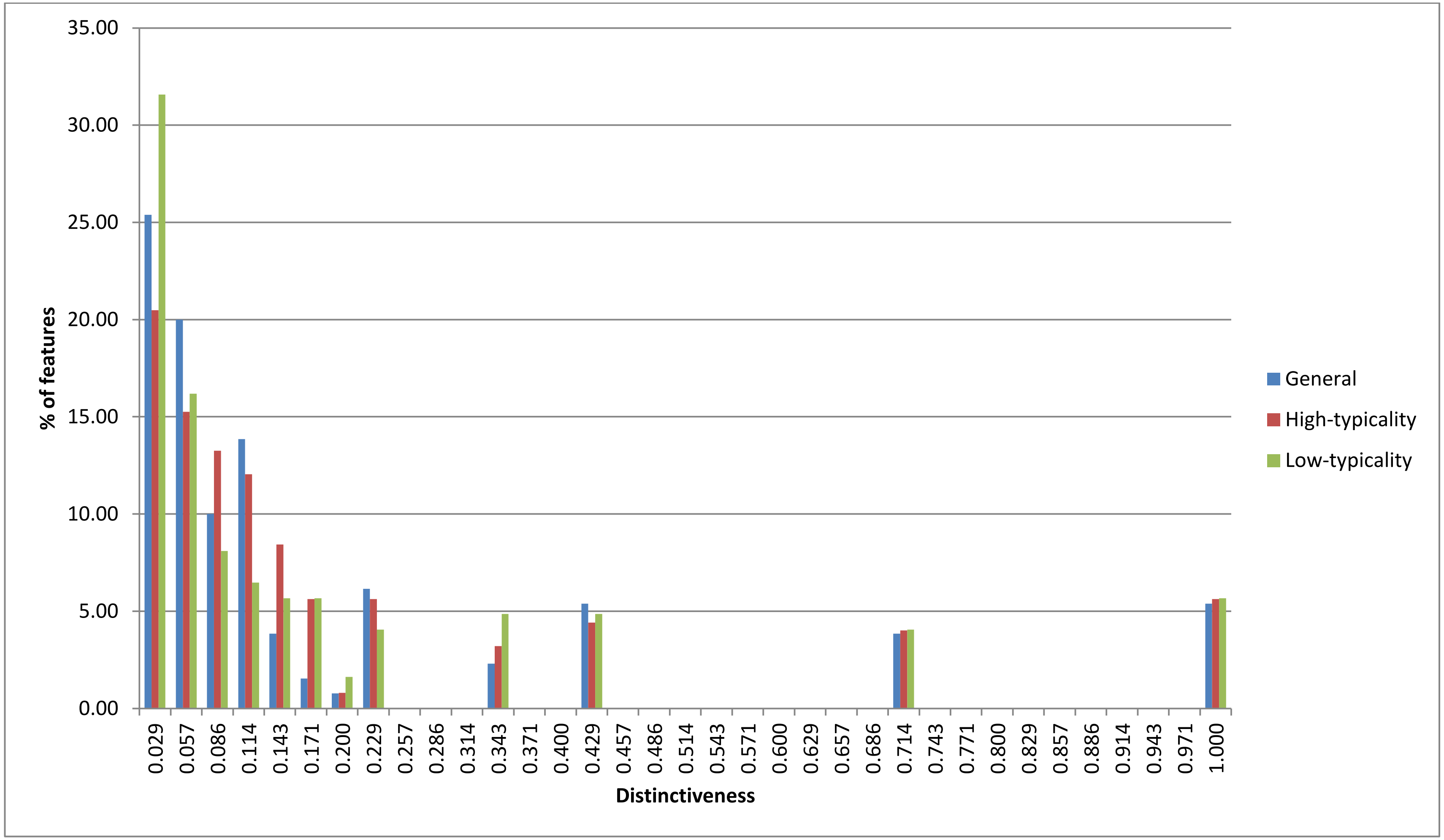

Figure 3.23 Feature distinctiveness and percentage proportions for general, high-, and low-typicality verbs

3.7. Discussion

3.7.1. Summary of main findings

This chapter aimed to investigate the semantic similarity of verbs in terms of speaker perception and decomposition into semantic features. A pairwise similarity rating task showed that: (1) participants' ratings of verb similarity is consistent with performance in category listing tasks in that some categories of verbs appeared clustered within a discrete space (i.e. cooking and making) whereas other categories blended in semantic space (i.e. breaking and cutting); and (2) participants ratings of similarity within categories also reflected rated typicality with a tendency for more highly typical verbs to be positioned at the centre of the category with less typical verbs positioned more distant from the centre.

An analysis of the feature composition of verbs across and within semantic categories showed that: (1) the majority of features that were listed for individual verbs were given by very few participants (i.e. low production frequencies); (2) the mean number of features associated with verbs varied between categories; (3) motoric features were the most prevalent feature type overall although there was variation between categories and at differing levels of feature distinctiveness; and (4) the majority of features were highly distinctive and only associated to one or very few different verbs. In addition, there was little evidence to suggest featural differentiation between general and specific verbs and also between verbs of high- and low-typicality.

3.7.2. Discussion of main findings

Rated similarity is consistent with category listing

The finding that rated similarity showed parallels with category listing suggests that speakers are able to perceive discrete categories of verbs but that certain categories also show a degree of overlap most likely attributable to the polysemy of verbs associated with the categories. The current analysis therefore contradicts slightly with the analysis using self-organising maps on the basis of featural properties (i.e. Vinson & Vigliocco, 2002) where no discrete verb categories emerged. This finding may be dependent on the number of verbs which are included within the analysis or alternatively it may be a consequence of the differential methods used to represent verbs within a simulated semantic space.

Rated similarity is consistent with rated typicality

While there was an overall effect of rated typicality, whereby high-typicality category members were located more central to the category core (based on similarity ratings) than low-typicality category members, this was not consistent across all individual categories. This finding is therefore similar to that of Romney et al (1996) where typicality was a better indicator of rated semantic similarity for some categories than it was others (e.g. where members of the category *vehicles* showed a correlation between feature composition and typicality whereas members of the category *vegetables* did not). In order to investigate this claim in greater detail and to allow greater validity of findings, it would be useful to include a greater number of category members that would allow correlational analyses with rated typicality and distance from category core. It may be that the current analysis whereby a two-way typicality distinction was made (i.e. high-, low-typicality), increases the chances of finding a typicality effect and that a correlation analysis based on more observations would be more valid and robust if significant relationships are found. Despite this potential limitation, this finding does further highlight the potential validity of typicality as an organisational principle within verb semantic categories.

Few features are strongly associated with verbs

The fact the there were few features that were strongly associated with individual verbs was indicated by the fact that there were very few verb-feature pairs that had high production frequencies. A large proportion (42%) of verb-feature pairs were generated by only a single participant out of the 20 total participants completing the Vinson & Vigliocco (2008) feature listing task and 83 percent were given by five or fewer participants. These values are generally greater than those found by Garrard et al (2001) who found that of 2969 noun-feature pairs across their 62 word set approximately 270 (9%) were only generated by a single participant. Approximately 38 percent were then given by one or two participants out of their total of 20 participants who listed semantic features and approximately 68 percent were given by five or fewer participants. Although there is in general relatively little consistency in feature listing regardless of whether features are listed for verbs or for nouns, these results therefore suggest that listing semantic features is a more difficult task for verbs than it is for nouns. Or at least, feature listing for verbs shows more variation between participants because of the nature of their semantic representations (e.g. possibility of looser conceptual fit and greater polysemy making the interpretation of the written verb highly variable between participants).

Verbs are represented by motoric features

The finding that verbs in general were represented in large part by motoric features, and also the fact that feature type was variable between semantic categories are both consistent with previous research (i.e. Vinson et al, 2003) which has performed feature type analysis on the complete Vinson & Vigliocco (2008) feature norms. One difference between the current analysis and previous research is that there was a greater proportional representation of visual and other perceptual features in the current analysis. These differences are likely due to the differential methods of analysing features rather than differences in the categories employed (although these did differ). The current analysis was conducted on the basis of simple presence or absence of features whereas previous research (i.e. Vinson et al, 2003) has analysed features when adjusted for feature weight (i.e. the production frequency of features) which gives an additional consideration of how strongly each feature is associated with a verb. However, given that the majority of features associated with the 55 verbs in the total sample had low production frequencies and were highly distinctive, the implications of analysing feature type using these different methods deserves further attention. The variability between semantic categories observed here and also by Vinson et al (2003) also contrasts with the relative consistency found in their investigated object categories where visual features and other features tended to be most prominent across the majority of categories and especially within categories of living things (e.g. *animals*, *fruit*/*vegetables*).

Most features are highly distinctive

Given that the majority of verb-features pairs were produced with low frequency, it is perhaps unsurprising that the majority of features across the 55 verb set and within semantic categories were also highly distinctive (i.e. shared by few if any other verbs within the same category). Only the feature <*action*> was shared across all verbs. Other features that were often shared by verbs within categories were features which were consistent with the hypothesised category (e.g. <*break*> which was shared by 54 percent of verbs in the *breaking* category; <*cut*> which was shared by 75 percent of verbs in the *cutting* category; and so on). It is worth noting that in previous studies involving feature listing for nouns, such ‘category’ features are have generally been excluded from the analysis as they are assumed not to a feature as such but expressing the categorical relation and are therefore redundant within the analysis. As verbs show a greater degree of polysemy than nouns and categories are not assumed to exist in the same structured manner as with nouns, these features were retained in the current analysis. Apart from these features, other shared features (i.e. low distinctive features) tended to highlight common actions or manners of performing such actions (see Appendix L). It is also interesting to note that very few of these appear to overlap with what might be termed semantic features from an abstract decompositional perspective (e.g. Jackendoff, 1983; Pinker, 1989). For example, while <*intentional*> was present in some categories, it was still not possessed by all category members (e.g. 46% of *breaking* members, and 50% of *cutting* members). This presents an interesting discussion in terms of retrievability of such features as this is unlikely to reflect that only 50 percent of cutting actions are intentional with the remaining 50 percent being unintentional (although some would naturally occur unintentionally). Therefore, feature listing studies which identify levels of distinctiveness of features may be further combined with feature verification tasks where it may be clearer to dissociate feature associations (i.e. whether a verb is associated with a feature) compared to ease of retrieving features (e.g. through patterns of yes/no responses and response times).

A finding which may be worth further investigation is the finding that prominence of feature types may vary according to the level of distinctiveness within verb categories. For example the finding that visual and motoric features tend to be less prominent as features become more distinctive (i.e. less shared between members of the same category) and also that functional features become more prominent as features become more distinctive. This may therefore parallel findings of Garrard et al (2001) who found feature type distinctions at differing levels of distinctiveness in the domains of living and nonliving things (e.g. more encyclopaedic features in both domains as features became more distinctive but an increase in sensory features as feature became more distinctive only in the domain of nonliving things). The difference here compared to Garrard et al's (2001) is obviously the size of domain under investigation. Whereas living and nonliving each cover what may be considered a fairly broad range of concepts, the categories of actions/verbs investigated here are likely to cover a more restricted range of concepts (even allowing for issues of polysemy of verbs, i.e. that one verb can be used to express a greater range of meanings than can a single noun). Therefore, further investigation may also consider the size of 'category' of actions that is under investigation.

Possession of features on its own does not dictate discrete categories

The results of the hierarchical cluster analysis showed a lack of discreteness in comparison to cluster analyses previously performed with nouns (e.g. Cree & McRae, 2003; Garrard et al, 2001; McRae & Cree, 2002). These previous analyses yielded clusters that were clear with regard to outlining category structures that were consistent from an intuitive sense and also in terms of mapping onto patterns of dissociation observed in cases of language impairment (e.g. living/non-living things). While the cluster analysis for verbs does show some clustering based on the hypothesised semantic categories (e.g. the tendency for *talking* verbs to cluster together, the majority of *walking* verbs, and the majority of *cooking* verbs), the overall patterns appear relatively difficult to interpret with confidence. There is no evidence of broad subdivisions between general clusters of verbs and the general pattern of that clustering occurs at fairly distant levels of semantic similarity (i.e. with clusters branching towards the left extreme of the resulting dendrogram).

The lack of discreteness in the dendrogram might be expected given the fact that verbs tend to be more polysemous than nouns and it might be expected that discreteness would only emerge in such an analysis if verbs had been selected within fewer and more opposing categories. For example, the clusters that appear most obviously include those to do with *talking* and *walking* which conceptually and thematically only require a single participant, although *talking* is usually done in communication exchanges with two or more people. It may therefore be that verbs which conceptually and thematically require two or more participants are more difficult to tease apart on the basis of featural composition as things that may be participants and which may be encoded as features of verbs could be *broken* or *hit* or *cut* or *made*. Even taking into account polysemy of verbs, there are still some patterns which may be unexpected based on category listing data and intuition, for example, with *cleaning* verbs such as *spray* and *shine* apparently clustering with the majority of other *cooking* verbs while *cook* itself appears in complete isolation to all other verbs not clustering directly with any other *cooking* verbs or any other verbs within the *making* category (i.e. the category in which it was a member).

No evidence of a featural distinction between general and specific verbs

The finding that there was little to differentiate general and specific verbs did not corroborate the notion of richer semantic representations for specific verbs compared to general verbs where richness equates to the number of features possessed (e.g. Barde et al, 2006; Breedin et al, 1998). However, this finding is consistent with those of Marques (2007) who found little difference in terms of number of features between nouns at differing levels of specificity. These results therefore present a potential further dissociation between abstract featural knowledge (i.e. Jackendoff, 1980; Pinker, 1989) and featural knowledge that is consciously available in feature listing experiments. Semantic features obtained in listing experiments are not assumed to be an accurate and unbiased representation of conceptual-semantic knowledge (see McRae et al, 2005, for discussion) which may potential limit their usefulness in inferring principles of semantic organisation. However, abstract features suffer from the reverse limitation in that their presence and validity cannot be ascertained as by their nature they are abstract and their presence is most regularly inferred from how verbs behave within sentences (e.g. by considering their argument structure and thematic role assignment conventions).

An alternative explanation may be that the level of analysis employed in the current study was not sensitive enough to allow differentiation between the featural distinction between general and specific verbs. One possibility may be that specific verbs may have a greater reliance on correlated features. Specific verbs are assumed to be associated with a more limited range of contexts and if participants tend to list related objects or other thematic participants (e.g. instruments, locations, and so on), there may be greater consistency between participants and such features may be more dominant (i.e. higher production frequencies) and show stronger associations with co-occurring features. Even though the overall strength of correlated features has been shown to be lower in action verbs in comparison to object nouns (e.g. Vinson et al, 2003) it may be that particular semantic subsets of verbs show stronger correlations.

No evidence of a featural distinction between high- and low-typicality verbs

The evidence for a lack of featural distinction between high- and low-typicality category members (and also from their respective superordinate, i.e. general category verbs) came via a lack of a consistent relationship been computed family resemblance and typicality and also through the numbers and distinctiveness of features between high- and low-typicality category members. These findings are contrary to what might be expected based on previous object category research (e.g. Rosch & Mervis, 1975) but are consistent with contemporary research (e.g. Marques, 2007). It may also be that the level of analysis was not deep or sensitive enough to highlight any crucial featural distinction between verbs but it is perhaps more likely that other factors influence rated typicality and that computations of featural properties alone cannot account adequately in all categories (e.g. as consistent with Romney et al's,1996, analysis of object categories).

3.7.3. Limitations and further research

The similarity rating procedure and subsequent analysis is limited in that the seemingly neat separation of categories may change if more verbs were entered into the analysis. This does however, pose problems in that the number of pairwise comparisons to be made would quickly become unmanageable and a vastly increased participant sample size would be required. Comparing just 32 verbs yields a total of 496 pairwise comparisons where it was deemed suitable in the current experiment to break this into 4 blocks. This was not done because of excessive time-demands that the task placed on participants (as this was not expected to exceed 25 minutes) but due to the fact that the majority of comparisons were expected to receive low similarity ratings as most pairwise comparisons were expected to show little similarity based on the category listing data (see Chapter two). There may have been a subsequent risk that participant's levels of attention to the task may have decreased if they found themselves constantly giving low similarity values.

Further research may look to apply similar methods but may focus on within category distributions of similarity, for example by focusing on and selecting more verbs from a single category, or those categories which may be expected to overlap. This may allow further in-depth investigation and identification of the actual dimensions along which verbs are distributed within category and may also be used to further explore the relation between rated similarity and rated typicality.

The analysis of featural composition was limited in that analysis focused on an existing collection of feature norms which did not contain a number of verbs that would have allowed a more confident analysis, especially within-categories, based on data gathered in the previous category listing task (see Chapter two). It would have been preferable to collect feature norms as part of the current investigations to allow for deliberate selection of verbs to more effectively investigate the patterns of categorisation and typicality demonstrated in Chapter two. This indeed was attempted as part of the current investigation but proved impractical due to the number of participants required and poor response rate from potential participants whom had been issued with feature listing response workbooks. Therefore, further research could focus on the collection and subsequent analysis of verbs which were selected within various categories and to represent the full range of typicality values within categories. Such methods would also allow a more thorough interpretation of the relation between feature composition and similarity ratings and the dimensions along which verbs are perceived to differ. In the present investigations there is limited overlap in the verbs employed in the similarity rating task and in the investigation of features which would make any conclusions regarding dimensions of similarity entirely speculative.

The current investigations also neglected the area of feature correlations. This was again partly due to their being limited scope for interpretation with the small numbers of verbs within categories but also because of the previous findings that verb correlations are not as prominent in featural composition of verbs as they are with nouns. In fact, it may be hypothesised that because there are so few features which are shared between large numbers of verbs, correlations may actually play a more significant role with verbs even if the number of correlations may be fewer. It may also be expected that feature correlations are especially prominent in the feature type 'other' which appeared to include a number of objects that may be associated with the actions, with feature correlations potentially reflecting common thematic partners. By extension to this hypothesis, it may be expected that correlations may be more abundant in specific verbs as these are assumed to occur with a smaller set of thematic roles (i.e. objects/nouns) which may be more strongly associated within semantic memory.

3.8. Conclusions

This chapter has presented an investigation into semantic similarity between verbs. It appears that participants perceive verbs as similar in a manner that is consistent with how they list verbs in categories and also how they rate the typicality of verbs within categories. This, to an extent, further validates the investigation of verbs using these experimental methods as it appears that the notion of semantic category can be extended into the domain of actions and verbs, although these categories may be less rigid than noun categories with greater overlap. More significantly, the analysis of similarity rating provides further validity to the notion that actions can be organised along a typicality gradient within categories, or at least in relation to a more general action. The analysis of semantic features provided little conclusive insight into how participants use this information in judging semantic similarity and further investigation is warranted. It is likely that speakers use additional experiential knowledge when being asked to rate typicality and/or similarity between actions, and this is likely to be to a greater extent that speakers do for objects (e.g. as in the role of ‘theories’ in categorisation; e.g. Murphy & Medin, 1985).

Chapter 4 Online Psycholinguistic Investigation of Action/Verb Organisation in Semantic Memory and Language Processing

4.1. Aims of Chapter

This chapter presents an investigation of the organisation and semantic processing of actions/verbs using two online psycholinguistic experimental methods with healthy adult speakers of English. The two previous chapters have reported investigations using offline methods which have been effective for understanding the content of semantic representations (i.e. category listing and feature listing) and have insights into speakers' perceptions of organisation (i.e. typicality and similarity rating). The current chapter follows up these themes and reports the use of two online experiments: (1) category verification, where participants gave a yes/no judgement as to whether a presented verb or noun belonged within a pre-identified semantic category; and (2) a semantically primed picture naming task where participants named a picture of an action or object following prior unconscious exposure to a written verb or noun.

Both tasks reported here involve semantic processing but they impose differential demands on language comprehension and language production. Where category verification relies on comprehension (i.e. processing up to semantic processing), semantically primed picture naming involves both comprehension and production. The use of these tasks therefore attempted to identify how verbs and nouns are similarly and differentially processed. Within each task, verb and noun retrieval was probed within discrete experimental blocks but all participants completed the tasks with both verb and noun stimuli. The verb and noun stimuli were selected according to the same semantic principles (i.e. categorical relations) according to data reported in Chapter two to allow direct comparison. As with the previous two chapters, much of the previous related research has been conducted in relation to the investigation of noun retrieval. There appears to have been no investigation into verb retrieval within a category verification task, while there has been some limited exploration of verb retrieval within semantic priming tasks. However, these investigations have not usually drawn direct comparisons between verb and noun retrieval.

This chapter continues with further discussion of the use of online psycholinguistic experimentation in the investigation of semantic memory and introduces some of the principle methods that have been applied. The chapter will then present further discussion of category verification as an online method and then describe the methods and results of a category verification task investigating effects on verb and noun processing. This will be followed with a discussion of semantic priming as a second online task and then describe a series of experiments using semantically primed picture naming to investigate verb and noun processing. In contrast to the previous chapters, individual discussion will be presented following the reporting of each experimental task and the chapter will end with general conclusions.

4.2. Background

4.2.1. Online investigation of semantic memory

While the experiments reported in Chapters two and three (i.e. category listing, typicality rating, similarity rating, and semantic feature listing) offer insight into the content of semantic memory, they also have limitations. Each of these tasks is limited to revealing content of semantic memory when participants have conscious awareness and even control over their own performance, i.e. they can afford time to be deliberate before either rejecting or committing to a response. This may provide data that is a biased representation of the content and organisation of semantic memory. For example, it has been argued that participants tend to list semantic features of objects that tend to be distinctive and they tend not to list more obvious features despite them clearly being relevant to the concept (see Cree & McRae, 2003; Medin, 1989, for discussion). These tasks are therefore limited in terms of how they can be interpreted as illustrating actual processing as and when it occurs.

In order to complement the findings from offline experiments, online experiments can be employed. Online methods involve the experimenter imposing time restrictions on participants and their responses. As such, participants are generally instructed to respond ‘as quickly and accurately as possible’ and the dependent variables are usually response times and error rates with these being compared across experimental conditions. The results can then be interpreted as a closer representation of processing as it occurs in real-time within models or theories of semantic processing. So, if response times are longer in a particular experimental condition, then it may be assumed that this involves more complex processing or that processing has been inhibited for some reason. Response times are usually measured from the time that a critical stimulus is presented to the time that the participant gives a response and these responses may be oral (e.g. naming), a button press (e.g. to indicate a yes/no decision), or any behaviour that can be timed. An increase in error rates may additionally be interpreted to reflect interference in processing (e.g. competition from co-activated conceptual/lexical representations).

Online methods have frequently been used to reinforce and refute theories of semantic and language processing which have often been developed from findings using offline methods. Smith et al (1974) used a category verification task to provide evidence for a two-stage feature-comparison theory within comprehension processing. Meyer & Schvaneveldt (1971) conducted a semantically primed lexical decision task and results from this and similar investigations led to the development of semantic network and spreading activation theories of semantic memory (e.g. Collins & Loftus, 1975). Levelt, Roelofs & Meyer (1999) used a variety of online investigations as a basis for the development of a complete model of lexical access and its representation within a computational model (i.e. WEAVER++). In addition, online tasks are becoming more frequently used in populations with language impairments in order to provide further tests of models of semantic memory and language processing to see whether they effectively account for patterns observed when processing is impaired (e.g. Kiran, Ntourou & Eubank, 2007; Wilshire, Keall, Stuart & O'Donnell, 2007).

4.3. Category Verification of Verbs and Nouns

4.3.1. Background

Category verification explores the organisation and encoding of categorical relations within semantic memory. Participants may be required to verify the validity of sentences (e.g. '*a canary is a bird,*' or, '*a canary is a fish*'; e.g. Collins & Quillian, 1969) or may have to judge whether two words are in a categorical relationship (e.g. *bird-canary,* or, *bird-fish*). Category verification may explore relationships along a vertical dimension of categorisation (e.g. superordinate-subordinate, as above) or may look at the horizontal dimension (i.e. to verify whether two items belong to the same category or are from different categories; e.g. *canary-robin*). Generally, researchers manipulate the relationship of the positive items (i.e. where a verification response should be 'yes' to indicate the two items share a categorical relationship) in order to investigate what factors influence the speed of positive decisions.

Category verification research appears to have focused exclusively on exploring categorical relations of objects/nouns with no apparent extension into the domain of actions/verbs. This may be a reflection of the assumption that verbs do not enter into hierarchical categorical organisation in the same way as nouns (e.g. Graesser et al, 1985; Huttenlocher & Lui, 1979). However, as discussed in Chapter two, resources such as the WordNet lexical database (Miller & Fellbaum, 1991) have adopted hierarchical principles to illustrate how verbs enter into similar entailment relations as nouns. For example, it is possible to state that '*washing is a way of cleaning*' just as it is possible to say '*apple is a type of fruit*', whereas it is incorrect to state that '*cleaning is a way of washing'* just as it is to state '*fruit is a type of apple*'. These one-way entailment relations also clearly differ from statements where no relation exists between the two concepts (e.g. '*washing is a way of jumping*', and, '*apple is a type of bird*'). So even if actions/verbs do not enter into hierarchical and categorical relations in precisely the same manner as objects/nouns, there appears to be some potential validity in using a category verification task as it may highlight which factors, if any, influence speed of processing for verbs where they do enter similar entailment relations as nouns do. This is especially pertinent given the similarities that have been highlighted earlier in this thesis where participants' performance with actions/verbs has paralleled that with objects/nouns (e.g. category listing and typicality rating).

Collins & Quillian (1969) were among the first to report the use of a category verification task in providing evidence for hierarchical organisation of objects/nouns. While this did not explicitly investigate the within-category factors that influence response time, their results are insightful in demonstrating the validity in using such a task to test experimental hypotheses. This study demonstrated that participants were faster to verify statements that involve traversing fewer levels of hierarchical structure. Participants were fastest to verify the truth of '*a canary is a canary*' (which does not involve traversing hierarchical levels; $M \approx 1000$ msecs), compared to '*a canary is a bird*' which involved traversing at least one level of hierarchical structure ($M \approx 1170$ msec), and this was also faster than '*a canary is an animal*' which involved traversing at least two levels of hierarchical structure ($M \approx 1240$ msec). This was taken as evidence that semantic and lexical search processes occur in a strict hierarchical manner where searching must exhaust all information encoded at lower levels before progressing onto higher levels of organisational structure.

Smith et al (1974) used category verification to investigate the role of typicality in determining response time. They found that category members that were rated as highly-typical of the category were verified faster than category members that were less-typical. They used this result as evidence for a two-stage feature comparison process whereby potential category members' semantic features were compared with those specified by the category as either defining or characteristic. Subsequent research has also used category verification to suggest that the distinction between defining and characteristic features is unnecessary as a metric of overall featural similarity was enough to correlate with typicality and therefore predict response times in category verification (e.g. Hampton, 1979).

Armstrong, Gleitman & Gleitman (1983) disputed the finding that typicality was an influential factor in category verification response time and in fact argued that typicality was not a valid organisational principle of semantic memory. They found that participants were able to rate some members of so-called well-defined categories as more typical than others and that these typicality ratings did indeed predict response time in category verification. These categories included those whereby a classical definition of 'necessary and sufficient' features could be applied (e.g. *even numbers, odd numbers, females,* and *plane geometric figures*), and which should not therefore show a typicality gradient in internal category structuring. They also observed the same finding in more natural categories, such as *fruits*, *sports*, *vegetables*, and *vehicles*. Therefore, Armstrong et al (1983) argued that typicality, as identified in previous research, could not possibly represent the phenomenon that researchers had considered it to and that it likely represented an as yet unknown variable or set of variables. In response to Armstrong et al (1983), Larochelle, Richard & Soulières (2000) conducted a similar study and found that typicality effects were only present in well-defined categories when these had a large number of category members (e.g. *languages*, *numbers*, *parts of the human anatomy*) where typicality was likely to be determined on the basis of familiarity and frequency of exposure. In smaller well-defined categories (e.g. *seasons*, *continents*, *planets*), no such typicality effects were observed in either typicality rating or category verification. In comparison, typicality effects were present in both large and small natural categories. This therefore provided evidence for the validity of typicality as an organisational principle within categories and consequently the validity of typicality as a predictor of response time in category verification tasks.

The studies of Armstrong et al (1983) and Larochelle et al (2000) demonstrate the need to understand and control, or account for, potentially confounding variables. In order to address this issue, Hampton (1997) suggested that an ideal method of exploring category verification would be to use matched conditions where all potentially confounding variables are held constant with the exception of the independent variable of interest. So, if typicality was the variable of interest then the stimuli would be matched for all other lexical and psycholinguistic variables (e.g. lexical frequency, familiarity, associative frequency, and so on). However, it was also conceded that this is increasingly difficult to achieve as increasing numbers of variables and interactions between variables are found to influence psycholinguistic processing in a variety of ways.

Where a matched sets design may be difficult to effectively achieve, an alternative approach has been to use regression analysis. This permits a relatively free selection of experimental stimuli, allowing the researcher to obtain insight into which variables may influence response time. Although regression analysis itself is not as statistically powerful as matched sets analysis, it may be an insightful first step to allow researchers better understanding of potential confounding variables which would need to be controlled in further matched sets designs. Perhaps more importantly, it also gives insight into which variables do not need to be controlled. This was the approach taken by Hampton (1997) who first conducted a regression analysis of verification response times with inclusion of variables including typicality, familiarity, and category dominance (i.e. production frequency within the category). Both typicality and category dominance were found to be influential predictors of positive (i.e. 'yes') response times whereas familiarity was not. Therefore, in a second experiment using a matched sets design, one condition held typicality constant while category dominance was free to vary, then in a second condition, category dominance was held constant with typicality free to vary. It was found that both typicality and production frequency influenced response time in that more typical and more dominant category members were responded to faster than less typical and less dominant category members respectively. These independent and unique effects of typicality and category dominance, in conjunction with no effect of familiarity, led Hampton (1997) to conclude that semantic processing underlying category verification decisions relied on principles of association and similarity but not on the degree of exposure to an item or concept.

A further study that used regression analysis, by Larochelle & Pineau (1994), investigated the effects of several variables within and across several semantic categories of objects/nouns. The variables included: typicality, category dominance, instance dominance (i.e. the strength of association from a category member to its category; in effect the reverse of category dominance), familiarity, and lexical frequency. Within individual categories, typicality was found to be the most significant predictor of positive responses whereas familiarity was found to be the most significant predictor variable for negative response times (i.e. 'no' decisions) where the more familiar the item, the quicker the decision was made. However, when regression analysis was conducted across categories, the only significant predictor for both positive and negative responses was found to be familiarity. This study also contradicted some findings of previous work by Casey (1992) who found typicality to be a unique predictor of positive response times and familiarity to be a predictor of negative response times.

From a brief review of category verification, it can be seen that results can vary as a consequence of methodological factors and the variables that are considered and accounted for. Typicality has frequently, although not always, been found to be influential in predicting response time and this effect has been seen to persist in speakers with language impairment (e.g. Kiran & Thompson, 2003). These findings are generally compatible with prototype perspectives on categorisation where category members are judged against an 'ideal' category prototype (e.g. Rosch & Mervis, 1975). What has yet to be investigated however is the influence of typicality and other variables on category verification decisions to actions/verbs. If similar patterns are observed, i.e. that typicality influences response times to verifying category membership of verbs, this would strengthen the argument that actions/verbs tend to cluster around a core meaning (i.e. category) and that some actions/verbs are closer to this core meaning than others. This would also strengthen the argument for a unitary semantic system which imposes comparable principles of representation for both objects and actions (e.g. Vigliocco et al, 2004).

The following sections report the method and results of a category verification experiment for actions/verbs (e.g. *bending-breaking*; *trotting-breaking*) and objects/nouns (e.g. *orange-fruit*; *elephant-fruit*). This was used to investigate the respective influence of several psycholinguistic variables on participants' response times to positive verifications and to errors in positive verifications (e.g. responding 'no' when 'yes' is expected). The variables included in the analysis were: typicality, production frequency, familiarity, lexical frequency, mean rank, and number of letters. The analysis used regression methods to provide a preliminary investigation of the following questions:

1) Do particular psycholinguistic variables influence participants' performance within a category verification task:
    a. With actions/verbs?
    b. With objects/nouns?

2) If particular variables influence performance, is their relative influence comparable between word classes?

3) Are influential variables consistent between a group level analysis and analysis of individual participants?

4) Are influential variables consistent across and between individual semantic categories?

4.3.2. Method

Participants

Ten participants completed this experiment (*M age* = 21, *SD* = 1.9, *range* = 18-25). All participants were native English speakers recruited from a University psychology department subject pool. All gave written consent and were paid for their participation. The sample included four males and six females (9 right-handed, 1 left-handed).

Stimuli selection

Targets were selected from four verb categories and four noun categories to serve as positive response stimuli within the category verification task (i.e. targets which would be congruous with the presented category and which would require a ‘yes’ response). There were difficulties selecting an equal number of targets from each category (primarily with verb targets) due to there being limited normative data available for each predictor variable that would be entered into regression analyses. Therefore, an unequal number of targets were drawn from each category which possessed values for each variable. The categories and their respective numbers of targets were: *breaking* (n = 16 targets); *making* (n = 18); *cleaning* (n = 14); and *talking* (n = 12). In order to create equal sized presentation blocks, additional positive response targets were included according to data obtained in within category listing (see Chapter 2) but these were not included within regression analyses as they did not possess values for all predictor variables (generally, either lexical frequency and/or familiarity were unassigned). The four noun categories of *birds*, *clothing*, fruit, and *furniture* each had 18 positive target responses for which all predictor variables had values. A complete list of experimental stimuli is provided in Appendix N (verb stimuli) and Appendix O (noun stimuli).

Whereas the selection of verb stimuli was restricted to a relatively small set of targets for which normative data were available, there was potential for a wider selection of noun targets. The selection of noun targets attempted to ensure primarily an even spread of targets in terms of production frequency, typicality, and lexical frequency although no systematic method of selection was chosen to ensure this. In fact, direct equivalence of noun and verb stimuli was impossible as values for predictor variables (i.e. lexical frequency, typicality, and familiarity) were gathered from different sources (see below).

All verb targets which were entered into regression analyses were associated with values for each of the following independent variables: number of letters (based on progressive form of verb); production frequency, expressed as a proportion v; mean rank position; typicality (i.e. all from Chapter 2); raw lexical frequency (i.e. combined frequency of infinitive, progressive, past tense, and third person singular form of verb obtained from the British National Corpus); and familiarity (obtained from the MRC database whereby familiarity is rated according to the participant's familiarity with the written word; Wilson, 1988).

All noun targets entered into regression analyses were associated with values for the same variables as verb targets although their source was occasionally different. Raw lexical frequency values were obtained by combining frequency counts of singular and plural forms of the noun from the British National Corpus. Typicality and familiarity ratings were taken from Hampton & Gardiner (1983; where familiarity was rated according to the participant's familiarity with the meaning of the written word in the context of other members of the same category – this may therefore be inferred as a slightly different measure of familiarity in comparison to those used for nouns). Ideally, familiarity ratings would also have been taken from the MRC database as was done with verb targets; however, a number of noun targets did not have familiarity ratings within the MRC database. In addition, the range of the Hampton & Gardiner (1983) typicality scales was from 1 to 5 rather than from 1 to 7 as were obtained for verbs. Given that the aim of this category verification task was to provide a preliminary investigation of the relative contribution of each independent variable on response time, the differential sources of data were not considered to be problematic.

Negative targets that were also presented with each category were selected randomly from the entire corpus of words that were elicited during category listing (i.e. Chapter 2). The only selection criterion was that the negative responses had not appeared in the category it was to be associated with or any of the remaining experimental categories (e.g. as a number of verbs had appeared in more than one category during the category listing experiment).

Design and procedure

The experiment was programmed and run using DMDX experimental software (see Forster & Forster, 2003). Each word class (i.e. verb and noun) was presented in discrete experimental phase with a short break between. The order of word class presentation was counterbalanced between participants. Within each word class phase, a blocked design was used so that participants saw one category name followed by a series of words on which to make category verification decisions. A blocked design was preferred over a random design to avoid potential strategic effects within participants' response behaviours (see Hampton, 1997). Within each word class phase, DMDX was programmed to present the four categories in a different random order for each participant. DMDX also presented the 36 target words associated with each category (i.e. 18 positive and 18 negative target items), in a different random order for each participant. When each category was completed, participants moved onto the next category in their own time by pressing the spacebar.

Participants were instructed that they would see a word representing a category across the upper half of the display. They were told that a series of words would then appear underneath the category and that they should indicate whether they felt the word was associated with the category. They were instructed to indicate 'yes' using the Ctrl key on the side of the keyboard of their preferred hand (e.g. if right handed then they should press the right Ctrl key) and to indicate no with the Ctrl key on the side of their non-preferred hand. Also, before each experimental phase (i.e. nouns and verbs), participants completed two practice category blocks of the relevant word class each containing 10 items (5 positive and 5 negative items). The two categories and associated target items (both positive and negative) appearing in the practice blocks did not appear at any stage during the experimental blocks. During these practice blocks, participants had the chance to ask any questions and the experimenter was able to provide any feedback (e.g. to remind participants to keep their fingers on the Ctrl keys while the experimental blocks were running).

Participants were seen individually and none reported any difficulties understanding or carrying out the task. Each experimental session lasted approximately 25 minutes.

4.3.3. Results

Predictor variables

Table 4.1 and Table 4.2 present correlation data for the predictor variables associated with positive noun and verb target items that would be entered into regression analyses. Also discussed are the tolerance values for each variable. The tolerance value gives an indication as to the variable's interdependence on linear combinations with other variables allowing the extent to which each variable can make a unique contribution to the regression model to be determined. A tolerance value of 0 would suggest that the variable is fully predictable, and therefore dependent, on a linear combination with other variables. A tolerance value of 1 would suggest that the variable is fully independent from other variables.

	# letters	Prod freq	Mean rank	Typicality	Lex freq	Familiarity
# letters	1					
Prod freq	-.178	1				
Mean rank	.279**	-.494***	1			
Typicality	.004	-.553***	.534***	1		
Lex freq	-.308**	.290**	-.293**	-.131	1	
Familiarity	.211*	-.509***	.303**	.583***	-.240*	1

Note.: * p < .05; ** p < .01; *** p < .001

Table 4.1 Pearson correlation matrix of independent variables (across nouns categories)

	# letters	Prod freq	Mean rank	Typicality	Lex freq	Familiarity
# letters	1					
Prod freq	.033	1				
Mean rank	-.211	-.470***	1			
Typicality	-.236*	-.348**	.450***	1		
Lex freq	-.039	.029	.041	-.049	1	
Familiarity	.004	.066	-.137	-.026	.300**	1

Note.: * p < .05; ** p < .01; *** p < .001

Table 4.2 Pearson correlation matrix of independent variables (across verb categories)

There was a greater degree of intercorrelation among predictor variables for the noun targets than for verbs with 12 of 15 possible correlations reaching significance compared to 5 of 15 for verbs. Despite the presence of intercorrelations, tolerance values suggested that each predictor variable was able to make a unique contribution to regression models as the lowest values were .445 (typicality) for nouns and .602 (production frequency) for verbs.

In addition to demonstrating that predictor variables have the potential to make unique contributions to regression models, Larochelle & Pineau (1994) stated that it is important to demonstrate that target items reflect as full a variety of value for each variable as possible. For example, it is important for target items with high values for production frequency, typicality, and so on, to be included within the data set in addition to target items with moderate and low values for production frequency, typicality, and so on. Table 4.3 presents descriptive statistics for each predictor variables including the mean value, standard deviation and minimum and maximum values. While the data for nouns and verbs are not necessarily directly comparable (as different sources or methods of data collection were used to calculate typicality, lexical frequency, and familiarity), variables are reasonably distributed and are not concentrated around the mean. However, as is the case with some scales, such as typicality, values tend to be skewed towards the positive end of the ratings scale and so it is difficult to include target items which are truly atypical of a category but which are nevertheless category members.

	Nouns				Verbs			
	M	SD	Min	Max	M	SD	Min	Max
# letters	5.97	1.99	3	12	7.7	1.14	6	11
Prod freq	0.45	0.28	0.09	1.00	0.29	0.22	0.09	0.86
Mean rank	10.18	3.77	1.57	22	5.89	2.11	1.96	11
Typicality†	1.82	0.69	1	3.59	2.69	0.83	1.1	4.55
Lex freq†	1856	3623	5	21594	2994	4354	26	23173
Familiarity†	1.47	0.45	1.00	3.16	545	40.3	449	632

†Different measurements used between noun and verb measures

Table 4.3 Descriptive data for predictor variables

Errors and data replacement

Analysis was based exclusively on responses to positive target items (i.e. targets where the expected verification response was ‘yes’ to indicate that the target did belong to the respective category). As with similar studies (e.g. Casey, 1992; Larochelle & Pineau, 1994), extreme response times were assumed to be anomalous and were transformed to minimise their impact on subsequent analysis. Responses under 250 msecs were replaced with the participant’s mean response to correct positive verifications. Responses over 2500 msecs were trimmed to 2500 msecs. Also in accordance with previous studies, incorrect responses were also replaced with the participant’s mean response to correct positive verifications.

Across the entire data set of positive responses, the combined effect of data replacement and transformation procedures led to 33 noun responses being replaced (5.16% of total positive responses) and 127 verb responses being replaced (19.84%).

Descriptive statistics

Table 4.4 presents descriptive statistics on response times to positive target items that were entered into regression analyses both across and within noun and verb categories.

	M	SD
Across nouns	765.98	117.48
Birds	789.09	122.55
Clothes	749.83	82.78
Fruit	713.40	113.29
Furniture	811.60	130.23
	M	SD
Across verbs	972.34	152.75
Breaking	980.92	149.52
Cleaning	905.61	130.04
Making	1077.24	123.35
Talking	881.42	135.50

Table 4.4 Mean response times for positive verifications

Across category analysis

Regression analysis was employed to investigate influence of predictor variables on response time and error proportions across categories. As with Larochelle & Pineau (1994), the analyses were conducted both without and with the inclusion of dummy variables. Four sets of binary dummy variables were introduced to the data set to allow for any particular effects on response time of each semantic category. For example, there may be something particular about one semantic category that makes response times faster relative to responses for the other three categories. Therefore, for one set of dummy variables, all members of one semantic category (e.g. *breaking*) were assigned a binary value of 1 whereas all other members of the other three categories were assigned a value of 0. The same was done for all four categories within each word class. Therefore, if there are any particular effects of semantic category on response time which are not captured by the other independent variables, dummy variables would enter into regression models.

Semi-standardised regression coefficients. This analysis employed a method of regression where all predictor variables were entered into the model and their respective influence could be assessed by calculating semi-standardised coefficients (e.g. Casey, 1992; Larochelle and Pineau, 1994). Semi-standardised coefficients were calculated by multiplying the non-standardised coefficient value by the standard deviation for the relevant predictor variable. Within response time analysis, the semi-standardised coefficient value represents the amount of change in response time (in msecs) that occurs following a change of 1 SD in the predictor variable. Similarly, within the error analysis, the semi-standardised coefficient represents the change in error proportion following a 1 SD change in the predictor variable.

Table 4.5 presents the semi-standardised coefficient values for all predictor variables on the group mean response time to positive category verification for both nouns and verbs. Separate rows report the coefficients when regression models were derived without and with the inclusion of dummy variables. Table 4.5 also includes the associated t-values for each predictor variable.

Table 4.6 presents the semi-standardised coefficients and t-values for predictor variables' influence on the error proportion of each target item. Error proportion was calculated over all 10 participants where a value of 1 indicated that no participants made an incorrect positive verification decision on an item and a value of 0.3 (the lowest proportion recorded) indicated that only 3 participants gave a correct positive verification decision (i.e. 7 participants indicated 'no' when the expected response was 'yes'). Where Tables 4.5 and 4.6 present semi-standardised coefficient values, the following points (*i*-*iv*) present statistics on the reliability of the regression models that derived these coefficient values.

i) Response time (without dummy variables)

The regression model derived for nouns yielded a significant *F*-statistic (F (6, 71) = 4.85, $p < .001$) although the model was only able to account for 25% percent of the total variance (adjusted $R^2 = .245$). The regression model for verbs was not significant (F (6, 59) = 1.08, $p = .389$) and accounted for less than 1% of the total variance (adjusted $R^2 = .008$).

ii) Response time (with dummy variables)

The model for nouns remained significant when dummy variables were included (F (9, 71) = 3.98, $p < .001$) and accounted for 27% of the total variance (adjusted $R^2 = .274$). The model for verbs also produced a significant model (F (9, 59) = 3.34, $p = .003$) and accounted for 26% of the total variance (adjusted $R^2 = .263$).

iii) Errors (group analysis without dummy variables)

The model for nouns was significant (F (6, 71) = 3.02, $p = .011$) but accounted for only 15% of the total variance (adjusted $R^2 = .146$). The model for verbs was also significant (F = (6, 59) = 8.99, $p < .001$) and accounted for 45% of the total variance (adjusted $R^2 = .448$).

iv) Errors (group analysis with dummy variables)

The model for nouns remained significant (F (9, 71) = 2.62, $p = .012$) but still accounted for just 17% of the total variance (adjusted $R^2 = .171$). The model for verbs also remained significant (F (9, 59) = 7.23, $p < .001$) and accounted for 49% of the total variance (adjusted $R^2 = .487$).

	# letters	production freq	mean rank	typicality	lexical freq	familiarity
	semi-standardised coefficients					
without dummy variables						
Nouns	-4.55	-69.88	-32.83	18.85	-3.62	-10.58
Verbs	-14.9	-18.29	-23.54	42.94	8.71	-0.89
with dummy variables						
Nouns	-7.86	-58.38	-20.17	16.97	-10.87	-0.76
Verbs	6.78	-4.17	-14.2	55.36	8.71	11.29
	# letters	production freq	mean rank	typicality	lexical freq	familiarity
	t-values					
without dummy variables						
Nouns	-0.334	-4.359	-2.052	1.038	-0.244	-0.653
Verbs	-0.699	-0.704	-0.929	1.739	0.474	-0.042
with dummy variables						
Nouns	-0.573	-3.311	-1.110	0.952	-0.862	-0.044
Verbs	0.351	-0.178	-0.594	2.374	0.390	0.599

Table 4.5 Semi-standardised coefficients and t-values of predictor variables for group mean response times (positive responses)

	# letters	production freq	mean rank	typicality	lexical freq	familiarity
	semi-standardised coefficients					
without dummy variables						
Nouns	0.02	0.01	<0.01	-0.03	<0.01	<0.01
Verbs	0.01	<0.01	-0.01	-0.13	<0.01	<0.01
with dummy variables						
Nouns	0.03	<0.01	-0.01	-0.03	<0.01	0.04
Verbs	<0.01	-0.01	-0.01	-0.14	<0.01	0.04
	# letters	production freq	mean rank	typicality	lexical freq	familiarity
	t-values					
without dummy variables						
Nouns	2.229	0.767	0.136	-2.038	0.29	-0.001
Verbs	0.625	0.076	-0.383	-5.533	0.577	-1.341
with dummy variables						
Nouns	2.552	0.185	-0.696	-2.032	0.949	0.171
Verbs	-0.025	-0.196	-0.54	-5.476	0.622	-1.753

Table 4.6 Semi-standardised coefficients and t-values of predictor variables for group error proportion (positive responses)

Within category analysis

In addition to across category analysis, regression models were calculated for response times for each category individually. Naturally, these models were calculated on a small number of observations but such analysis could yield indications as to how predictor variables' influence is consistent or variable between categories of target items. This is especially worthy of consideration given that a number of dummy variables entered into final models when they were included as predictor variables in regression analyses.

Semi-standardised regression coefficients. Table 4.7 and Table 4.8 present the semi-standardised coefficients and associated t-values for noun and verb categories respectively. For noun categories, production frequency continues to have a consistent influence in each category both in terms of size of effect and direction, although the effect is larger for *fruit*. For each category, an increase in 1 SD of production frequency leads to a reduction in response time ranging from 43.3 to 67.8 msecs. Other predictor variables generally had a smaller effect of response time and are more variable in terms of the direction of effect. For example, an increase in 1 SD unit of familiarity leads to a reduction of 53.9 msecs in *fruit* but an increase of 50.1 msecs in *birds*.

For verb categories, typicality has a relatively large effect on all categories except *making*. In the three categories that it affects, an increase in 1 SD unit of typicality (indicating that the typicality rating is higher and therefore the target is less typical) leads to an increase in response time ranging from 75.5 to 124.1 msecs. Again, with the remaining predictor variables there is more variation with the size and direction of effects in different verb categories. For example, an increase of 1 SD unit of mean rank leads to a large increase in response time for *talking*, a smaller increase for *cleaning*, a small reduction in response time for *breaking*, and a larger reduction for *making*.

	# letters	production freq	mean rank	typicality	lexical freq	familiarity
	semi-standardised coefficients					
birds	-2.36	-45.01	50.07	-41.49	-30.46	50.08
clothes	-9.31	-47.87	-29.67	8.05	11.65	-3.25
fruit	10.68	-67.80	-51.99	88.80	-23.85	-53.89
furniture	-3.63	-43.32	-13.36	20.96	-29.89	9.92

	# letters	production freq	mean rank	typicality	lexical freq	familiarity
	t-values					
birds	-0.086	-1.153	1.363	-1.043	-0.990	0.952
clothes	-0.262	-1.382	-0.950	0.256	0.344	-0.087
fruit	0.932	-1.596	-1.308	1.960	-0.854	-1.485
furniture	-0.086	-0.627	-0.243	0.376	-0.626	0.171

Table 4.7 Semi-standardised coefficients and t-values of predictor variables for group mean response time within noun categories (positive responses)

	# letters	production freq	mean rank	typicality	lexical freq	familiarity
	semi-standardised coefficients					
breaking	8.48	72.68	-6.45	124.05	36.87	-20.72
cleaning	25.74	36.86	62.79	75.51	-17.76	5.77
making	9.54	-10.27	-70.30	-12.09	17.66	65.12
talking	6.37	31.30	134.59	94.71	10.10	127.41

	# letters	production freq	mean rank	typicality	lexical freq	familiarity
	t-values					
breaking	0.187	0.917	-0.084	1.563	0.606	-0.347
cleaning	0.578	0.362	0.834	1.343	-0.284	0.118
making	0.296	-0.212	-1.775	-0.195	0.582	1.719
talking	0.144	0.706	2.256	2.111	0.201	2.432

Table 4.8 Semi-standardised coefficients and t-values of predictor variables for group mean response time within verb categories (positive responses)

Identifying influential predictor variables. Table 4.9 and Table 4.10 presents the correlations with group mean response time and predictor variables for noun categories and verb categories respectively.

	Across categories	Birds	Clothes	Fruit	Furniture
# letters	-.021	.002	-.097	.036	.138
Prod freq	-.501***	-.657**	-.268	-.579**	-.537*
Mean rank	.070	.334	-.053	.264	.386
Typicality	.292**	.251	.163	.571**	.401*
Lex freq	-.106	-.385	-.025	.162	.423*
Familiarity	.220*	.505*	.042	.180	.486*

Note.: * $p < .05$; ** $p < .01$; *** $p < .001$

Table 4.9 Pearson correlation values with mean response time and predictor variables (across- and within- noun categories)

	Across categories	Breaking	Cleaning	Making	Talking
# letters	-.134	-.187	.101	.001	-.139
Prod freq	-.069	-.125	-.453	.196	-.361
Mean rank	.054	.223	.412	-.430*	.428
Typicality	.286*	.471*	.501*	-.215	.551*
Lex freq	.064	.331	-.270	.120	-.331
Familiarity	.018	.037	.021	.345	.057

Note.: * $p < .05$

Table 4.10 Pearson correlations with mean response time and independent variables (across- and within- verb categories)

For nouns, production frequency continued to play an influential role as it correlated significantly with response time in all categories except *clothes.* For verbs, typicality correlated strongly with response time in three of the four categories.

4.4.4. Discussion

Summary of main findings

In across category analysis of nouns, production frequency was the most consistent predictor of response time both in group and individual analyses such that the more strongly a noun was associated with its category, the faster it was verified as a category member. Also, number of letters and typicality were significant predictors of error rates within group level performance. These findings were consistent regardless of whether dummy variables were entered into the regression analyses or not. The across category analysis of response time to verbs showed that typicality was the most consistent predictor. This indicated that, as verbs became more typical within the category, the faster they were verified. Typicality also predicted error rates with verbs. Again, these findings were consistent regardless of whether dummy variables were entered into analyses.

In within category analysis of nouns, production frequency was again most consistently a predictor variable of response time with the exception of the *fruit* category where response time was more effectively predicted by typicality. Within verb categories, typicality was consistently the most predictive variable with the exception of the *making* category where response time was more effectively predicted by mean rank and familiarity.

Discussion of main findings

Noun response times are most consistently predicted by production frequency. This finding is consistent with previous research by Hampton (1997) who concluded that association strength (i.e. production frequency) is an influential predictor of response time to positive verification decisions. However, Hampton (1997) and other researchers (e.g. Casey, 1992; Larochelle & Pineau, 1994) also found a unique influence of typicality on positive verification times for nouns within object categories.

Noun errors are most consistently predicted by typicality. Again, this finding is consistent with that of Hampton (1997; Experiment 1) who found that typicality was the only significant predictor of the probability of a 'yes' response to a category member when a 'yes' response was expected.

Verb response times are most consistently predicted by typicality. This finding, in conjunction with the finding that production frequency did not predict performance, clearly indicated that category verification for nouns and verbs was influenced by different variables. This is an intriguing result especially given that typicality was not found to influence categorisation response times to nouns even though it may have been expected to do so. The effect of typicality was also robust enough to persist when dummy variables were introduced into regression analyses. A further intriguing finding was with regard to the impact of including dummy variables within this particular analysis. The variation accounted by the regression model improved from less than 1% to 26% once dummy variables were included within regression models. Given that dummy variables were included to partial out effects that membership of individual semantic categories may have on response time, this increase in the percentage of variance accounted for, suggests that for verbs, the semantic category may influence response time to some degree.

Verb errors are most consistently predicted by typicality. This finding is again consistent with that of Hampton (1997) who found typicality to be the only predictor of errors within noun category verification. Given that there were more errors within verb categorisation than noun categorisation, this finding may be more robust than for the nouns.

Limitations and further research

In general, while the regression models were mostly significant (i.e. *F*-values), they were also relatively poor at accounting for the overall variances present. This is potentially due to either (1) a small sample of responses on which regression models were derived, or (2) variables that were not considered and entered into regression analyses, or both. Given that the variables entered into the analyses included those which have most frequently been found to influence performance in category verification tasks (i.e. production frequency/association strength, typicality, and familiarity), it is likely that the relative poor performance of the models was due to a small sample size. The largest number of items entered into regression analysis was 72 (across noun category analysis) which is the same as the number used by Larochelle & Pineau (1994) for across category analysis but inferior to the likes of Hampton (1997) who presented between 68 and 110 items (including positive and negative items) for each of 12 categories. In the current experiment, the selection of verb stimuli was hampered by the lack of data regarding familiarity and so the maximum possible number of verbs was selected according to the presence of data for all variables. Consequently, the number of verbs that were selected also dictated the number of nouns selected (as the purpose of the current study was to provide as equal a comparison as possible). Therefore, further research may consider including a larger number of stimuli on the proviso that familiarity rating data was available and also that familiarity ratings were collected according to the same principles as for nouns (see section 4.3.2 for description of how familiarity ratings used here may have been measuring qualitatively different phenomena).. This would hopefully lead to the derivation of superior regression models and allow greater confidence in findings.

An issue that could be explored further relates to the error rates. Errors were fairly uncommon in noun category verification but were prevalent in verb category verification. This was insightful in that it allowed regression models to be derived for errors, but it also impacted negatively on the response time analysis as error responses were replaced with a mean response time value. Again, this may be an issue that is overcome more effectively if a greater number of verbs are included within the experimental design.

The notion that verbs have a 'looser' fit with their conceptual referent may give an indication as to possible other variables which may be influencing response time and error rates. As verbs tend to be more polysemous than nouns, it may be that additional variables, such as 'number of meaning senses' would have a greater impact on verb category verification. If verbs have a greater number of senses, they may be more difficult to verify within one specific category within the context of an online task which emphasises that participants should make responses as quickly as possible. In such situations, participants may be cautious about giving a positive verification if their initial processing of the target (or indeed the category) triggers activation of non-target conceptual representations that may conflict with those presented in the task. Category decisions may therefore be easier for nouns as categorisation tends to be more discrete with objects.

4.4. Semantically Masked Prime Picture Naming

4.4.1. Background

The term 'semantic priming' here refers to a paradigm within psycholinguistic research which attempts to elucidate the processes involved in word recognition and word retrieval. More specifically, semantic priming experiments aim to investigate whether prior exposure to one stimulus affects subsequent processing of a second stimulus where the two stimuli share some kind of semantic relation. For example, prior exposure of one stimulus (i.e. a prime) may, under certain circumstances, be found to facilitate (i.e. prime; speed-up) subsequent processing of a second stimulus (i.e. a target). Conversely, under different circumstances, exposure to primes may lead to subsequent inhibited (i.e. slowed) processing of targets. Meyer & Schvaneveldt (1971) were among the first to demonstrate effects of semantic priming when they found participants were faster to make lexical decisions (i.e. to decide if a string of letters was a legal word) when primed by a semantically related word (e.g. *bread-BUTTER*; *M* = 855 msec) compared to when primed with an unrelated word (e.g. *bread-DOCTOR*; *M* = 940 msec). This was argued to reflect that related concepts were stored proximally closer within semantic memory and so the prior activation of a related word and concept meant that the distance to the target was shorter than if the target was unrelated.

The conditions under which semantic priming is investigated affect the direction of priming effects. The various factors which may influence outcomes include: the task being used to elicit priming effects; the specific type of semantic relationship (and strength of that relation) between the prime and target stimuli; the stimulus onset asynchrony (SOA), or, the intervening time between presentation of the prime and the target stimuli; the duration of the prime stimuli; and the presentation method of the prime in terms of whether participants are aware of this or are unaware of its presence in the task. By considering these factors, researchers can infer processing and organisation within conceptual-semantic memory and how this interacts with language processing (see Tulving & Schacter, 1990, for discussion).

Tasks

There are various tasks which have been used to investigate the effect of semantic priming and some general trends which can be identified in terms of the effects that these tasks lead to (i.e. whether they lead to facilitation or inhibition effects). These tasks vary in terms of their processing demands and the responses that participants are required to give but they are consistent in that they look to investigate the influence of prior processing of primes on subsequent processing of targets.

Lexical decision tasks require participants to make a yes/no response as to whether a visually presented letter string constitutes a legal word or not. Such letter strings are seen following presentation of a prime stimulus and it has been shown that prior exposure to primes leads to faster lexical decision to semantically related target words (e.g. Meyer & Schvaneveldt, 1971). Such tasks therefore do not necessitate retrieval of semantic representations of lexical items (as only a decision based on lexicality is required) but the fact that semantic relatedness appears to influence behaviour has been argued to reflect spreading activation within a semantic network in that prior and residual activation of semantically related nodes makes it easier to achieve threshold activation for subsequent related items. Effects in the lexical decision tasks have also been observed in a direct comparison between verbs and nouns. Rösler, Streb & Haan (2001) found that for both verbs and nouns, lexical decisions were fastest when verb and noun prime-target pairs were strongly semantically related (e.g. *sweep-DUST*; *magazine-NEWSPAPER*) compared to prime-target pairs which were weakly associated (e.g. *clean-DUST*; *text-NEWSPAPER*) which were themselves faster than unrelated prime-target pairs (e.g. *stamp-DUST*; *merchant-NEWSPAPER*).

Reading aloud, or pronunciation tasks, require participants to read aloud written words, again following prior exposure to written word (or picture) prime stimuli. This task requires a verbal response as opposed to a button press response as in lexical decision but similarly prime stimuli have been shown to lead to facilitated responses to semantically related target stimuli and that this effect can be modulated by the strength of semantic relation between the two so that targets that are more strong semantically related to the prime will be named faster than targets that are more weakly related to the primes where such effects have been explained in terms of 'post-lexical' expectancy generations where participants (consciously or unconsciously) generate their own expected responses following the prime and before being exposed to the target (e.g. Keefe & Neely, 1990). However, facilitation effects have also been observed under circumstances where participants have been (consciously) unaware of the presence of prime stimuli which does suggest some level of automatic semantic processing as accounting for such effects (e.g. Hines, Czerwinski, Sawyer & Dwyer, 1986).

Picture naming tasks require participants to verbally name pictures following exposure to words, pictures or prior naming of pictures (i.e. paired or blocked picture naming). In picture naming, Lupker (1988) found that picture naming was facilitated when targets were preceded by prior naming of categorically related picture or word primes (e.g. *fox-ELEPHANT*) in comparison to unrelated primes (e.g. *canoe-ELEPHANT*). These findings were argued to again reflect the automaticity of semantic processing as a common processing mechanism in both variants of the task (i.e. when primes were either pictures or words to be named) even when primes could be processed non-semantically (i.e. when reading words aloud). However, within a blocked picture naming task (i.e. where participants are required to name a longer series of pictures sequentially), the naming of categorically related pictures (e.g. mouse, spider, snake, fish, duck) has also been shown to lead to inhibited (i.e. slowed) naming in comparison to sequential naming of unrelated pictures (e.g. Damian, Vigliocco &

Levelt, 2001). Inhibition effects in picture naming have also been observed when primed by categorically related written words when the written words did not even require naming (e.g. Alario, Segui & Ferrand, 2000).

Finally, picture-word interference tasks are variants of picture naming tasks although the presentation of prime and target are presented simultaneously with a word being superimposed over a picture (n.b. although the onset of prime presentation may be slightly before or following presentation of the target with variable effects) and where the participant is instructed to ignore the word and name the picture. Generally, picture-word interference tasks, as the name suggests, have shown inhibitory effects of semantically related primes (i.e. word stimuli) on the naming of targets (i.e. picture stimuli) in comparison to unrelated prime-target pairs (e.g. Hantsch, Jescheniak & Schriefers, 2005; Vitkovitch & Tyrrell, 1999) although it has been suggested that these effects are strongest when primes and targets are presented within a short time of each other (i.e. with an SOA of 0 to 100 msecs, e.g. Glaser & Düngelhoff, 1984). Such effects have been argued to arise from lexical level competition between simultaneously activated representations which are derived from parallel processing routes (i.e. lexically mediated via word stimuli and semantically mediated via picture stimuli) where the delayed naming represents the additional time it takes to resolve competition when the items in competition are more closely semantically related. However, despite the general trend for inhibitory effects in these tasks, facilitation effects have also been observed when prime and target share a categorical relation which conflicts with the suggestion of lexical competition in such tasks (e.g. Mahon, Costa, Peterson, Vargas & Caramazza, 2007). Picture-word interference tasks are also significant for the current purposes as this is the task which has been applied most consistently to investigate verbs in addition to nouns. The general finding with many of the studies using picture-word interference with verbs is again that semantically related prime-target pairs (e.g. *eat-DRINK*) leads to inhibited (i.e. slowed) naming in comparison to unrelated conditions (e.g. *sneeze-DRINK*; e.g. Collina & Tabossi, 2007; Roelofs, 1993; Schnur, Costa & Caramazza, 2002; Tabossi & Collina, 2002). However, it should be highlighted that the presence of inhibition effects between studies, and sometimes between experiments reported in the same study, are variable leading for some to claim "a better understanding of how verbs are semantically related is needed in order to evaluate the cause of the transient semantic interference effects" (Schnur et al, 2002:18).

While general trends and mechanisms can be identified with each of the tasks which aim to investigate semantic priming, there are also numerous examples of contradictions to the general trends. Such contra-results can often be attributed to more fine-grained methodological details concerning the stimuli used within the tasks are how the task is actually presented to participants, as, even when using a particular type of task, there is still potential for great variation with how the task is administered.

Semantic relatedness of stimuli

Studies investigating semantic priming make a distinction between primes and targets that are semantically related (e.g. *boat-TRAIN*) and those that are associatively related (e.g. *nest-BIRD*) (e.g. see for example Alario, Segui & Ferrand, 2000; Bueno & Frenck-Mestre, 2008; McRae & Boisvert, 1998). This is primarily because these two types of relation have been shown to be influenced differently by different experimental manipulations such as stimulus onset asynchrony (see below). Such studies have generally tended to show that priming effects for semantically related pairs tend to occur earlier and when participants are given less time to consider the prime stimuli in comparison to effects for associatively related prime-target pairs (e.g. they occur at shorter SOAs, at shorter prime durations, and these effects are not present for associative pairs). In the current experiments, semantic relatedness refers to primes and targets that share some kind of category-based relation whether it be a category coordinate relation (i.e. members of the same category at the same level of categorical abstraction, e.g. *apple-BANANA*; *frying-BAKING*), or in a hierarchical superordinate-subordinate relation (e.g. *fruit-BANANA*; *cleaning-BAKING*). This therefore assumes some level of semantic featural overlap which previous studies have employed as their criteria for deciding whether prime-target pairs are indeed 'semantically' similar/related (e.g. Bueno & Frenck-Mestre, 2008; McRae & Boisvert, 1998) Therefore, any effects that may be observed would be more likely due to semantic level processing than lexical level processing which is more likely the case with associative pairings.

Several studies using various tasks have investigated the effects of category coordinate primes to responses to target nouns and also verbs. The majority of studies with verbs have found inhibition effects using a picture-word interference task in comparison to unrelated conditions (e.g. *eat-DRINK* versus *sneeze-DRINK*, *cry-LAUGH* versus *fall-LAUGH*; e.g. Collina & Tabossi, 2007; Roelofs, 1993; Schnur, Costa & Caramazza, 2002; Tabossi & Collina, 2002). There has been more varied use of tasks and findings associated with the investigation of noun category coordinates. However, with regard to the current experiment (i.e. a primed picture naming task) the most relevant studies are those involving the naming of pictures where the prominent finding has also been of inhibition effects (i.e. slowed naming responses; e.g. Damian, Vigliocco & Levelt, 2001; Alario, Segui & Ferrand, 2000) although there are exceptions where facilitation has been observed (e.g. Mahon, Costa, Peterson, Vargas & Caramazza, 2007).

In addition to investigating effects of within category coordinates, further research has investigated the effect of related items at differing levels of hierarchical organisation (e.g. subordinate, basic, and superordinate levels of categorisation). The majority if this research has been conducted with regard to nouns due to the difficulty in establishing how verbs enter into hierarchical category relations. The general finding when participants have been required to name pictures within the experimental task (e.g. picture-word interference tasks) has been of inhibition effects in related conditions (e.g. *dog-POODLE*; *flower-LILY*) compared to unrelated conditions (e.g. *cat-POODLE*; *flower-BABOON*) (e.g. Vitkovitch & Tyrrell, 1999; Hantsch, Jescheniak & Schriefers, 2005). Such effects have been found regardless of the assignment of subordinate and superordinate category members to primes and targets (i.e. whether the prime is *flower* and target is *LILY* or whether the prime is *lily* and the target is *FLOWER*; see for example Hantsch, Jescheniak & Schriefers, 2005)

Stimulus onset asynchrony (SOA)

The stimulus onset asynchrony (SOA) refers to the time between the onset of the prime stimulus and the onset of the target stimulus and reflects the time that the participant is able to process the prime before needing to make a response to the target. The length of the SOA impacts on the hypothesised level of processing that a task may be tapping. Where SOAs are considered to be short (i.e. $\leq$ 250 msecs) this is generally assumed to tap automatic semantic processing. This implies that the participant does not have sufficient time to consciously process the prime stimuli completely before they are exposed to the target although there is sufficient time for the prime's semantic representations to be unconsciously processed. Where SOAs is long (i.e. > 250 msecs) other processes in addition to automatic semantic processing are likely to be implicated in any observed priming effects. At longer SOAs, participants may have opportunity to generate expectancies of what the target stimulus may be. For example, if they have time to recognise and process the prime as an animal (e.g. *cat*) they may generate some expectancies that it may be likely that the target is also likely to represent an animal (e.g. *dog*, *lion*, and so on). If their generated expectancies are consistent with the actual target, then their response time may be faster than under other circumstances due to them already partially activating the target's semantic representation. Alternatively, if the target is not in fact another animal (e.g. *hammer*, *jacket*, etc.), then response time may be slower than under other circumstances due to the participant generating expectancies and activating semantic representations of unrelated concepts which must subsequently be inhibited when they are not consistent with the actual target. When SOA is long a further consideration is the associative relation between prime and target. If these are in an associative relation, these may benefit from processing that is non-semantic as they may be frequently co-occurring lexical items (e.g. *tea-coffee*, *fish-chips*, and so on).

Prime duration

Similar to considerations with SOA, the prime duration (i.e. the length of time the prime is displayed for) may also impact on the level of processing that may give rise to any effects on response time to the target. Broadly, primes may be presented in one of two ways, firstly, for a duration that is so short that it occurs before conscious recognition can take place, to the extent that participants cannot report being aware of seeing a prime stimulus; secondly, for a longer duration where participants are able to report seeing a prime stimulus. While there is individual variation of the duration threshold of conscious awareness of stimuli, a prime duration of around 30 msecs has been found to be below threshold levels for the majority of individuals (e.g. REF). Despite participants not being consciously aware of the presence of prime stimuli, such conditions are still able to produce semantic priming effects within priming tasks (e.g. REF). This therefore indicates that under such conditions, automatic semantic processing can still be implicated and affected by semantic relatedness of prime and target. As with SOA, a short prime duration (i.e. 30 msecs) was used in the current experiment in efforts to ensure than automatic semantic processing could be interpreted as being responsible for any observed priming effects.

Prime presentation

The presentation method of the prime refers to whether the prime appears masked or unmasked. Where primes are masked this is a situation where the prime is presented following presentation of a string of other (usually non-alphabetic) characters, such as hash marks (e.g. ##########). Where this other string is present before presentation of the prime, this is referred to as forward masking, where masking also occurs following presentation of the prime, this is referred to as backward masking.

Masking is used in order to mask the orthographic pattern of the visually presented prime and is in contrast to where primes are presented on an otherwise blank screen where it may be possible to identify particular letter patterns due to there being less visual distraction. Masking was used in the current experiment, again to strengthen claims that any priming effects were due to automatic semantic processing as opposed to orthographic pattern recognition and matching (i.e. post-lexical processing systems).

4.4.2. The current investigations and research questions

The research literature that reports the use of priming tasks and their effects on verb processing is sparse and inconsistent. This has often been attributed to poor knowledge with regard to the nature of verbs' semantic representations and how they are inter-related within the semantic system and wider systems associated with language processing. Despite this, there are suggestions that both verb and noun processing can be subject to similar effects of priming under similar circumstances (e.g. Rösler, Streb & Haan, 2001). If this suggestion is upheld across a broader range of experimental tasks then this may be interpreted as further evidence for comparable levels of semantic representation and organisation for verbs and nouns.

The current experiment used a picture naming task with masked written word primes. This task was hypothesised to involve a degree of semantic processing when accessing conceptual-semantic representations via pictures and generating a verbal label for this. As the task also involved processing at multiple modalities of input and output processing, any effects that are observed could potentially be attributed to semantic level mechanisms (as semantic level processing would be common to all modalities involved). To further increase confidence in interpreting any observed effects as semantically mediated, a number of methodological considerations were implemented.

A short SOA (i.e. 130 msecs) between initial presentation of prime and target stimuli was used in order to avoid priming effects that may be based on associative rather than purely semantic relations (e.g. McRae & Boisvert, 1998). The prime stimulus was presented for a duration that is generally considered to be below the level of conscious awareness (i.e. 30 msecs) which has also been argued to strengthen the interpretation of automatic semantic processing (e.g. Van den Bussche, Van den Noortgate & Reynvoet, 2009). Primes were masked by a string of non-alphabetic characters (i.e. ########) in order to reduce potential influence of orthographic pattern recognition and matching (e.g. Holender, 1986). Finally, experimental items were presented just once in a single modality and in the presence of a number of unrelated fillers items. This was an attempt to reduce a possible 'response congruency' effect (e.g. Kuipers et al, 2006) whereby participants may employ subconscious strategies if they become aware that a number of items are drawn from a small number of particular semantic categories.

The following two experiments report the use of a semantically primed picture naming task to investigate respectively the following questions:

1) Do related category coordinate verb and noun primes affect subsequent processing of nouns and verbs respectively when elicited by pictures of actions and objects?

2) Do related category superordinate verb and noun primes affect subsequent processing of nouns and verbs respectively when elicited by pictures of actions and objects?

Following the two experiments that directly investigate these two questions, a further analysis is presented that investigates the interaction of prime relation (i.e. coordinate, superordinate) and prime condition (i.e. related, unrelated) across word classes (i.e. verb, noun). The following sections present the methods and results of the three analyses and these will then be followed with a discussion of the main findings.

4.4.3. Category coordinate semantic priming

Introduction

This first semantically primed picture naming experiment investigates the effects of category coordinate word primes on the production of verbs and nouns elicited through pictures of actions and objects respectively. Based on previous research and the applied methodology which attempted to encourage automatic semantic priming, predictions about participants' performance, in terms of response times, could be made. For nouns, related category coordinates may be expected to inhibit production of noun targets in comparison to unrelated primes (e.g. Alario et al, 2000; Damian et al, 2001). For verbs, related category coordinate primes may also be expected to inhibit naming of verb targets in comparison to unrelated primes (e.g. Collina & Tabossi, 2007; Roelofs, 1993; Schnur et al, 2002).

Method

Participants. Ten participants took part in the experiment (*M* age = 26; *SD* = 8.6; *range* = 19-45). This sample included two males and eight females with nine participants being right handed and one left handed. All participants had normal or corrected vision and gave written consent.

Stimuli selection. Within each word class condition (i.e. nouns and verbs), 24 words were selected as experimental stimuli. The 24 verb stimuli were canonically transitive verbs and were drawn equally from six of the semantic categories for which category norm and typicality had previously been obtained (see Chapter two). The verbs were drawn from the categories *breaking*, *cleaning*, *cooking*, *cutting*, *hitting*, and *making* and had all received high-typicality ratings and had high production frequencies within the respective category. In addition to experimental stimuli, 96 verbs were selected to act as fillers. These included canonically intransitive, transitive, and ditransitive verbs. Fillers were selected so that they were not categorically related to experimental stimuli although some fillers did share categorical and/or associative relations (e.g. *eat*, *drink*). All verbs were paired with picture stimuli with high levels of naming agreement for the target word (i.e. from the Newcastle Aphasia Therapy Resources; Webster, Morris, Whitworth & Howard, 2009; with additional pictures drawn by an illustrator).

The 24 noun stimuli were drawn equally from the six categories of *clothes*, *fruit*, *furniture*, *tools*, *transport*, and *vegetables*. All targets were rated as high-typicality members of their respective categories, according to the Hampton & Gardiner (1983) norms, and had high production frequencies (see Chapter two). As with verb stimuli, 96 nouns were selected to act as fillers. Fillers were selected so that they did not share any category relations with the experimental stimuli but some filler stimuli did share category and/or associative relations (e.g. *arm, leg*). Nouns were paired with pictures drawn from the 'Snodgrass and Vanderwart-like' set[2] of coloured materials (Rossion & Pourtois, 2004). A complete list of experimental stimuli is provided in Appendix R.

Design and procedure. Each word class condition (noun, verb) was presented to each participant with one of two counterbalanced presentation lists (i.e. List A and List B). Presentation lists were counterbalanced for relatedness so that where a target was

[2] Images courtesy of Michael J. Tarr, Brown University. http://www.tarrlab.org/

paired with a related prime in List A, the same target was paired with an unrelated prime in List B. Participants only named target pictures in one of the two presentations lists per word class condition.

Each presentation list consisted of 120 prime-target pairs and each presentation list was divided between two 60 item presentation blocks. Participants were able to take a short break between blocks. Each block included six related prime-target experimental trials (one target per semantic category) and six unrelated prime-target experimental trials (one target per semantic category). The remaining 48 trials were filler prime-target pairs. For both experimental items and fillers, where an item appeared as a prime in one presentation block, it subsequently appeared as a target in the other presentation block. Filler primes and targets were randomly paired for each presentation block and for each experimental list.

Each participant completed one presentation list per word class and the order of word class presentation lists (i.e. noun-verb, and verb-noun) was counterbalanced between participants. Between word class presentation lists, participants completed an unrelated task, which lasted approximately two minutes, in an attempt to minimise any carry-over effects.

Experimental prime-target pairs were presented with between three and five intervening filler prime-target pairs and each presentation block began with at least three filler prime-target pairs. Given this condition, the order of experimental and filler prime-target pairs was random within each presentation block.

Before participants began each experimental list (i.e. noun and verb lists), they completed two practice blocks of 10 prime-target pairs. Neither the words nor the picture stimuli used in the practice blocks were used within the experimental presentation lists. Words used as primes within the first practice blocks were subsequently used as targets in the second practice block and vice-versa, although primes and targets were randomly paired for each practice block. Any necessary feedback was given to participants following completion of each practice block.

Participants were seated in front of a computer display with a microphone placed approximately 12 inches from their mouth. The threshold for voice trigger activation was individually calibrated for each participant. The experiment was presented using DMDX software on a flat screen display running display settings of 1280, 1024, 1024, 32, 50.

One experimental trial consisted of a blank (white) display for 2000 msecs followed by the presentation of a fixation cross '+' in the centre of the display for 100 msecs. This was followed by a pattern mask of '########' for 1000 msecs. The pattern mask was replaced for 30 msecs by the prime which was subsequently backward masked with '########' for a further 100 msecs. The target picture was then presented in the centre of the display for a fixed duration for 3000 msecs. The voice trigger was activated on presentation of the target picture and clocked the time to the onset of the participant's verbal response. The experiment proceeded automatically and immediately onto the next trial. All text was presented in black 32 point Arial font. Verb pictures were presented as black and white line drawings and noun target pictures were presented as colour shaded drawings.

Participants were verbally instructed that they were going to see pictures on the display and that they should name the pictures as quickly and accurately as possible. Within the noun condition, they were told that the pictures would be of various objects and that they were to name the object shown. In the verb condition, participants were told they would see pictures of various actions and that they should name the action using one word.

Results

Data exclusion and replacements. Before data were analysed, a subset of response timings were checked using PRAAT speech analysis software (see Boersma, 2001). These included all items which received timings of less than 500 msecs and all responses which were recorded as timeouts (i.e. > 3000 msecs). Errors in timing were adjusted as appropriate. Following this, all non-target responses and responses which were greater than 3000 msecs were automatically excluded. Subsequently, all responses which were greater than two standard deviations from each participant's grand mean were excluded. In total, the data exclusion procedure lead to the exclusion of 121 noun condition responses (10.08% of total) and 268 verb condition responses (22.33% of total). Any experimental items which were excluded were replaced with the participant's mean obtained from all other experimental items. Data replacement was required for 15 noun condition responses (6.25% of experimental items) and 65 verb condition responses (27.08% of experimental items). Only correct experimental responses and replaced experimental values were subjected to reaction time analysis.

Response time analysis. Mean reaction times by condition are presented in Table 4.11. Reaction time data were analysed with a two-way within-participants ANOVA with factors of word class (noun, verb) and relatedness (related, unrelated). Analysis by participants produced a significant main effect of word class (F_1 (1,9) = 10.98, p = .009, η_p^2 = .550) but no significant main effect of relatedness (F_1 (1,9) = 3.03, p = .116, η_p^2 = .252) nor a significant word class by relatedness interaction (F_1 (1,9) = 0.04, p = .84, η_p^2 = .005). Analysis by items produced a significant main effect of word class (F_2 (1,23) = 38.94, p <.001, η_p^2 = .629) but no significant main effect of relatedness (F_2 (1,23) = 1.45, p = .240, η_p^2 = .059) nor a significant word class by relatedness interaction (F_2 (1,23) = 0.65, p = .430, η_p^2 = .027).

Differences between relatedness conditions were also investigated within each word class individually using paired samples t-tests (one-way). By participants, neither nouns (t(9) = 1.108, p = .297, d = 0.247) nor verbs (t(9) = 0.909, p = .387, d = 0.164) showed significant differences between means in related and unrelated conditions. By items, neither nouns (t(23) = 1.079, p = .292, d = 0.287) nor verbs (t(23) = 0.416, p = .681, d = 0.036) showed significant conditions between means in related and unrelated conditions.

	Nouns		Verbs	
Related	874	± 113	1096	± 171
Unrelated	836	± 192	1069	± 160
Difference	38		27	

Table 4.11 Mean response time (msecs) by word class and relatedness (by participants)

4.4.4. Category superordinate semantic priming

Introduction

This second semantically primed picture naming experiment investigated the effects of category superordinate word primes in the production of verbs and nouns elicited through pictures of actions and objects respectively. Previous research using word naming (i.e. reading aloud) may well lead to the prediction that noun targets should be facilitated by related superordinate primes (e.g. Keefe & Neely, 1990), however, there is also considerable research using picture-word-interference tasks that have found inhibited naming following superordinate primes (e.g. Kuipers et al, 2006; Vitkovitch & Tyrrell, 1999). For verbs, predictions based on previous research are difficult to make as previous research has not directly investigated effects of superordinate primes on verb processing. This condition therefore, provides a crucial test of semantic representations and organisation. If organisational principles are comparable between verbs and nouns, then this may be evidenced by comparable patterns of priming (e.g. superordinate verb primes lead to facilitated target verb production). In contrast, if organisational principles differ between verbs and nouns then it may be predicted that superordinate verb primes may lead to comparable effects as coordinate verb primes (i.e. inhibition) and a dissociation with effects of superordinate noun primes on target noun processing.

Method

Participants. Ten participants took part in the study (*M* age = 20.8 years; *SD* = 1.8; *Range* = 19-24). This sample included 5 males and 5 males and all participants were right handed and had normal or corrected vision and gave written consent.

Stimuli selection. The experimental stimuli acting as experimental targets were the same as those used in the category coordinate naming experiment. Within the related condition, verb targets were paired with primes which were congruous superordinate level verbs (i.e. *breaking*, *cleaning*, *cooking*, *cutting*, *hitting*, and *making*). In the unrelated condition, targets were paired with unrelated verbs assumed to be at similar superordinate levels of specificity (i.e. *jumping*, *moving*, *putting*, *running*, *talking*, or *walking*). In the noun word class condition, targets were also paired with appropriate superordinate level nouns both in the related condition (i.e. *clothes*, *furniture*, *fruit*, *tools*, *transport*, or *vegetables*) and the unrelated condition (i.e. *bird*, *fish*, *flower*, *music*, *sport*, or *weapons*). A complete list of experimental stimuli is provided in Appendix R.

As there were only six different noun and verb primes used within each relatedness condition, each of these primes appeared twice in each presentation list and each was presented once in each presentation block.

Design and procedure. The design and procedure was the same as previously used in the category coordinate naming experiment. The only methodological difference was the selection of experimental prime-target pairs. All filler stimuli and practice prime-target pairs were also the same as those used previously.

Results

Data exclusion and replacements. Data were excluded and replaced according to the same procedures used previously. This lead to the exclusion of 142 responses within the noun condition (11.83% of total) and 293 responses within the verb condition (24.42% of total). Data replacements were required for 20 noun experimental condition responses (8.33% of experimental items) and 84 verb experimental condition responses (35% of experimental items).

Response time analysis. Mean reaction times by condition are presented in Table 4.12. Reaction time data were analysed with a two-way within-participants ANOVA with factors of word class (noun, verb) and relatedness (related, unrelated). Analysis by participants produced significant main effects of word class (F_1 (1,9) = 66.58, p < .001, η_p^2 = .881) and relatedness (F_1 (1,9) = 5.98, p = .037, η_p^2 = .399) but no significant word class by relatedness interaction (F_1 (1,9) = 0.59, p = .461, η_p^2 = .062). Analysis by items produced a significant effect of word class (F_2 (1,23) = 90.66, p < .001, η_p^2 = .798) but no significant main effect of relatedness (F_2 (1,23) = 2.02, p = .169, η_p^2 = .081) and no significant word class by relatedness interaction (F_2 (1,23) = 0.23, p = .638, η_p^2 = .01).

Differences between relatedness conditions were also investigated within each word class individually using paired samples t-tests (one-way). By participants, neither nouns (t(9) = -1.312, p = .222, d = 0.224) nor verbs (t(9) = -1.967, p = .081, d = 0.339) showed significant differences between means in related and unrelated conditions. By items, neither nouns (t(23) = -0.938, p = .358, d = 0.242) nor verbs (t(23) = -0.995, p = .330, d = 0.248) showed significant conditions between means in related and unrelated conditions.

	Nouns		Verbs	
Related	821	±112	1059	±141
Unrelated	847	±124	1112	±173
Difference	26		53	

Table 4.12 Mean response time (msecs) by word class and relatedness (by participants)

4.4.5. Semantic Priming: Combined Analysis

Introduction

Given that the two previous experiments had suggested that coordinate primes inhibit picture naming regardless of word class and that superordinate primes facilitate picture naming, also regardless of word class, the data from both experiments were combined in order to investigate the possibility of a dissociation between prime type (coordinate versus superordinate) and relatedness condition (related versus unrelated).

Method

The data from the category coordinate and category subordinate naming experiments were combined and reanalysed within a mixed ANOVA design with one between participant factor of prime type (coordinate, superordinate), and two within participant factors of word class (noun, verb) and relatedness (related, unrelated).

Results

Overall, there was a significant main effect of word class (F_1 (1,18) = 40.56, $p <$ .001, η_p^2 =.693) and no significant main effects of relatedness (F_1 (1,18) = 0.09, $p <$ .768, η_p^2 =.005) or prime type (F_1 (1,18) = 0.028, p = .868, η_p^2 = 002). The statistic of interest was the interaction between prime type and relatedness which produced a significant interaction (F_1 (1,18) = 8.53, p = .009, η_p^2 = .332) and this interaction is presented in Figure 4.1. Targets in related coordinate conditions were, on average, named 32 msecs slower than items in the respective unrelated conditions. In comparison, targets in related superordinate conditions were named, on average, 40 msecs faster than items in respective unrelated conditions.

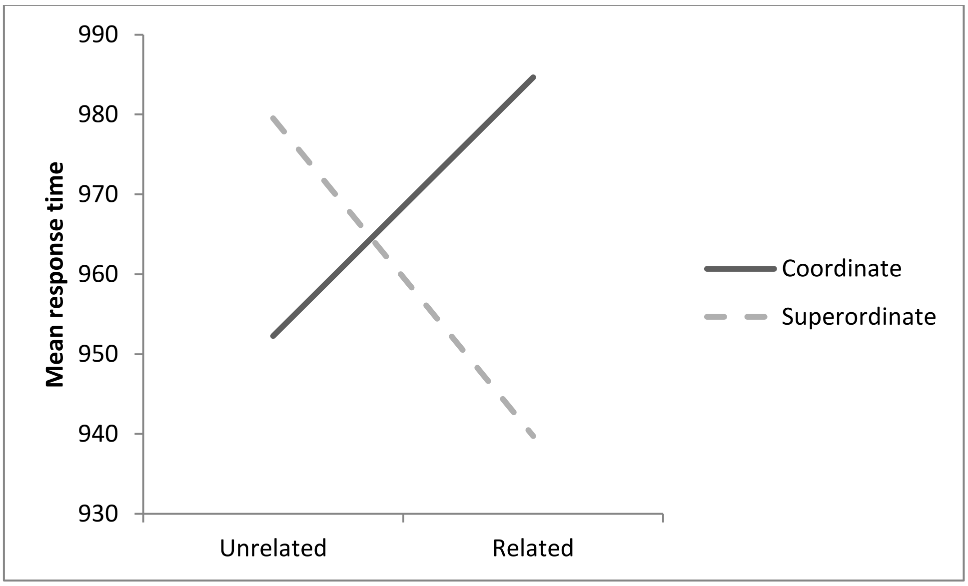

Figure 4.1 Interaction of Prime type and Relatedness (by participants)

4.4.6. Discussion of semantic priming experiments

Summary of main findings

Within the category coordinate priming experiment, participants responded significantly faster to nouns than to verbs. Trends in performance were apparent that did not reach significance, a factor likely to be contributed to by the sample size, but which are worthy of reporting. Noun targets were produced on average 38 msecs slower when primed by a related category coordinate although this was not a significant difference. Verb targets were produced on average 27 msecs slower when primed by a related category coordinate but again this was not significant.

Within the superordinate priming experiment, participants were again faster to respond to nouns than verbs. There was an overall significant effect of relatedness with target items being responded to significantly faster when primed with a related superordinate compared to when primed by an unrelated superordinate. When each word class was considered individually, nouns were responded to an average 26 msecs faster when primed by a related superordinate although this was not significant. Verbs were responded to an average 53 msecs faster when primed by a related superordinate but again this failed to reach significance.

The combined analysis of both priming experiments emphasised the dissociation between prime type (i.e. coordinate and superordinate) and relatedness condition (related and unrelated) that the experimental conditions, which attempted to tap automatic semantic processing, gave rise to. This dissociation was present regardless of the different word classes that were being investigated and this provides evidence for comparable semantic processing operations underlying the processing of verbs and nouns within these experimental conditions.

Discussion of main findings

Noun production was inhibited by related coordinate noun primes. The finding that noun production was inhibited by related coordinates was predicted on the basis of previous research with similar findings within a primed picture naming task using a short SOA (i.e. Alario et al, 2000). Alario et al (2000) observed an average 33 msec inhibition effect when primes were related coordinate nouns. This is comparable to the 38 msec effect found here although the current effect was not found to be significant. This was most likely to be due to insufficient participant numbers (i.e. 10 compared to 20 used by Alario et al, 2000). However, a power calculation derived using GPower 3.1

(see Faul, Erdfelder, Lang & Buchner, 2007) indicated that a sample size of 90 participants would be required to ensure the current difference was significant ($p < .05$).

Whilst noun coordinates have also been found to facilitate naming of related targets (e.g. Hines et al, 1986; Lupker, 1988), the methods were not directly comparable as prime and target have been presented in the same modality rather than cross-modality as is the case in the current experiment and with Alario et al (2000). However, Hines et al (1986) did find facilitation of word reading aloud when related word primes were masked and presented below levels of conscious awareness and thus when effects may be attributable to automatic semantic priming.

Verb production was inhibited by related coordinate verb primes. The finding that verb production was inhibited by related coordinates was predicted based on previous research that has employed picture-word interference tasks. The mean 27 msec effect observed here is again comparable to previous effects of Collina & Tabossi (2007; M = 32 msec effect) and Roelofs (1993; *M range* = 22-25 msec). However, having previously highlighted that such methods with nouns produce results that contrast to those found here, the validity of comparing to picture-word interference tasks here is debateable. However, no studies appear to have employed picture naming when previously primed by a written word as opposed to simultaneous presentation of both stimuli. The inhibition effect of related verb coordinates observed here was again not significant and power calculations indicated that a sample of 133 participants would have been required for this difference to be significant ($p < .05$). Therefore, in isolation, the current results do not present a convincing argument for a genuine effect.

It has been found that priming effects tend to increase as response times increases (e.g. Hines et al, 1986). So, as nouns were responded to faster than verbs, if priming effects were present for nouns, a parallel effect might be expected to be greater for verbs. However, it would need to be ascertained as to the reason why verbs are responded to slower than nouns. It is likely that interpretation of action pictures imposes greater processing demands than interpretation of object pictures and so a simple comparison of response times and priming effect sizes is unlikely to be valid.

Despite the lack of statistical support, the descriptive data do offer intriguing differences between related and unrelated prime conditions, especially as the differences occur in the same direction as with priming with noun coordinates (i.e. related primes lead to slower naming of target items).

Noun production was facilitated by related superordinate noun primes. The finding that noun production was facilitated by related superordinates was predicted based on previous research that used a method where participants read targets aloud following exposure to a written prime (i.e. Keefe & Neely, 1990). The 26 msec facilitation effect observed here is comparable to the 32 msec effect found by Keefe and Neely (1990) but was again not significant in the current experiment. A power calculation determined that a total sample of 65 participants would be required to allow the observed difference to reach significance ($p < .05$). However, when ANOVA analysis combined data with nouns and verbs, there was an overall significant effect of relatedness ($p = .037$).

The current finding contrasts with previous research that has investigated cross-level priming between the basic and subordinate levels of noun categorisation. Vitkovitch & Tyrrell (1999) found that related primes at differing levels of categorisation to the target inhibited naming of the target within a picture-word interference task.

Verb production was facilitated by related superordinate verb primes. There was no previous research that allowed a direct prediction of the finding that verb production was facilitated by related superordinates. Roelofs (1993) did investigate cross-level priming of verb targets with related subordinate verbs (e.g. *gorge-DRINK*) within a picture-word interference task. No effects were found in one experiment while facilitation effects were found within a second experiment. However, this can still not serve as a reliable comparator as Roelofs (1993) used primes at a lower level of categorisation to the target which was the reverse to the current experiment.

The facilitation effect of 53 msecs again did not reach significance, although it was the individual comparison that came closest to reaching significance (p = .081) and again, when combined with noun data within ANOVA analysis, this led to the overall significant effect of relatedness for superordinate primes (p = .037). A power calculation indicated that a total sample of 30 participants would have been required in order to make this difference significant.

There was a dissociation between prime type and relatedness. The observed dissociation between prime type (i.e. coordinate, superordinate) and relatedness of prime validates a claim for the presence of priming effects in the current study. This result suggests that regardless of word class, related coordinate primes lead to inhibited picture naming whereas related superordinate primes lead to facilitated picture naming.

Limitations[3] *and further research*

The number of participants who took part in each experiment was generally lower than the number used in the majority of priming studies although is comparable to both Damian et al (2001) who used 10 participants in one of their two experiments and Roelofs (1993) who used 9 participants in two experiments with both studies reporting significant effects of priming. As indicated by power calculations, the number of participants was relatively small in the current experiments and insufficient to lead to significant effects, however the majority of calculations suggested that the number of participants needed would greatly exceed the number used in previous reports. Only the calculation associated with verb superordinate primes, which suggested 30 participants, was comparable with previous studies. There was a large amount of variation in mean response time between participants which may have impacted on group level analyses by increasing the variance in the response time data. A potential solution to this would be to exclude the data from participants with the slowest and fastest mean response times in order to account for potential outlying responses. However, a more robust solution would be to include more participants in order to conduct separate analyses of participant with ‘slow’ mean responses and participants with ‘fast’ mean responses (e.g. Hines et al, 1986). This would potentially reduce variance between participants’ response times within these groups and would also allow further clarification on whether the size of facilitation/inhibition effects between word classes is equivalent, i.e. whether the *d*-value is equal for effects associated with verbs and nouns for participants who are generally fast responders and for participants who are generally slower responders.

[3] A limitation not discussed here is the need to collect information regarding participants’ awareness of the presence of prime stimuli. Although primes were presented for a duration which was hypothesised to be below conscious awareness for the majority of participants (i.e. 30 msecs), some participants did report seeing words and ‘flashes’ at some time throughout the experiment once the experiment was concluded. Such data was informally collected from all participants (i.e. did you notice anything unusual about the experiment; did you notice a flash; did you notice any words; do you remember any of the words?), unfortunately this data was lost during the data analysis stages and was not recoverable. Further research may look to record and report such information and compare the results of participants who were aware of primes (at some point during the experiment) to those who report no awareness of the presence of prime stimuli. Alternatively, pre-testing for each participant may be conducted to determine the duration at which primes may be presented in order to ensure they are below the level of conscious awareness. This may be lead to the prime duration varying between participants but this should not pose confounding issues for further analysis and interpretation.

As with category verification, the verb data in primed picture naming was subject to a large number of replacements for error responses. This has implications for the data analysis in that a large proportion of response times may not have accurately reflected true values as they were representative of the mean response time to all experimental stimuli. While data replacements were fairly evenly spread across experimental conditions, it would be unknown as to what effect this may ultimately have had on data analysis and therefore interpretation of results. Therefore, future work should aim to reduce the number of data exclusions as much as possible so that levels are comparable to other priming research (e.g. $\leq$ 5% of total responses being errors due to incorrect naming responses) The high number of exclusions for verb naming was partly anticipated given the inherent difficulty and variability in naming pictures of actions compared to naming pictures of objects. In the current experiment, any response that was non-target was excluded, so perhaps, future experiments may look to only exclude responses that are non-target and non-acceptable alternatives in an effort to reduce the potential effects that replacing errors with a mean value would have. While the picture stimuli used in the current experiments had generally high levels of naming agreement, these data were gathered in an offline context (i.e. with less time pressure). Therefore, further data regarding non-target acceptable responses, within an online context, may be identified before conducting the experiments in order to avoid possible biases. This could be obtained in separate naming agreement phases where control participants name the stimuli pictures in online conditions without the presence of prime stimuli (the naming agreement data used to select stimuli was conducted in an offline context with no time pressure). While it may have been a possibility to identify non-target yet acceptable responses based on experimental participants' responses, where participants consistently named a picture with a non-target this may reflect inadequacies of the picture stimuli to elicit the target within an online experimental context, or, alternatively it may reflect influences of prime stimuli which may lead to the production of non-target responses. A further methodological consideration to reduce errors may be to conduct a pre-training phase in the experiment where participants are exposed to picture stimuli alongside their target responses for participants to either repeat or read aloud.

One further issue that should be considered in further research is attention to other psycholinguistic variables. Within the current experiments, stimuli selection, particularly for verbs, was based on criteria of having a relatively high association strength and high typicality value within the respective category but also on being able to represent the referent action within a nameable picture. This severely restricted the possible stimuli that could be used and therefore, no particular attention was paid to controlling other variables which may influence performance within online tasks (e.g. lexical frequency, familiarity, imageability, and so on). Ideally, such variables would be controlled in such experiments, especially those which may impact on semantic processing (e.g. imageability) although, given that the experiment was conducive to tapping automatic semantic priming, the influence of variables which are less likely to be semantic in nature (e.g. lexical frequency) should hopefully be minimised.

There are currently very few studies that have looked to directly compare the effects of priming between word classes, with Rösler et al (2001) providing one of the few examples but who intriguingly found parallels in performance between word class conditions within a lexical decision task. Considering that the current experiments have highlighted potential similarities in the processing of verbs and nouns within a specific semantic priming task, this would suggest further investigation is warranted to compare the influence of semantic priming on verb and noun processing within other priming tasks (e.g. picture-word interference, lexical decision, semantic decision) when similar principles are applied to the selection of stimuli between word classes. However, given that the effects observed here are far from conclusive this also suggests that further investigation using primed picture naming is also warranted.

4.5. Conclusions

The two online experimental tasks reported in this chapter highlight the usefulness of applying the same experimental techniques to compare the retrieval of verbs and the retrieval of nouns in healthy speakers. The results revealed contrasting patterns between the two tasks with verb and noun retrieval being influenced by different variables within category verification but with semantically primed picture naming showing that both verb and noun retrieval can potentially be both facilitated or inhibited depending on the semantic relatedness of stimuli that has been processed immediately prior.

Such findings from healthy speakers are insightful from a theoretical perspective in relation to models of semantic memory and language processing in general. The parallel findings with regard to verbs and nouns and the differing directions of priming effects associated with differential categorical relations in particular demonstrates the potential that concurrent activation of verbs through different modalities may be able to cause inhibition effects (when the two items may conflict and compete for selection to describe pictures) and also facilitation effects (where the two items overlap and are not in conflict). However, what may potentially give rise to difficulties with such an interpretation is that for verbs, there may not be a clear candidate to name the pictures used (as indicated by greater errors/non-target responses). This may imply that such experiments with verbs (including unprimed confrontation picture naming) are naturally more susceptible to competition effects. However, an interpretation of competieition effects for verbs would not be inconsistent given the previous literature using picture-word interference tasks where inhibition effects have also been observed when primes and targets have shared a semantic categorical relation (e.g. Collina & Tabossi, 2007; Roelofs, 1993; Schnur, et al, 2002; Tabossi & Collina, 2002).

Having considered how verbs and nouns are retrieved by healthy speakers in offline and online psycholinguistic tasks, the following chapter will move on to consider verb and noun retrieval within speakers with language impairment, specifically within an intervention study that aims to improve verb and noun retrieval. On the basis of experiments thus far, in particular those reported in the current chapter, it may be predicted that speakers with language impairment may show comparable effects with improvements in verb retrieval and noun retrieval following intervention. This would be so on the basis that while healthy speakers have generally had more difficulty in completing tasks with verbs than with nouns (as indicated by slower response times and more errors, etc.), similar patterns of behaviour arise which may be assumed to reflect that similar processes are in operation. Therefore a key issue is whether impairment to normal language processing leads to a continuance in these similar patterns of behaviour between verb and noun processing, or whether disruption leads to different observable patterns, or whether investigation of speakers with language impairment leads to a reinterpretation of how the observable similar pattens arise (i.e. whether they are only superficially similar and are accounted for by different processing mechanisms).

Chapter 5 An Intervention Study to Improve Retrieval of Verbs and Nouns in Speakers with Aphasia

5.1. Aims of Chapter

So far, this thesis has investigated the semantic representations of verbs through the use of offline and online behavioural experiments with healthy adult speakers of English and has sought to compare these findings with those into the semantic representations of nouns. However, as Chapter one highlighted, a great deal of current understanding with regard to semantic and language processing has come from observation and testing of psycholinguistic behaviours with speakers with impaired semantic and language processing abilities. This chapter builds on this line of investigation of verbs' semantic representations by reporting the results of an intervention study for participants with acquired language impairment (i.e. aphasia) as a consequence of cerebrovascular accident (CVA). To continue the theme of the current thesis, the intervention study sought to compare the effectiveness of therapy when targeting improved verb retrieval to when targeting improved noun retrieval. By making these comparisons it allowed further insight into whether and how action/verb and object/noun semantic representations are similar and also how they may differ.

Unlike psycholinguistic investigations with healthy speakers, in the field of aphasiology, there exists an extensive research literature investigating impairments associated with verb processing (e.g. Berndt, Mitchum, Haendiges & Sandson, 1997; Breedin & Martin, 1996; Breedin, Saffran & Schwartz, 1998). There have also been direct comparisons between impairments of verb processing and impairments of noun processing (e.g. Bird et al, 2003; Caramazza & Hillis, 1991; Zingeser & Berndt, 1988). However, for the most part, this extensive literature on the nature of verb impairments has yet to be applied within appropriate theory that can inform effective therapies for the remediation of verb impairments (see Conroy, Sage & Lambon Ralph, 2006, for discussion). Impairments in verb processing can be particularly detrimental to an individual's ability to communicate through language given verbs' central role in understanding and producing sentences (e.g. Berndt, Haendiges, et al, 1997). There is currently a growing body of literature reporting intervention studies aiming to remediate verb retrieval. Hence, this chapter aims to compare the effectiveness of a semantic-based therapy for the remediation of verb and noun retrieval. By comparing the effectiveness of the same therapy, it is argued that this can inform theoretical discussion as to the status of the semantic representations of objects/nouns and actions/verbs. So far, this thesis has demonstrated a number of areas where objects/nouns and actions/verbs show similarities, leading to the hypothesis that a single therapy approach should be equally effective in remediation of impairments affecting noun and verb processing. However, there have also been indications in the current thesis that verb processing is more difficult, or more demanding than noun processing, for example, with more errors made and slower response times.

This chapter begins with a discussion on the relationship between theories of semantic and language processing at a single-word level (as opposed to sentence level processing) and the investigation of speakers with language impairments. This discussion will draw out how these strands are mutually informative within the field of cognitive-neuropsychology. More specific discussion will then focus on the investigation of verb retrieval and its disruption in speakers with language impairment and how these investigations have informed both theory and therapy. The chapter will then describe in more detail what ‘semantic therapies’ are, followed by a review of three broad types of semantic therapy. This will include description of their therapy protocol, discussion of the underlying assumptions in terms of what processing components each targets, and a review of significant findings of these therapies when they have been applied with speakers with language impairment. This explicitly focuses on comparing findings of similar therapy approaches when aiming to improve noun retrieval and those aiming to improve verb retrieval. This literature review is followed by a description of the methods and results of an intervention study for five participants with aphasia. The intervention studies utilised a cross-over design whereby the same therapy approach was applied to nouns and verbs in discrete phases with the order of phases being counterbalanced between participants. The main findings will then be considered in relation to previous research and their implications will be discussed.

5.2. Background

5.2.1. Theory informing therapy; therapy informing theory

The development of cognitive neuropsychological models based on theories of semantic and language processing has been a useful asset in the investigation of language impairment (see Basso & Marangolo, 2000; Hillis, 1998; Nickels & Best, 1996, for discussion). By observing where individuals have difficulty in language tasks and by recognising the types of errors they make, cognitive neuropsychological models allow the identification of specific loci of breakdown. For example, impaired performance in comprehension based tasks across different modalities (e.g. spoken-word-to-picture matching and written-word-to-picture matching) may be an indicator of impairment within the semantic system as a unitary semantic system is assumed to underlie all input modalities within the cognitive neuropsychological model of language processing (i.e. Patterson & Shewell, 1987; Whitworth et al, 2005: see Figure 5.1). Despite the usefulness of cognitive neuropsychological models to the assessment of language impairments, it has also been suggested that such models are underspecified in terms of informing how to actually remedy language impairments in therapeutic contexts (e.g. see Howard & Hatfield, 1987, for discussion). Stated simply, cognitive neuropsychological models can allow interpretation of observed patterns of impaired performance across language-based tasks, however, they currently cannot specify which tasks must be employed within intervention to improve this impaired performance: “a theory of therapy remains to be developed” (Nickels & Best, 1996:22). Despite this, there appears to be an assumption that performing tasks that can target specific levels of breakdown is also beneficial and can act to remediate some of the impairment. This remediation may occur through reactivation or relearning of impaired representations or systems, or reorganisation so that intact representations or systems may take over the processing previously associated with now damaged representations or process (see Nickels & Best, 1996).

Just as cognitive neuropsychological models of language processing have developed, in part, from observations of language impairment, outcomes of intervention studies can (re)inform the theory of these models. For example, therefore, if actual outcomes of therapy are consistent with the predicted outcomes, a researcher or clinician may draw the conclusion that the intervention affected change in the hypothesised manner and that the model or theory on which the intervention was based is sound. If actual outcomes do not match with the predicted outcomes, this should lead the researcher or clinician to re-evaluate their hypothesis regarding either: (1) the participant’s level/s of impairment; (2) the demands of the task (e.g. maybe it was not tapping the behaviour expected); or (3) the model or theory underpinning the task.

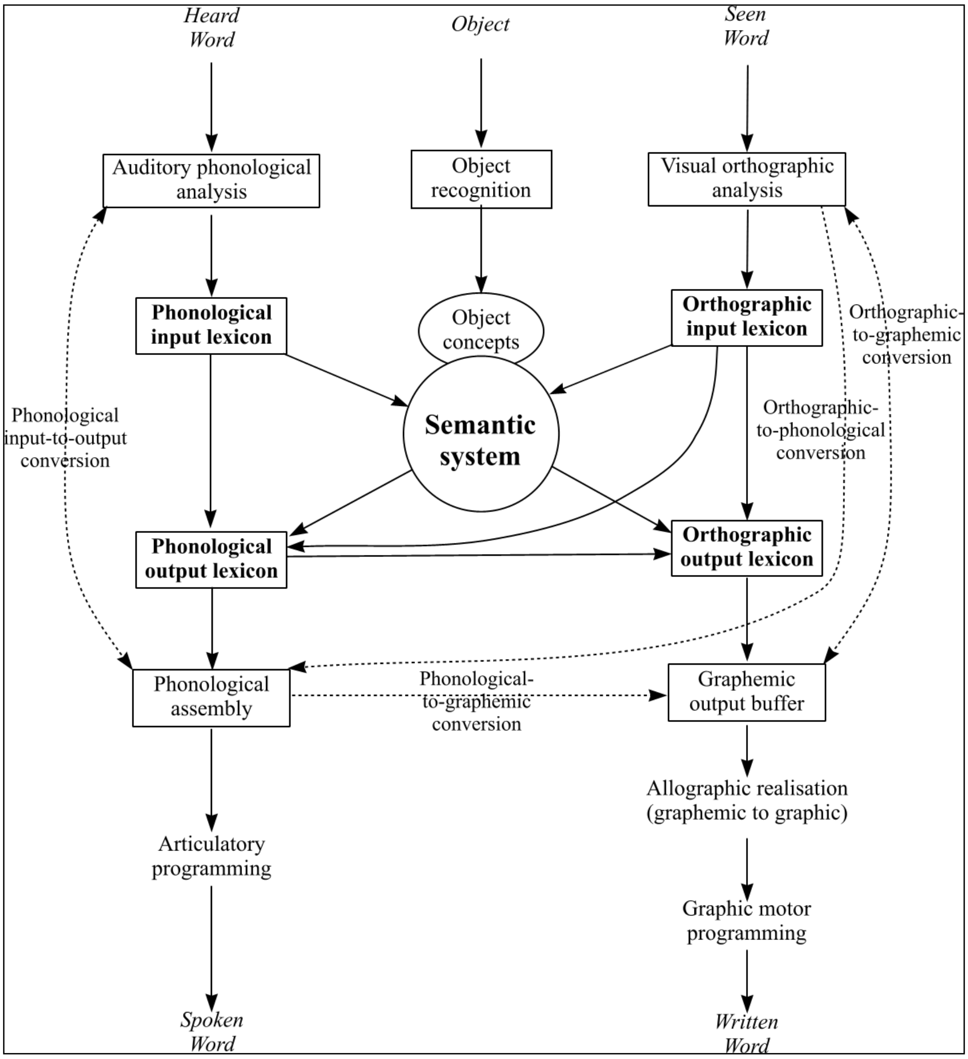

Figure 5.1 Cognitive neuropsychological model of language processing (from Whitworth et al, 2005)

Interventions for language impairments should be ‘model-guided’ (see Horton & Byng, 2002, for further discussion) so that therapy tasks should attempt to target hypothesised areas of language breakdown. As such, if a therapy approach affects change in the way it is assumed to, the results of therapy should in some sense be predictable. Naturally, various other factors need to be considered when selecting therapy tasks, not least a participant’s motivation to undergo therapy and their ability to undergo intervention (e.g. other health and/or cognitive issues). However, assuming that a participant is able to undergo therapy and there is a clearly hypothesised level of impairment, a therapy protocol can be developed whereby outcomes can assess the therapy’s effectiveness in bringing about the expected change in particular language behaviours.

When considering outcomes in therapies aiming to remediate word retrieval difficulties, outcome measures generally consist of assessment of naming performance, most usually from picture stimuli. Within this task, there are a number of ways of monitoring changes to language performance. Firstly, improvement can be measured in the items that have been explicitly used within the therapy protocol (i.e. the treated items). Secondly, improvements can be measured to items that have not been used within the therapy protocol (i.e. untreated). Untreated items may then be further subdivided into those that share some kind of relation to those items that have been treated and those that share no relation to those items that have been treated. This three-way distinction between item sets can form the basis of predictions of improvement following intervention. For example, if conducting a phonological therapy for participants with impairment in phonological assembly, it may be predicted that improvements would be observed in treated words and also words that share similar phonemes to the words that were treated. If conducting a semantic-based therapy for participants within impairment to semantic representations, it may be predicted that treated items would improve in addition to semantically related items on the basis that they share a certain proportion of semantic representational encoding (e.g. see Plaut, 1996, for discussion). If it is predicted that intervention may lead to participants adopting a conscious strategy to assist word-retrieval, then it may also be predicted that improvements may be observed in untreated items that share no relation to treated items (e.g. by promoting a self-cueing strategy; DeDe, Parris & Waters, 2000).

The majority of cognitive neuropsychological models and theories of single-word processing have been developed from observations and investigations of noun processing, whereby breakdown at any component or access route between components (see Figure 5.1) may lead to predictable patterns of behaviour. For example, breakdown early on in language processing (i.e. pre-semantic or at the level of semantics or in access from semantics) may result in semantic-type error behaviours. This is in comparison to breakdown at later or post-semantic levels which may give rise to phonological-type errors (see Whitworth et al, 2005, for a comprehensive account of characteristic patterns of behaviour following breakdown of differing processing components). In comparison, the processing of verbs has almost exclusively been considered in relation to cognitive neuropsychological models of sentence processing (e.g. Garrett, 1982). It is not contentious that verbs are central to sentence processing and they are appropriately considered in relation to such models of sentence processing. However, investigation into language impairment frequently considers how speakers retrieve verbs at the level of single words (i.e. in isolation from explicitly realised sentence contexts). This is often achieved, as with nouns, through picture naming tasks where participants are asked to name pictures of actions using a single word. In addition, there have been reports that speakers with language impairments may show similar error patterns with verbs as they do with nouns (e.g. Marshall, Pring, Chiat & Robson, 1996; Marshall, Chiat, Robson & Pring, 1996). Further to this, there have been numerous reports of intervention studies that have aimed to remediate processing of verbs at a single word level, i.e. the aim of therapy being to produce a single verb elicited through picture naming. While there are numerous reports such as these looking into the nature and remediation of verb impairments, it is unclear as to how such investigations fit with cognitive neuropsychological models of single word processing (e.g. Patterson & Shewell, 1987; Whitworth et al, 2005) which have been developed to account for processing of nouns. The following section provides an overview of the recent findings with regard to verb investigations in the field of language impairments.

5.2.2. Verb processing in aphasia

Poorer performance (e.g. slower processing and/or more error responses) with verbs relative to nouns is a pattern observed in tasks with healthy speakers (e.g. Almor et al, 2009; Kohn, Lorch & Pearson, 1989; Mätzig, Druks, Masterson & Vigliocco, 2009; Ramsay, Nicholas, Au, Obler & Albert, 1999). This therefore suggests that there is something particular about verb processing that makes this intrinsically more complex than noun processing. Further to this, research into populations with language impairments, has often shown that verbs can be selectively impaired beyond what may be expected based on overall word-retrieval difficulties (e.g. Caramazza & Hillis, 1991; McCarthy & Warrington, 1985; Miceli et al, 1984; Williams & Canter, 1987). It is important to highlight that the reverse pattern whereby nouns are impaired to a greater extent than verbs has also observed (e.g. Bi et al, 2007; Zingeser & Berndt, 1988) although a review by Mätzig et al (2009) found that 75% of reported cases of speakers with language impairment showed a selective verb retrieval deficit. In comparison to many of the areas discussed throughout this thesis, there has been substantial research into verb processing in the field of aphasiology. However, this has mostly focused on identifying how verb processing is disrupted and the factors that underlie such disruption. It has been claimed that the knowledge gained from these investigations has so far failed to be translated into effective interventions for the remediation of verb impairments (see Conroy et al, 2006, for discussion).

When specific verb impairments have been identified, these have been associated with deficits at various levels of single-word and sentence level processing. Breedin & Martin (1996) described how verb processing may be impaired in either one or any combination of: (a) comprehension or production of a verbs' action (i.e. its conceptual-semantic representation); (b) the thematic role information that the verb specifies; or (c) the subcategorisation frame information that the verb specifies (i.e. argument structure properties). Kemmerer & Tranel (2000) investigated stimulus (i.e. picture), lexical (i.e. verb being elicited) and conceptual (i.e. semantic and thematic representation) factors that are likely to influence ability to produce verbs when naming action pictures in a group of left-hemisphere brain-injured patients. Overall, they found that naming performance was influenced by factors including familiarity, image agreement, name agreement, whether the verb has a homophonous noun form, and whether the verb specifies that the undergoer of the action changes location. However, individual participants varied greatly in terms of which factors influenced their performance. These studies, and others, highlight the complex nature of verb impairments and the various factors, and combinations of factors that have potential to influence performance.

Various semantic factors have been argued to influence verb processing in speakers with language impairments. Barde et al (2006) and also Breedin et al (1998) have found that some speakers with language impairment tend have greater difficulty in processing semantically simple verbs, where complexity is considered in relation to verbs' semantic featural composition (i.e. in terms of Jackendoff's, 1983, interpretation of semantic features). Breedin et al (1998) found that six out of eight participants who showed a disproportionate impairment of verbs in relation to nouns had greater difficulty producing lighter and semantically simple verbs (e.g. *go*) compared to heavier, or more semantically complex counterparts (e.g. *run*). Barde et al (2006) extended this finding by demonstrating that a heavy/complex verb advantage was only present for participants with agrammatic-type aphasia and not for participants who had a non-agrammatic-type aphasia. This was argued to reflect the fact that agrammatic aphasia was a consequence of impaired grammatical and syntactic representational information upon which semantically light and simple verbs are more reliant. This is because they can appear in a wider variety of sentence constructions with a wider variety of thematic participants compared to semantically heavier and complex verbs which are used in describing a more restricted range of specific contexts.

Other work has highlighted that argument structure properties may influence ability to retrieve verbs. Kim & Thompson (2000) found that, across a group of seven participants with agrammatic aphasia, performance in categorising and naming verbs was influenced by the obligatory number of arguments that the verb was associated with. A hierarchy of difficulty was found such that participants performed best with one-argument verbs (e.g. *listen*) followed by two-argument verbs (e.g. *catch*) with worst performance on three-argument verbs (e.g. *give*). This was argued to demonstrate impaired access to lexical-syntactic representations of verbs at a lemma level of representation (i.e. Bock, 1995; Levelt et al, 1999). Collina, Marangolo & Tabossi (2001) extended the argument structure complexity hypothesis by investigating a single agrammatic participant's naming performance in one-argument and two-argument verbs and also argumental nouns (i.e. those that express relations between entities; e.g. *arrest*) and non-argumental nouns (i.e. the majority of nouns that do not express relations between entities; e.g. *medal*). The participant performed worse within verb naming and showed poorer performance with two-argument verbs compared to one-argument verbs. However, within noun naming, the participant performed worse with argumental nouns (73% of responses were errors) compared to non-argumental nouns (5% of responses were errors). This led Collina et al to conclude that argument complexity may be confounded with grammatical class but that each factor may produce independent effects. Therefore, depending on the level of impairment word class dissociations may be an artefact of argument complexity.

There has been a recent increase in the number of intervention studies being published that attempt to investigate how various factors may be incorporated into therapies to remediate verb retrieval. An important distinction that needs to be made however is between those therapies that attempt to remediate verb retrieval at the level of single words and those where verb retrieval is targeted within the context of sentences. Intervention studies targeting remediation of verb retrieval at single word level focus on verbal production of verbs in isolation, with outcomes most usually measured in terms of ability to produce verbs from pictures. These studies have generally shown that verb retrieval can be improved but that these effects, as with nouns, are usually limited to those verbs that have been employed within the therapy task itself (i.e. treated items; e.g. Boo & Rose, 2011; Edwards & Tucker, 2006; Fink, Martin, Schwartz, Saffran & Myers, 1992; Raymer & Ellsworth, 2002; Wambaugh, Cameron, Kalinyak-Fliszar, Nessler & Wright, 2004). Intervention studies targeting retrieval of single verbs have also shown limited improvement in sentence processing abilities except when sentence level impairments are associated specifically with difficulties in verb retrieval (e.g. Marshall, Pring & Chiat, 1998). Intervention studies targeting verb retrieval within sentences may aim to raise awareness of argument and thematic properties of verbs and integrate these explicitly within sentence contexts, i.e. verbal production of sentences. Such studies have also found that improvements are limited to verbs used in therapy but they do also suggest that improvements may also be observed in general word-retrieval (i.e. in noun retrieval) and also in sentence processing (e.g. Edmonds, Nadeau & Kiran, 2009; Kim, Adingono & Revoir, 2007; Raymer & Kohen, 2006; Webster & Gordon, 2009; Webster, Morris & Franklin, 2005). These intervention studies have employed a variety of differing protocols and approaches in attempts to improve verb, and in some cases, sentence processing abilities. Many of these studies used approaches that go beyond the scope of the current discussion, for example by incorporating the use of gesture (e.g. Rose & Sussmilch, 2009) or by focusing on written verb retrieval (e.g. Murray & Karcher, 2000); therefore, the following sections will firstly describe semantic therapy as a broad approach to remediation of word retrieval impairments at single word level (n.b. but see Conroy et al, 2006, for a broader review of verb therapy approaches). Following this, three specific examples of semantic therapy tasks will be discussed in more detail, including how they have been applied in attempts to improve noun retrieval and verb retrieval.

5.2.3. Semantic therapy

Therapy approaches that aim to remediate impaired processing within semantic memory can generally be termed semantic therapies. These may be operating to reactivate, reorganise, or re-teach semantic representations or links between semantic representations (e.g. Nickels & Best, 1996). The use of the term semantic therapy can sometimes be misleading however as some have claimed that *semantic therapy* is used as an umbrella term for therapy approaches that rely on semantic processing but which may not be necessarily aiming to remediate semantic processing itself (e.g. Horton & Byng, 2002; Nickels, 2000). Nickels (2000) identifies three broad types of task which may affect semantic processing: (1) tasks which aim to remediate word-finding difficulties regardless of whether the nature of the impairment is semantic or post-semantic (e.g. phonological); (2) tasks which do not explicitly aim to target semantic processing but where semantic processing may occur due to the nature of the task; and (3) tasks which aim to remediate semantic processing itself, for example, by enabling greater or improved access to semantic representations which can impact upon performance in tasks across modalities.

The impact of a particular task may also vary depending on the underlying impairment. If aiming to remediate semantic memory processes for individuals with hypothesised impairment within semantic memory, therapy should be theoretically motivated to affect change to either reactivate or reorganise the impairment (e.g. Best & Nickels, 2000; Horton & Byng, 2002; Plaut, 1996). If not model-guided, a semantic task may be arbitrarily selected based on an intuitive recognition that individuals may present with semantic difficulties. Task that involve elements of semantic processing involved in comprehension may include: sorting items into categories, identifying the odd-one-out from an array of items, matching words to pictures, identifying words from definitions, verifying attributes and semantic features of items. Tasks involving semantic processing involved in production may include producing words from picture stimuli (i.e. naming or picture description), or verbalising semantic information (e.g. providing definitions, listing attributes and semantic features). A further aspect of any of the above tasks will involve the use of cueing hierarchies and feedback from the clinician and/or researcher on participants' performance which may provide further reinforcement and facilitate the hypotheses reactivating and reorganising potential of the intervention. Therefore, given the variety of tasks which have potential to target semantic representations and processing, it is important to consider the specific demands of the task in relation to an individual's level of impairment when designing a therapy protocol.

Semantic therapy tasks differ from phonological therapy tasks which attempt to reactivate, reorganise, or re-teach phonological information (i.e. post-semantic processing). Here, tasks may aim to raise awareness and offer practice in producing appropriate phonemes for treatment items and may involve repetition, production from cues (e.g. when given the initial phoneme), making judgements about whether two words rhyme, and so on. However, despite what might appear a clear distinction, semantic therapy tasks frequently involve aspects of phonological processing (e.g. through naming of items) and phonological tasks will frequently involve aspects of semantic processing (e.g. just by being exposed to words or pictures it is argued that semantic processing will be activated; Howard, 2000). Therefore, there is not a strict correlation between type of impairment and type of therapy task that should be employed and again this should be theoretically motivated based on all available information (e.g. level of impairment, participant factors, and so on). Despite this, there are some general observations regarding the effectiveness of semantic and phonological approaches to the remediation of word-finding difficulties. Wiseburn & Mahoney (2009) report the results of a meta-analysis of intervention studies aiming to remediate word-finding difficulties in participants with aphasia. They found that, overall, both semantic and phonological approaches were associated with positive treatment effects. Effects of semantic therapies tended to be smaller but they were more stable between studies and participants whereas effects of phonological therapies were more variable, although they did have the potential to be larger than those effects associated with semantic therapies. It was also found that semantic therapies were more effective than phonological therapies in terms of leading to generalised improvement within untreated items. For example, calculated effect sizes for untreated items that were untreated and not exposed in therapy (i.e. were not used as distractors) were 0.57 for semantic therapies and 0.37 for phonological therapies. When untreated items were exposed in therapy and were related to treated items the effect sizes for semantic and phonological therapies were 1.99 and 1.26 respectively. These general patterns are compatible with prominent individual studies that have looked to compare the effectiveness of these two approaches (i.e. semantic and phonological therapies) when aiming to improve noun retrieval (e.g. Howard, Patterson, Franklin, Orchard-Lisle, & Morton, 1985) and also when aiming to improve verb retrieval (e.g. Raymer & Ellsworth, 2002) although with verbs no differences in effectiveness of therapy types was observed.

5.2.4. Semantic therapy tasks and their effects

While there are numerous types of therapy task that may involve aspects of semantic processing, there are three broad tasks that are prominent in terms of them being developed with the aim of directly improving semantic processing with the aim of improving word retrieval. These are: (1) word-to-picture matching tasks; (2) semantic feature verification and discrimination tasks; and (3) semantic feature analysis tasks. Each of these tasks imposes differing processing demands both within semantic memory and also within general language processing. The following subsections will describe each type of task, give an indication as to the main processing components that are assumed to be underlying success in the task, and also review some of the findings for when the task has been applied to both nouns and verbs (where possible).

Word-to-picture matching

Word-to-picture matching is a common task employed both as an assessment measure and also as an intervention tool. Participants must select a corresponding picture when hearing a spoken word or reading a written word. Target pictures are presented within an array whereby distractors are manipulated in terms of their semantic relatedness to the target. The number of semantically related distractors in comparison to number of unrelated distractors and also the closeness of semantic relatedness can be controlled in order to target representational information at a particular depth of semantic processing. For example, a target item (e.g. *lion*) may be presented alongside other members of the same category that would share a great deal of semantic overlap (e.g. *tiger*, *cat*). This encourages the need to make fine-grained semantic distinctions, and hence facilitates relatively deep semantic processing. Alternatively, the target may be presented with category members that may not overlap in terms of semantic representations, or even presented with members of other categories (e.g. *dog*, *horse*, *apple*, etc.) where shallower semantic processing is sufficient to ensure success in the task.

Success in word-to-picture matching tasks relies on intact input processing via the object/conceptual input modality in order to recognise picture stimuli and input processing via either auditory or written input modalities (depending on whether the task uses spoken or written words; see Figure 5.1). Following this, the task requires intact semantic processing in order to identify a semantic representation that is associated with both the picture stimulus and the lexical stimulus. At this point, the task can be successfully completed by pointing to the correct picture (assuming the task does not also require repetition and/or reading aloud of the lexical item).

Marshall, Pound, White-Thompson & Pring (1990) used a written-word-to-picture matching task in an intervention study targeting noun retrieval for three single cases and within a group of participants with aphasia. The participants showed various patterns of impairment related to the semantic system itself and in accessing phonological representations from the semantic system. The distractor items included semantically related and unrelated items and participants were additionally required to repeat the target in order to ensure phonological representations were accessed. Within the single case studies, two of the three participants showed improved naming of pictures following intervention. Of the two who improved, one showed impaired semantic processing while the other showed impaired ability to access phonological representations from semantics. However, the participant who did not improve also showed impaired semantic processing. This therefore demonstrated the versatility of word-to-picture matching in facilitating improvements in participants with different underlying impairments while also showing that it does not guarantee improvement across participants who appear to have similar levels of impairment. Within the group study, it was found that improved naming was observed in treated items (i.e. those that served as targets) as well as semantically related items that were seen as distractor items within the array of pictures. No improved naming was observed with unrelated items that had appeared as distractors nor with items that had not been used as distractors. These improvements were maintained one month following the end of therapy and these findings were argued to suggest that the therapy task had affected change by strengthening associations between semantic and phonological representations in addition to facilitating processing within the semantic system. It is important to note that further studies have also shown that for such improvements to be observed with word-to-picture matching tasks, the crucial element of the task lies in the verbal production of the target item as improvements are not observed when the task is carried out without a productive element (e.g. Le Dorze, Boulay, Gaudreau & Brassard, 1994).

Marshall et al (1998) report the application of word-to-picture matching focused on improving verb retrieval with a single participant identified as having specific verb retrieval impairment associated with impairment in accessing verbs' phonological representations from their semantic representations. Matching tasks were completed in the presence of semantic, phonological and unrelated distractors and also in conjunction with other semantic based tasks (e.g. identifying the odd-one-out from an array and naming verbs from definitions). Following therapy, the participant showed a significant improvement in naming treated verbs from picture stimuli and a non-significant improvement in naming untreated verbs that were thematically related to those that were treated. As a consequence of improved verb retrieval, the participant also demonstrated improvement in sentence production abilities which confirmed that the therapy had affected change in the manner predicted, i.e. by improving access to phonological representations from semantic representations.

As has been highlighted above, a difficulty with comparing the effectiveness of word-to-picture matching tasks between different word classes and even within the same word class is that there is great variability in how the task is carried out. For example, the number of degree of semantic relatedness of distractors may be controlled within each particular study but without knowing the precise stimuli it is difficult to ascertain whether the demands on semantic processing are comparable between studies.

Also, as highlighted within the Marshall et al (1998) study, word-to-picture matching is often utilised within an intervention that comprises of other elements so again it is difficult to ascertain the precise effect that the word-to-picture matching component has on the speaker's impaired language system.

Semantic feature verification and discrimination

There are a collection of different therapy tasks which are often conducted in parallel and which all require similar processing demands. Such tasks may include sorting items into categories or according to presence or absence of particular semantic features. For example, participants may be required to sort items between general categories of *living* and *non-living* or between more specific categories such as *large animal* and *small animal* (e.g. Warrington, 1975). A further task is feature matching where participants are required to match presented features to a concept presented within an array. Further variations on this also include yes/no verification tasks where participants must respond 'yes' or 'no' as to whether a concept belongs within a particular category or possesses a particular feature. Such studies may potentially be seen to be comparable to the group of semantic-based studies that Boyle (2010) identifies as semantic feature review approaches.

Being able to complete semantic feature and discrimination tasks relies on intact input processing in accordance with how the task is presented (e.g. through pictures, spoken words, written words, or a combination). The task then requires intact semantic processing in order to make comparisons between two differing semantic representations (e.g. between a feature and a concept, or between a concept and a category, or between two concepts). This therefore differs from word-to-picture matching tasks which require convergence upon a single semantic representation. However, as with word-to-picture matching, the depth of semantic processing required may be manipulated from something where general discrimination is required (e.g. discriminating an animal from a non-animal) to where more specific discrimination is required (e.g. discriminating a large animal from a small animal; e.g. Warrington, 1975).

The majority of intervention studies using feature verification and discrimination tasks have done so with the aim of improving noun retrieval and have tended to focus on applying treatment within discrete categories. The rationale for this is that by focusing on category members, features verification and discrimination will entail attention to shared and distinctive features between concepts (i.e. some feature may be common to a large number of category members whereas other feature will be common to fewer members). A number of studies have applied this with the assumption that this will facilitate generalisation to untreated category members due to the overlap in semantic representations. Such studies have been conducted with focus on categories such as clothing and food (e.g. Kiran & Thompson, 2003), birds and vegetables (e.g. Stanczak et al, 2006), and furniture and clothing (e.g. Kiran, 2008). In general, the findings from these studies have converged to support Thompson, Shapiro, Kiran & Sobecks' (2003) Complexity Account of Treatment Efficacy (CATE). It has been found that generalised improvements in naming do occur following therapy, however, improvement only occurs for items that are less typical than those that are treated. For example, if treatment focuses on category members that are atypical, then improvements may be expected in naming of these treated items and also untreated items that are more typical within the same category. However, if treatment focuses on typical category members then improvement may only be expected within these treated items. This is argued to be because the treatment utilising more atypical category members entails greater semantic activation through a greater diversity of featural information that is highlighted through the therapy task (i.e. typical items tend to be composed of a core set of semantic features whereas atypical items possess more distinctive features). Lowell, Beeson & Holland (1995) provide further evidence of the effectiveness of feature discrimination and verification without explicitly focusing on category-based therapy stimuli. They found that improvements were observable in untreated items that were both semantically related and unrelated to the treated items. However, here it was found that only two out of three participants showed these improvements. The participant who did not show improvement was described as having more severe impairment to semantic and/or phonological processing which may have limited the potential to benefit from therapy.

While Marshall et al's (1998) study could be argued to contain aspects of semantic verification and discrimination, there appear to be no studies that have used this approach directly to focus on improving verb retrieval. A potential barrier to conducting such studies in order to make direct comparisons is the current lack of understanding of whether and how verbs constitute semantic categories or clusters as determined by their internal semantic feature composition.

The use of semantic feature verification and discrimination tasks within intervention studies does appear to be potentially effective in improving word retrieval, although it is currently unknown as to how effective this may be with verb retrieval.

Comparisons between different studies are again problematic as such tasks can be presented in different ways and many studies include a variety of different methods to draw attention to featural properties of treatment stimuli. Therefore, as with word-to-picture matching tasks it may be difficult to ascertain exactly what effect each component, or what each individual task, on its own has on the speaker's impaired language system.

Semantic feature analysis

Semantic feature analysis (SFA) is an approach where typically, participants are presented with a target picture which they are required to name and then subsequently generate a number of semantic features that are associated with the target. These studies are therefore consistent with what Boyle (2010) identifies as semantic feature generation which differs from tasks employing semantic feature review (i.e. semantic feature verification and discrimination tasks) in that the responsibility lies with the participant to actively generate featural responses. SFA approaches generally also avoid one of the problems with word-to-picture matching and semantic feature verification and discrimination studies, namely the variety in delivery between studies. SFA tends to be delivered as a single therapy task which comprises the whole intervention. As such, the effectiveness of SFA between studies is slightly easier to judge as the method tends to vary little, at least between the majority of earlier reports.

Success in SFA depends on intact processing through both input and output processing which is necessarily mediated by an intact semantic system. Input processing is necessary in order to recognise the picture stimuli and subsequently identify an appropriate semantic representation. Semantic processing is required in order to identify related semantic information which would be appropriate as features. Output processing is required in order to provide verbal responses for both the name of the target concept (i.e. the picture) and for the related semantic features. It has been assumed that SFA therapy tasks operate on principles of spreading activation within semantic networks as associations between concepts and their semantic representational information is reactivated or relearned and that this activation may spread to other concepts that may have overlapping semantic representations. This assumption may therefore predict that SFA approaches would be effective in promoting improvements to treated items and also to untreated items that are semantically related to those that are treated. Alternatively, SFA has been argued to work as an effective conscious strategy to combat word-retrieval difficulties, where participants may internalise a feature generation procedure in efforts to increase activation potential of the to-be-retrieved word (e.g. Boyle & Coelho, 1995). Under this assumption, SFA would therefore be effective in promoting generalised improvements to untreated items regardless of whether they are semantically related to the target or not.

Boyle & Coelho (1995) provided the first single case study of SFA targeting improved noun retrieval with a participant described as having Broca's-type aphasia with prominent word-finding difficulties. In this version of SFA, the participant was required to generate semantic features relating to the target's: Group (is a ...); Use (is used for/to ...); Action (does what?); Properties (has/is); Location (is found ...); and Association (reminds me of a ...). Following therapy, the participant showed improved naming ability of items that were treated and also that were untreated and unrelated to treated items (in the absence of an untreated and semantically related set of items) and these improvements were maintained at both one- and two-months following the end of intervention. Similar findings have also been observed when this same SFA protocol has been subsequently replicated (e.g. Boyle, 2004; Coelho, McHugh & Boyle, 2000).

Wambaugh & Ferguson (2007) used SFA with a focus on improving verb retrieval with a single participant described as having anomic aphasia. The participant was required to generate semantic features for: Subject (who usually does this?); Purpose of action (why does this happen?); Part of body or tool used to carry out action (what part of the body or what tool is used to make this happen?); Description of physical properties (tell me what it looks like); Usual location (where does this action usually take place?); and Related objects or actions that reminded the participant of the target verb (what does it make you think of?). Following intervention the participant was reported to show improved naming of treated verbs from between 30-40 percent accuracy to 60 percent accuracy (in 10-item treatment sets) although there was no parallel improvement in naming of untreated items.

A further study by Faroqi-Shah & Graham (2011) reports the use of SFA within an additional component of sentence generation with treatment items selected according to Levin's (1993) verb categories of *verbs of cutting* and *verbs of contact*. This was done on the assumption that verbs within these categories share some semantic features within their respective abstracted event templates (i.e. contact, motion, and action). Therefore, it was predicted that generalisation may be observed to untreated items on the basis of this semantic overlap. Where one participant showed no improvement at all, the second participant showed improvement only to treated verbs. However, both participants did show improvement in naming action pictures within An Object and

Action Naming Battery (OANB; Druks & Masterson, 2000) and they also showed improvements in spontaneous speech measures in the Western Aphasia Battery (Kertesz, 1982). Therefore, in comparison to reports of SFA focusing on noun retrieval which yields patterns of generalised improvement, SFA with verbs has so far only resulted in item specific improvement.

While there is evidence that SFA may facilitate generalised improvements in naming ability, one potential complication is that most reported studies have employed repeated probing of naming alongside therapy in order to monitor improvements as and when they occur. This is useful to ascertain the rate of improvement that an individual shows as a result of a particular therapy approach. However, given that this repeated probing is administered while therapy is ongoing, this makes it difficult to separate effects of pure SFA to effects of SFA plus this additional naming task which includes naming practice of both treated and untreated items (see Howard, 2000 for discussion). From a clinical perspective, this may indeed be ideal – to supplement SFA with additional naming to promote widespread semantic activation – however, from a research perspective it makes it difficult to truly understand the mechanisms of SFA-type therapies and how they affect change in semantic and language processing.

A further issue that is so far under-researched is the effect of SFA (and other therapy approaches) on aspects of comprehension processing. If SFA does affect reactivation or relearning of semantic representations, this may be expected to improve comprehension abilities of participants who show impairments in semantic processing pre-therapy.

For the current study, semantic feature analysis was chosen as the therapy intervention. This was primarily because of the claims regarding its ability to facilitate generalised improvement in naming but the as yet unclear explanation for how this generalisation is achieved. Boyle (2010) highlights that the fact that naming of untreated items being repeatedly probed alongside the therapy phases of intervention complicates the issue but there are other issues to consider. Most studies tend to rely on relatively small item sets on which to measure improvement, for example with Wambaugh & Ferguson (2007) measuring improvement within 10 item sets when baseline performance is already between 30-40 percent. This often means that it is difficult, if not impossible to report statistical change from pre- to post-therapy performance with improvements tending to be reported in terms of percentage change or in terms of effect sizes. While such descriptive measures are informative and may be clinically relevant, on small item sets, their validity and robustness may be questioned. It has also been theorised that generalisation, particularly as a result of semantic-based therapies may be most likely within semantic categories or at least where there is semantic overlap between treated and untreated items (e.g. see Plaut, 1996; Nickels, 200, for discussion). This has to an extent been demonstrated with those therapies applying semantic feature verification and discrimination (i.e. what Boyle, 2010, would identify as semantic feature reviews) but has yet to be fully investigated within semantic feature analysis (i.e. semantic feature generation) approaches. Without establishing whether improvements are observed in semantically related items or whether they are more wide ranging and observable in items that are not semantically related leads to a situation where any generalisation can only plausibly be explained as 'strategic' whereby participants internalise a strategy to employ when experiencing word-finding difficulties (e.g. as argued by Boyle & Coelho, 1995). This consequently entails that this does little to inform the theory on which semantic feature analysis methods are based. For example, it is argued that the approach takes advantage of spreading activation within semantic networks, hence it may be expected that semantically related untreated items may benefit more (if at all) compared to items that are unrelated. At present, this has clearly been an overlooked aspect in reports of such intervention approaches.

5.2.5. Comparing therapy effects for nouns and verbs

For the majority of reports of semantic-based therapy approaches, it is difficult to make comparisons of the effectiveness of the therapy on remediation of noun retrieval and the remediation of verb retrieval. This is because few studies have looked to directly compare the exact same therapy protocol and experimental design and its effects when targeted at noun retrieval and verb retrieval. A further complication is that studies have tended to employ differing intensities and durations of treatments, which again, makes direct comparison problematic. This is important for the purposes of the current argument as this makes it difficult to draw conclusions about the underlying effect of therapy and its consequences, along with the effectiveness of facilitating reactivation or relearning of semantic representations and processing of nouns and in particular verbs. However, a group of more recent research studies have emerged which attempt to apply comparable therapies for nouns and verbs within the same participants and compare their relative effectiveness. Few of these studies have employed explicitly semantic-type approaches but they are insightful as they tend to demonstrate that therapies can be equally effective in facilitating improvement in noun and verb processing.

Raymer et al (2007) used a combined semantic and phonological therapy approach within a group of eight participants with varying levels of impairment and severity of aphasia. This was applied to both nouns and verbs within a cross-over design and involved aspects of raising awareness of semantic (e.g. categorical and associative relations) and phonological (e.g. initial phoneme and rhymes) properties as well as repetition of target items. Following intervention, five of the eight participants showed significant and equivalent improvement in naming of treated nouns and verbs with no parallel improvement in untreated items in either word class. A further key finding was that degree of improvement was correlated with aphasia severity such that those participants whose aphasia was most severe showed the least improvement (i.e. none) whereas those with least severe level of impairment showed the greatest improvement. Similar findings were also observed by Conroy, Sage & Lambon Ralph (2009a and 2009b) who reported the application of errorless versus errorful and decreasing cues versus increasing cues therapy approaches within group studies. It was again found that naming of both treated nouns and treated verbs improved following intervention with no concomitant improvement in untreated items and again, degree of improvement was correlated with severity of impairment. Unlike Raymer et al's (2007) study where there was equivalent improvement in treated nouns and verbs, Conroy et al (2009a and 2009b) reported greater improvement of treated nouns over verbs which they attributed to a greater difficulty in relearning and/or reactivating verbs' representations and also in retaining this information once therapy was complete.

5.2.6. The current study and research questions

The discussion so far has claimed that insights into semantic representations and semantic processing can be obtained by observing patterns of language behaviour in speakers with language impairment. The following sections report an intervention study with five participants with aphasia, all of whom had differing patterns of impairment, but who all showed characteristics of word-finding difficulties affecting both nouns and verbs. By applying an intervention approach that was assumed to explicitly tap semantic representational information of nouns and verbs (i.e. Semantic Feature Analysis) it was hypothesised that by observing patterns of improvement within word classes, i.e. from nouns to other nouns and from verbs to other verbs, insights could be gained into the nature of the semantic representations of nouns and verbs. Improvement was measured in terms of ability to name pictures for pictures representing nouns and verbs that were: (1) explicitly treated in the study; (2) untreated and semantically related to those that were treated; and (3) untreated and unrelated to those that were treated. The outcomes were considered in both group and individual analyses.

If the intervention approach explicitly taps semantic representations, as was assumed, it was hypothesised that improvement may be observed in ability to name treated items and untreated items that were semantically related to the treated items. This was expected given that there should be semantic overlap, in terms of semantic features, between these sets of items and this may facilitate reactivation of relearning of lost or damaged semantic information. Improvement was not expected in naming of untreated unrelated items as these had little, if any, semantic overlap with the treated items. Although there is so far little evidence to suggest that such a pattern would be observed in verb naming performance, this was nevertheless the hypothesis for the current study because previous studies had not selected verb stimuli in a systematic manner to allow comparison between treated, untreated related, and untreated unrelated sets. Verb stimuli had tended to be selected according to performance during baseline assessment leading to there being little semantic overlap between and within treated and untreated sets (i.e. verbs tend not to be selected within discrete categories and semantic feature overlap has not been considered as a factor).

In addition to measuring improvement at a single-word level (i.e. through picture naming), improvement was also measured in sentence processing abilities (i.e. sentence comprehension and sentence production). As the intervention approach aimed to facilitate reactivation or relearning of semantic information that would aid word retrieval, it was appropriate to assess the potential for generalisation to other contexts which more closely resemble communication in real-life contexts (i.e. ability to retrieve words in sentences). At a theoretical level, improvements in sentence processing abilities were anticipated as the intervention approach may well facilitate access to thematic information associated with nouns and in particular verbs, which may in turn facilitate sentence level processing (e.g. Mitchum, Haendiges & Berndt, 1995). These improvements may nevertheless be highly dependent (even more so than single-word processing) on the nature of the underlying sentence processing impairment and therefore these analyses were only conducted at an individual level.

This study therefore aimed to address the following questions:

1) Does SFA therapy lead to improvement in ability to retrieve nouns and verbs? If so, are improvements observed to:
    a. Treated items?

b. Untreated items that are semantically related to treated items?
c. Untreated items that are unrelated to treated items?

2) Are any improvements observed equivalent between ability to retrieve nouns and ability to retrieve verbs?

3) For participants who show impairment in sentence processing, does SFA therapy lead to improved ability to understand and produce sentences?

4) For any improvements that are observed, what has been the underlying mechanism for improvements, i.e. how has SFA affected change in semantic and language processing?

5.3. Method

5.3.1. Design

The current study used a cross-over design whereby the order of therapy phases (i.e. verb phase and noun phase) was counterbalanced between participants. Background assessment for each participant was conducted during pre-therapy in order to ascertain hypothesised levels of semantic and/or language impairment. Each therapy phase was carried out over 10 one hour sessions at an intensity of two sessions per week. Naming performance on therapy items (treated and untreated) was probed twice during pre-therapy assessment and then again following the completion of each individual therapy phase and then again during a maintenance assessment phase following a 4-5 week break (i.e. to assess maintenance of therapy effects). The entire duration of each participant's involvement in the study was approximately 20 weeks (including the 4-5 week break). The major phases of the study are presented in Figure 5.2.

Pre-therapy assessment	• 6 x 1 hour sessions
Therapy phase 1	• 10 x 1 hour sessions
Post-therapy 1 assessment	• 1 x 0.5 hour session
Therapy phase 2	• 10 x 1 hour sessions
Post-therapy 2 assessment	• 4 x 1 hour sessions
Break	• 4-5 weeks
Maintenance assessment	• 2 x 1 hour sessions

Figure 5.2 Phases of intervention study

5.3.2. Pre-therapy assessment of semantic and language processing

Prior to the therapy phases, each participant completed a battery of assessments probing aspects of cognitive, semantic, and language processes in order to develop hypotheses about their respective levels of impairment.

Cognitive assessment

- Raven’s coloured progressive matrices (short version; Raven, Raven & Court, 1998), as an assessment of non-verbal cognitive reasoning. The participant is required to identify a missing piece from a visual pattern from a choice of six.
- Comprehensive Aphasia Test (i.e. CAT; Swinburn, Porter & Howard, 2004) Subtest 1 – Line bisection, as an assessment of visual attention/neglect. The participant is required to mark the mid-point on three lines of different lengths.
- CAT 2 – Semantic memory, as an assessment of semantic level processing independent of language processing. The participant is required to identify a semantic associate of a target picture from a choice of four.
- CAT 4 – Recognition memory, as an assessment of retention of visual/semantic information. The participant is required to identify pictures previously presented in CAT 2 from a choice of four.
- CAT 15 – Repetition of digit strings, as an assessment of short-term memory. The participant is required to repeat digit string of increasing length.

Semantic and language comprehension

- CAT 7 – Spoken word to picture matching, as an assessment of access to semantic representations via the auditory input modality (i.e. spoken nouns). The participant is required to identify a picture corresponding to a spoken word from a choice of four including semantic and phonological distractors.
- CAT 8 – Written word to picture matching, as an assessment of access to semantic representations via the written input modality (i.e. written nouns). The participant is required to identify a picture corresponding to a written word from a choice of four including semantic and phonological distractors.
- Pyramids and Palm Trees Test (i.e. PPT; Howard & Patterson, 1992) three-word, and three-picture versions, as an assessment of access to fine-grained semantic representational information of objects/nouns. Participants are presented with a target (e.g. *pyramid*) and must identify its semantic associate from a choice of two (e.g. *palm tree* or *pine tree*).
- Kissing and Dancing Test (i.e. KDT; Bak & Hodges, 2003) three-word and three-picture versions, as an assessment equivalent to PPT that probes semantic representations of actions/verbs. Participants are presented with a target (e.g. *kissing*) and must identify its associate from a choice of two (e.g. *dancing* or *running*).
- Sentence Comprehension and Production in Aphasia (i.e. SCAPA; Webster & Whitworth, in prep) comprehension subtest, as an assessment of sentence comprehension. The participant hears a spoken sentence and must select the matching picture from an array of four which included the target, reverse-role distractors and lexical distractors (i.e. depicting a different action/verb).

Language production

- CAT 12 – Repetition of spoken words, as an assessment of phonological output that is independent of semantic processing. The participant is required to repeat a series of words spoken by the researcher.
- CAT 20 – Reading written words aloud, as an assessment of access to and from the orthographic output lexicon. The participant is required to read aloud a list of written words.

- CAT 3 – Verbal fluency, as an assessment of verbal output when prompted by semantic (i.e. listing *animals*) or graphemic/phonological categories (i.e. listing *words beginning with 's'*).
- An Object and Action Naming Battery (i.e. OANB; Druks & Masterson, 2000), as an assessment of confrontation noun and verb naming when elicited by picture stimuli. The OANB additionally allows analysis of psycholinguistic variables that may affect naming performance (e.g. lexical frequency, familiarity, imageability, and so on).
- CAT 25 – Written picture naming, as an assessment of orthographic output via pictorial input and semantic processing. The participant is required to write the name of objects shown in pictures.
- SCAPA production subtest, as an assessment of sentence production elicited from pictures. The assessment elicits both active and passive sentences with potentially reversible and non-reversible thematic participants (e.g. *the man pulls the girl*: active reversible; *the boy kicks the box*: active non-reversible). In combination with the SCAPA comprehension subtest, this can help to differentially diagnose sentence processing difficulties associated with: verb retrieval difficulties, predicate argument structure difficulties, thematic role assignment difficulties, and mapping difficulties.

5.3.3. Pre-therapy selection of therapy items

During the pre-therapy assessment phase of the study, participants also completed two baseline administrations of both object/noun picture naming and action/verb picture naming in order to inform the selection of therapy items. Object naming was assessed using 166 pictures mostly drawn from the 'Snodgrass & Vanderwart-like picture set[4] (i.e. Rossion & Pourtois, 2004) with some additional pictures produced by an illustrator. These pictures included items from the categories: *animals* (n = 24), *clothing* (n = 17), *fruit* (n = 12), *furniture* (n = 14), *musical instruments* (n = 10), tools (n = 19), *transport/vehicles* (n = 16), *vegetables* (n = 14), as well as 40 pictures of other common objects. The target nouns that these pictures aimed to elicited included nouns of one (n = 74), two (n = 62), three (n = 22), and four syllables (n = 8). Action naming was probed using 100 action pictures mostly drawn from the Newcastle Aphasia Therapy Resources (Webster et al, 2009) with additional

[4] Images courtesy of Michael J. Tarr, Brown University. http://www.tarrlab.org/

pictures provided by an illustrator. These actions included items from the following categories (as determined by category listing data; see Chapter 2): *breaking* (n = 10), *cleaning* (n = 12), *cooking* (n = 9), *cutting* (n = 5), *hitting* (n = 8), *making* (n = 9), as well as 47 pictures of other common actions. A number of verbs also had potential to be associated with more than one category (e.g. *ripping* was associated with both *breaking* and *cutting* categories) although the numbers reported above reflect 100 different actions placed within the category they were most strongly associated with. All verb targets were canonically transitive requiring a subject (i.e. thematic agent) and a direct object (i.e. thematic patient or theme). The target verbs (in the uninflected form) included those of one (n = 84), two (n = 13), and three syllables (n = 3). A complete list of stimuli and relevant values for psycholinguistic variables are included in Appendix S (verbs stimuli) and Appendix T (noun stimuli).

Object and action naming was conducted separately and on two occasions each with a minimum of one week between administrations. On each administration, pictures were presented in a different random order and within two equal sized blocks to avoid assessment fatigue. For object naming, participants were instructed to name the object and for action naming participants were instructed to name the action using a single word. Responses were scored correct if they matched the expected target or were an acceptable alternative. Only first responses given within 10 seconds of initial picture presentation were scored, otherwise a no-response was recorded. False starts (e.g. single consonants with or without a short schwa) and hesitations were not considered as responses and so correct responses following these were scored as correct as long as they were produced within 10 seconds of initial picture presentation. For action naming, if participants gave a complete sentence, this was scored as incorrect even if the correct target verb was used (although this rarely happened). If participants gave a phrasal response this was correct as correct only when the phrase began with the target verb (e.g. *reading a book* for *reading*).

Based on each participant's performance in object and action picture naming, 60 noun targets and 60 verb targets were selected as therapy items. These were selected to mostly include items that were not named on either baseline administration of naming but some items were included that were named on one occasion and also on both baseline administrations. This was done to ensure some level of success during therapy but also to improve validity of statistical analysis by not having participants' performance at zero and so ensuring that relatively small gains would not lead to significant differences in statistical analysis. Items were selected so that baseline performance was generally between 20-40% success for both object naming and action naming for each participant.

Of the 60 items that were selected within each word class, 40 were drawn equally from five semantic categories and these were then divided equally into two sets of 20 items (i.e. there were four items from each of five categories in each set of 20). These two sets were then randomly allocated as either the treated set of an untreated related set (i.e. categorically related). The remaining 20 items were selected from the remaining items but did not overlap with the categories used in the treated and untreated related set; these made up an untreated unrelated set of items. As participants' naming performance differed, the allocation of items to sets differed, even to the extent that different participants' items were drawn from different semantic categories. These three sets of items within each word class were matched as far as possible for length and lexical frequency (according to data from the British National Corpus) although this was not always possible given the overriding criteria of selecting categorically based items that were matched for baseline naming performance (see Appendix U for frequency matching statistics). In general, treated and untreated related sets were usually matched for frequency but untreated unrelated sets tended to have higher mean frequencies than both of these.

5.3.4. Therapy protocol

A version of semantic feature analysis (SFA) therapy was used for both therapy phases (i.e. nouns and verbs). For both noun-SFA and verb-SFA, participants were shown a worksheet which presented the target picture and spaces for four semantic features (example worksheets are given in Appendix V). On initial presentation, participants were asked to name the picture using a single word (i.e. noun or verb). Regardless of whether they could name the picture correctly, participants were then required to verbally produce four semantic features that could be associated with the target. If participants were unable to give a verbal response spontaneously, the researcher offered a forced-choice alternative. Forced-choice alternatives aimed to present a close semantic distractor to encourage continued semantic processing (e.g. for the Location feature for the target *banana*, the forced choice may be '*does it grow on trees or in the ground?*'). If a participant was still unable to identify an appropriate feature, the researcher gave this and justified the selection. Once all four features had been produced, the participant was again asked to name the target picture. If they were unable to name the picture once all four features were produced, a forced-choice alternative was offered and if the participant was still unable to identify the correct option, this was given and justified. Again, forced-choice alternatives presented semantic distractors that were within the same category. All responses were written on the worksheet in the spaces provided as and when the participant gave them. After all features and target name were identified and written on the worksheet, these were reviewed before moving on to the next worksheet.

In a single therapy session, SFA was carried out with 10 target items. Their allocation to therapy sessions was random but was done so that all 20 treated set items were seen across two consecutive sessions. Once all 10 items and worksheets were completed within a session, these were again reviewed.

During noun-SFA therapy, participants were required to produce four semantic features, each of which was prompted by a keyword in addition to a carrier phrase or question that was also read aloud by the researcher to prompt verbal responses. These features were: Group (It's a type of ...); Location (Where could you find or see it?); Description (What does it look or sound like? / What is it made of?); and Action (What does it do? / What can you do with it?). During verb-SFA participants were required to generate features relating to: Purpose (It's a way of ... something); Tool (What could you use to do this?); Description (What does it look or sound like? / What movement is involved?); and Related object (What or Who could you do this to?).

5.3.5. Outcome measures

The primary outcome measure for the current study was performance on picture naming of the 60 object pictures and 60 action pictures that were selected as therapy items (treated and untreated). In addition to the two pre-therapy administrations, this was assessed following both therapy phases and again during the maintenance assessment phase of the study. Items were presented in different random orders within discrete blocks (i.e. actions and objects) at each administration and scoring was consistent with the procedure used during baseline administrations.

A secondary outcome measure was object and action naming as assessed by the OANB. This was reassessed following the second therapy phase and served as an independent measure of noun and verb production elicited via picture stimuli. Although there was some overlap in therapy items and the items probed on the OANB the picture stimuli were all different. Therefore, any improvement in OANB performance could not be attributed to an increased familiarity and exposure to the picture stimuli.

Improvement in the OANB may be interpreted as evidence of generalised improvement of word-retrieval.

Participants were also reassessed on both comprehension and production subtests of the SCAPA assessment following the second therapy phase. Improvements in sentence processing may be anticipated as a consequence of: (1) improved access to semantic representations of verbs word retrieval at single-word level; (2) generally improved word retrieval at single-word level; (3) improved awareness of thematic role information associated with verbs (e.g. as in mapping therapies; e.g. Schwartz, Saffran & Fink, 1994). Such improvements are likely to be dependent on an individual participant's level of impairment, for example, if background assessment demonstrates an intact semantic system then they would not be expected to show improvement attributable to improved access to semantic representations. It is also possible, if not highly likely, that participants may have additional levels of impairment specifically affecting sentence level processing which may negate any general improvement in word-retrieval when this is measured in sentence level tasks.

A control measure (i.e. one not expected to change as a result of therapy) was also intended to be selected according to each participant's performance during the baseline assessment phase. Repetition of digit string (i.e. CAT 15) was the only assessment that may not be predicted to change as a result of therapy and was an assessment where all participants' performance was at or below the cut-off for normal performance. This was not ideal however, as it was problematic to choose a language-based control measure if the therapy approached is hypothesised or assumed to potentially affect multiple processing components of language.

5.3.6. Participants

Five participants took part in the current study. All were referred by NHS speech and language therapists in the North East of England. All participants were monolingual British English speakers and all had suffered aphasia as a consequence of a cerebrovascular accident (CVA). Demographic information is presented in Table 5.1 while results from pre-therapy assessment are presented in Table 5.2.

	AB	GF	JA	RH	SH
Sex	Male	Male	Male	Male	Female
Age (years)	45	61	70	54	48
Time post-onset (months)	45	15	32	31	27
Lesion site	Left MCA	Left MCA	Right MCA	Left MCA	Left MCA
Prior occupation	IT project manager	Building firm contracts manager	Retired armed forces and delivery driver	Job centre worker	Nurse and volunteer counsellor
Handedness	Right	Right	Right	Right	Right
Hearing	Normal	Normal	Normal	Normal	Normal
Vision	Normal	Normal	Correct for reading (glasses)	Right hemianopia and glasses	Corrected for reading (glasses)

Table 5.1 Background information of participants

Assessment	Norms (where available)			AB	GF	JA	RH	SH
	n	M (SD)	cut-off					
Cognition								
Raven’s coloured progressive matrices (short version)	12		10	11	6	7	11	11
CAT 1 – Line bisection		+/- 1.32	+/- 2.5	-1.5	-1	0	disc	+0.5
CAT 2 – Semantic memory	10	9.81	8	9	10	6	8	10
CAT 4 – Recognition memory	10	9.7	8	9	10	8	10	10
CAT 15 – Digit string repetition		6.44	4	2	3	4	3	2
Semantic and language comprehension								
CAT 7 – Spoken word to picture matching	30	29.15	25	25	30	25	29	28
CAT 8 – Written word to picture matching	30	29.63	27	28	27	23	21	30
Pyramids and Palm Trees Test								
3 pictures	52	51.1 (0.8)		49	52	41	50	51
3 words	52	51.8 (0.6)		50	48	41	37	52
Kissing and Dancing Test								
3 pictures	52	50.4 (1.5)		48	47	39	45	49
3 words	52	51.4 (0.8)		43	38	45	39	51

Table 5.2 Results of pre-therapy assessment

Assessment	Norms (where available)			AB	GF	JA	RH	SH
	n	M (SD)	cut-off					
Semantic and language comprehension (cont)								
SCAPA – Comprehension subtest								
Sentences	60	59.7 (0.5)		31	36	29	38	39
Verbs	60			52	48	45	48	50
Language production								
CAT 12 – Word repetition	32	31.73	29	30	24	20	32	10
CAT 20 – Reading words aloud	48	47.42	45	16	12	36	18	25
CAT 3 – Verbal fluency		32	13	4	5	11	2	16
OANB								
Objects	162			106	80	119	72	115
Actions	100			37	36	53	42	63
CAT 25 – Writing picture names	21	20.15	15	15	4	1	4	n/a
SCAPA – Production subtest	60	59.5 (0.8)		3	2	6	4	30

Table 5.2 Results of pre-therapy assessment (continued)

AB

AB presented with functional everyday comprehension and moderate-to-severe non-fluent conversational speech. At a single-word level of processing, AB was hypothesised to have impairment at the level of the semantic system and also in accessing phonological representations from the semantic system.

Evidence of impairment to the semantic system came from reduced performance in comprehension assessment through all input modalities. AB's overall picture naming performance was significantly influenced by imageability (*Wald* = 17.74, $p < .001$). Errors in picture naming were generally non-responses or semantic-type errors. AB was also more accurate reading aloud high-imageability words compared to low-imageability words ($t = 2.24$, $p = .037$). Failures in word-retrieval were attributed to impaired access to intact phonological representations as AB could often correct name the target when given a phonemic cue and he could often spontaneously trace the initial letter with his finger (although this strategy was not effective in cuing correct naming).

Comprehension for both nouns and verbs was impaired with no difference in comprehension of pictures but a significant noun advantage in comprehension of words ($p = .016$). Nouns appeared better preserved than verbs in picture naming (65% correct versus 37% correct) although on a matched sets analysis (matched for psycholinguistic variables; Conroy et al, 2009b), AB scored equally in noun and verb naming (both 12/20 correct).

AB had additional sentence level difficulties with a significant reversibility effect in sentence comprehension (Fisher's exact, $p = .002$). He showed non-fluent agrammatic sentence production which consisted mostly of incomplete and abandoned sentences following production of either the subject noun of the main verb.

GF

GF presented with functional comprehension and severely non-fluent conversational speech which was also affected by residual difficulties associated with apraxia of speech. At a single word level of processing, GF was hypothesised to have impairment in accessing phonological representations from a relatively intact semantic system in addition to impairment at the level of phonological assembly which was in addition to apraxia of speech.

Evidence for impairment in accessing phonological representations from semantics came from GF's better ability to read aloud high-frequency words compared to low-frequency words ($t = 2.65$, $p = .033$). Evidence of impairment at the level of the phonological output lexicon came from GF's tendency to make phonological errors in word repetition (e.g. president → presents), reading words aloud (e.g. *position* → /posɪʃɪʃ/), and picture naming (e.g. *hammock* → /hanək/). Object picture naming was additionally influenced by number of phonemes (*Wald* = 4.6, p < 05). There was an additional suggestion of some level of impairment to semantics as GF was better at reading aloud high-imageability words compared to low-imageability words (t = 3.32, p = .007). GF also made a number of semantic errors in picture naming although both these characteristics could also be attributable to impairment in accessing phonological representations from semantics as opposed to solely semantic impairment.

GF showed a significant noun advantaged in single-word comprehension via pictures (p = .002). Nouns also appeared to be better persevered than verbs in picture naming (49% correct versus 36% correct) although in the matched sets analysis he showed no difference in performance (t = 1.27, p = .214).

In sentence comprehension, GF showed a non-significant trend towards a reversibility effect (Fisher's exact test, p = .082). In sentence production, GF showed output which was limited to isolated nouns with a high proportion of verbs either omitted or inappropriately substituted. Production was also characterised by perseveration and unsuccessful attempts to self-correct.

JA

JA presented with functional comprehension and mildly non-fluent conversational speech which was disrupted by occasionally severe word-finding difficulties. These difficulties frequently led to increased frustration which would further exacerbate word-finding problems. JA's production was also affected by residual difficulties associated with an apraxia of speech. He scored below normal levels on assessment of non-verbal reasoning which may be explained by him suffering a right-hemisphere CVA. At a single-word level of processing, JA was hypothesised to have impairment to the semantic system, in accessing phonological representations from the semantic system, and also in phonological assembly.

Evidence of semantic impairment came from reduced performance in comprehension assessment across different modalities and JA's naming performance was also influenced by both imageability (*Wald* 10.07, p = .01) and age of acquisition (*Wald* = 7.07, p = .01). JA's errors in picture naming tended to be no-responses and semantic-type errors with JA frequently showing an awareness of errors but an inability to self-correct (e.g. by commenting 'it's almost right but not quite'). This therefore showed impairment in accessing correct phonological representations even when correct semantic representations were accessed. Evidence of impairment in phonological assembly came from a tendency to make phonological errors in word repetition (e.g. *faith* → *face*), reading words aloud (e.g. *trout* → /kɹaʊt/), and picture naming (e.g. *tractor* → /ɹaktə/).

Comprehension for both nouns and verbs was equally impaired via both picture and word comprehension. In picture naming, JA appeared to perform better overall on nouns compared to verbs (74% correct versus 53% correct) but matched sets analysis showed a non-significant difference in performance (t = 1.27, p = .214).

In sentence comprehension, JA scored 29/60 correct with a mixture of reverse-role, word-order, and lexical (i.e. incorrect verb selected) errors; although no error type was associated with a significant influential effect on performance. Sentence production was characterised by frequent abandonments often preceding or following production of the main verb which was also frequently inappropriately substituted (i.e. 40/60 main verbs were innapropriate for the target picture, in addition to 8/60 that were omitted completely).

RH

RH presented with functional comprehension although he occasionally required repetitions and simplified verbal input for understanding task instructions. His conversational speech was moderately-to-severely non-fluent, being limited predominantly to single words and short phrases. RH was hypothesised to have impairment to the semantic system, in accessing phonological representations from semantics, and possible impairment to the phonological output lexicon.

Evidence of impairment to the semantic system came from reduced performance in comprehension across input modalities although it also appeared poorer via the written input modality suggesting further difficulties in accessing the semantic system via the written modality. RH's object naming was significantly influenced by imageability (*Wald* = 4.35, $p < .05$) and he was better at reading aloud high-imageability words compared to low-imageability words (t = 3.54, p = .002). When experiencing word-finding difficulties, RH would frequently spontaneously gesture an appropriate object or action and could occasionally benefit from phonemic cues, both of which suggest difficulty accessing phonological representations from intact semantic representations. Overall picture naming was also influenced by lexical frequency (*Wald*

= 4.41, $p < .05$) which may additionally indicate impairment in the phonological output lexicon but may equally be attributable to semantic level impairment.

RH's performance suggested that comprehension of nouns was better preserved compared to verbs although this difference only approached significance in comprehension via pictures ($p = .063$). He showed overall equivalent performance in picture naming (44% correct noun naming versus 42% correct verb naming) though the matched sets analysis showed a trend towards a significant advantage for noun production ($t = 1.94$, $p = .06$).

In sentence comprehension, RH showed a significant reversibility effect (Fisher's exact test, $p = .037$). Sentence production was characterised by a large proportion of incomplete and abandoned responses with high proportions of inappropriately substituted nouns and verbs.

SH

SH presented with good levels of comprehension and mildly non-fluent conversational production. She occasionally showed word-finding problems which she often overcame by consciously substituting the target for a similar word. SH was hypothesised to have impairment in accessing phonological representation from semantics and also in phonological assembly.

Evidence for impairments came from failures in picture naming where SH often commented 'I know it but can't say it'. On these occasions, she often benefitted from phonemic cues, suggesting that the phonological representations themselves were intact. SH did however show phonologically related errors in word repetition (e.g. *vine* → /vaɪm/), reading words aloud (*fraud* → /fɹɔg/), and picture naming (e.g. *table* → /neɪpl/) where errors were often characterised by conduite d'approche, although these attempts often failed to arrive at the target (e.g. piano → /pani, panli, plani, pəlani/). There was some evidence of semantic impairment as SH was better at reading aloud high-imageability words compared to low-imageability words ($t = 3.54$, $p = .002$).

SH showed equivalent comprehension of nouns and verbs that was within normal limits. In picture naming, she showed roughly equivalent overall performance with nouns and verbs (71% correct versus 63% correct respectively) although matched sets analysis showed significantly better verb naming ($t = 2.28$, $p = .03$).

In sentence comprehension, SH scored 39/60 correct with no significant effect of word-order or reversibility although 10 errors were lexical (i.e. selecting a picture showing a non-target verb). In sentence production, SH showed a trend towards a significant reversibility effect (Fisher's exact test, p = .085) suggesting that errors were more likely to occur when target sentences contained potentially eversible participants (i.e. where both subject and object were animate). Errors were characterised by an innapropriate substitution or omission of the main verb (n = 20) and/or reversed thematic/syntactic structure.

Summary of participants' impairments

Table 5.3 presents a summary of the hypothesised levels of impairment for each of the five participants.

	Access to semantic system	Semantic system	Access to output lexicon/s	Output lexicon/s
AB	✓	×	×	✓
GF	✓	✓	×	×
JA	✓	×	×	×
RH	×?	×	×	×
SH	✓	×?	×	×

Note.: ✓ normal; × impaired; ×? potential impairment

Table 5.3 Participants' hypothesised levels of impairment

5.4. Results

5.4.1. Ability to self-generate semantic features

A preliminary analysis was conducted in order to investigate the participants' ability to complete the SFA protocols within noun- and verb-SFA. This was analysed in terms of the number of semantic features that were spontaneously generated by participants as opposed to those identified by participants following either a forced-choice alternative, or failing this, through the researcher offering an appropriate response for the participant to repeat.

Figure 5.3 and Figure 5.4 summarise each participant's ability to spontaneously generate semantic features for treated items within each therapy phase as the sessions progressed. All 20 treated items were treated across two consecutive sessions within a single week and this gave opportunity to produce a total of 80 semantic features. As can be gathered from visual inspection of the data, feature production was consistently more spontaneous across participants through noun-SFA compared to verb-SFA, with all participants showing gradual increases in the proportion of features that were given spontaneously. A one-way within-participants ANOVA confirmed a significant increase in spontaneous feature production over time within noun-SFA (F (4,16) = 7.23, p = .002). Feature production in verb-SFA was more variable with generally lower proportions of spontaneous production and with only two participants showing continual increases as the phase progressed (i.e. RH and SH) although all but JA showed increases in week five compared to week one. A one-way within-participants ANOVA analysis of spontaneous feature production in verb SFA failed to reach significance (F (4,16) = 4.70, p =.096).

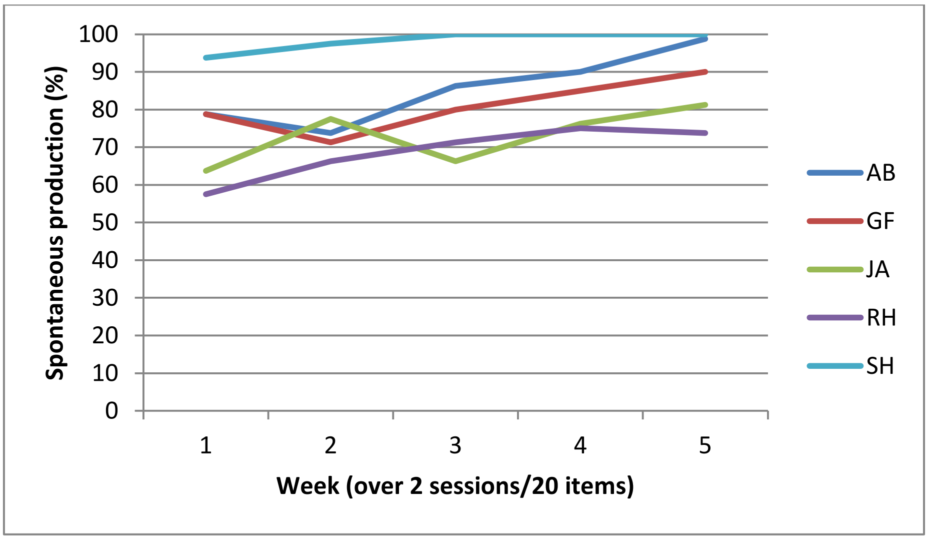

Figure 5.3 Spontaneous feature production (%) in noun-SFA therapy

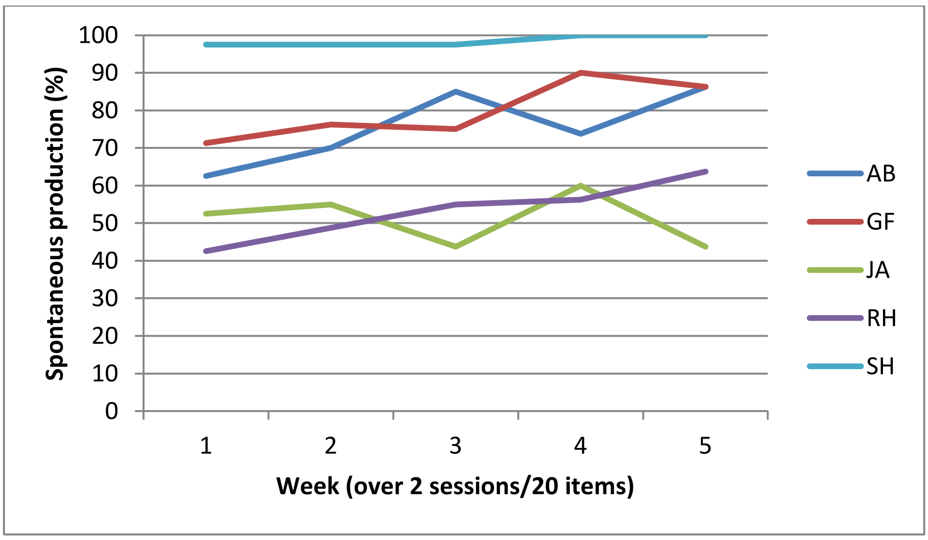

Figure 5.4 Spontaneous feature production (%) in verb-SFA therapy

Further analysis of feature production was conducted to investigate whether any particular types of semantic features appeared any easier or more difficult to spontaneously produce a response for than others. Figure 5.5 and Figure 5.6 present the percentage of spontaneous feature production across the entire therapy phase for each participant according to feature type for noun- and verb-SFA respectively. Visual inspection suggests that within noun-SFA, participants were equivalent in their feature production across feature types with the exception of the Group feature. There appears to be more variability in verb-SFA between feature types and between participants. It is particularly noticeable however that both JA and RH score below 50% in their ability to spontaneously produce features for Purpose and Description.

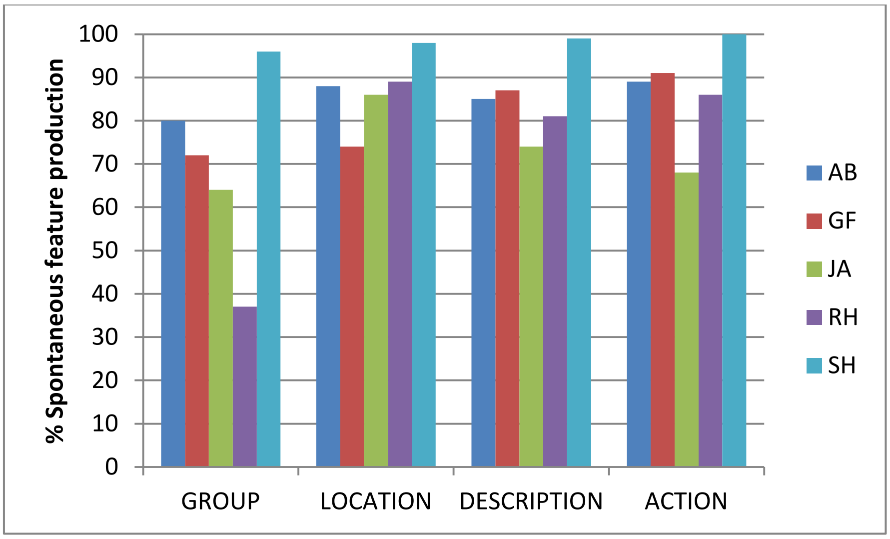

Figure 5.5 Spontaneous feature production (%) by feature type in noun-SFA therapy

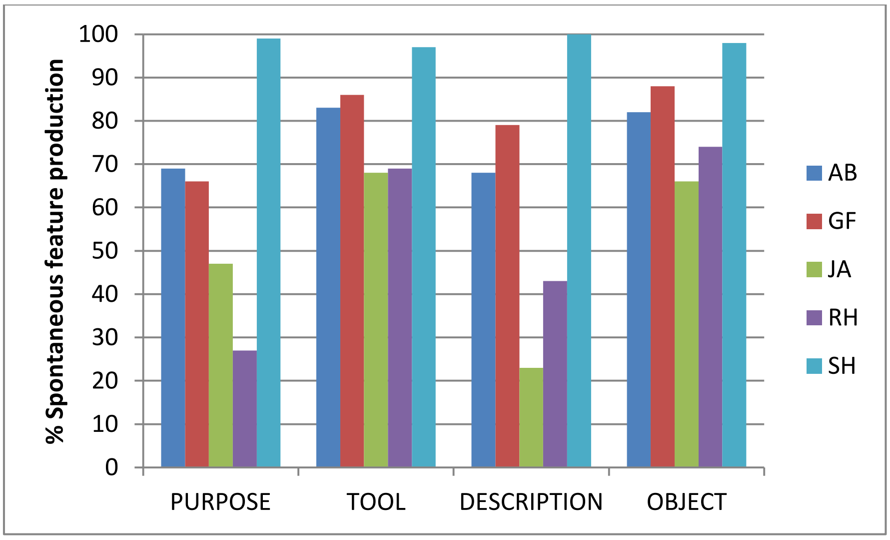

Figure 5.6 Spontaneous feature production (%) by feature type in verb-SFA therapy

5.4.2. Effect of total therapy on overall noun and verb naming

Group analysis

In a group analysis, naming scores on all items (n = 60) within each word class (noun and verb) at pre-therapy baseline 1 and post-therapy 2 were entered into a two-way within participant ANOVA with the factors time (2) and word class (2). The resultant mean scores are presented in Figure 5.7. This showed significant main effects of time (F (1,4) = 32.295, p = .005) and word class (F (1,4) = 22.479, p = .009) and also a significant interaction (F (1,4) = 29.824, p = .005) with nouns showing greater improvement than verbs. When the order of therapy (i.e. noun-verb; verb noun) was entered into the ANOVA as a between participant factor the significant main effects of time and word class and their significant interaction remained but there was no significant effect of order of therapy (F (1,3) = 6.442, p = .085) nor did order of therapy show significant two-way or three-way interactions with the other factors

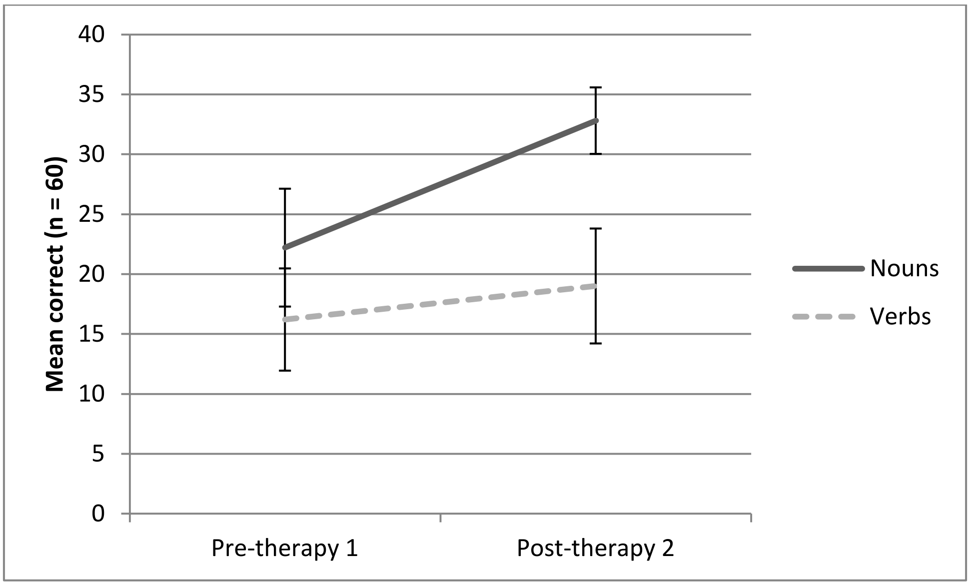

Figure 5.7 Mean correct (+/- 1 SD) noun and verb naming at pre- and post-therapy

To consider the maintenance of these observed effects, a further group ANOVA was conducted with the factors time (Pre-therapy 1, Post-therapy 2, Maintenance) and word class (noun, verb). This again showed significant main effects of time (F (2,8) = 10.176, p = .006), word class (F (1,4) = 13.384, p = .028), as well as a significant interaction (F (2,8) = 5.776, p = .028) suggesting that improvements were maintained above Pre-therapy 1 levels

Individual analysis

Detailed information regarding each participants' performance in noun and verb picture naming at all time points in the intervention study is presented in Appendix W.Figure 5.8 presents each participants (and group mean) scores in noun and verb picture naming at pre-therapy 1 and post-therapy 2. McNemer analyses suggested that of the individual scores only two comparisons reached significance: AB's pre- and post-therapy noun naming (p = .021), and SH's pre- and post-therapy noun naming (p = .011), with RH's pre- and post-therapy noun naming approaching significance (p = .054).

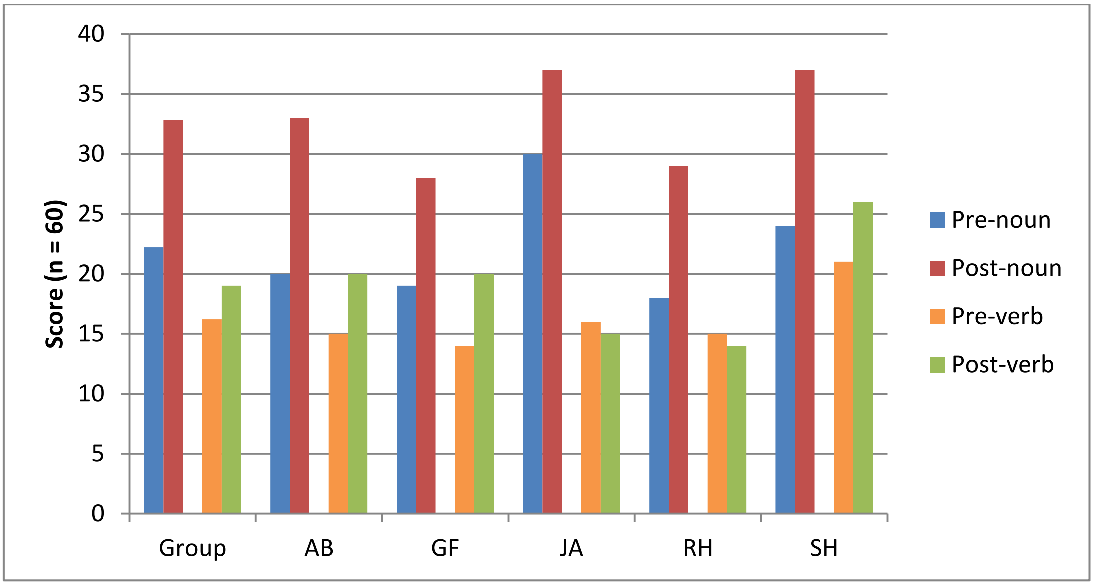

Figure 5.8 Noun and verb naming at pre-therapy 1 and post-therapy 2

5.4.3. Effect of total therapy on treated and untreated items

Group analysis

Naming performance on each set of items (i.e. treated, untreated related, and untreated unrelated) were compared within individual analyses for each word class within two-way within participant ANOVAs. The mean scores at pre-therapy 1 and post-therapy 2 for noun naming and verb naming are presented in Figure 5.9 and Figure 5.10 respectively. For nouns, there was a significant main effect of time (F (1,4) = 12.821, p = .032) but no significant effect of set (F (2,8) = 0.223, p = .805) nor a significant interaction (F (2,8) = 0.174, p = .713, Greenhouse-Geisser adjusted as the assumption of sphericity was violated). This suggested that naming of all three sets improved to a similar extent. For verbs, there was no significant main effect of time (F (1,4) = 5.565, p = .078) but there was a significant main effect of set (F (2,8) = 6.184, p = .024) and a significant interaction (F (2,8) = 12.962, p = .003) indicating that treated set verbs improved to a greater extent than both untreated sets.

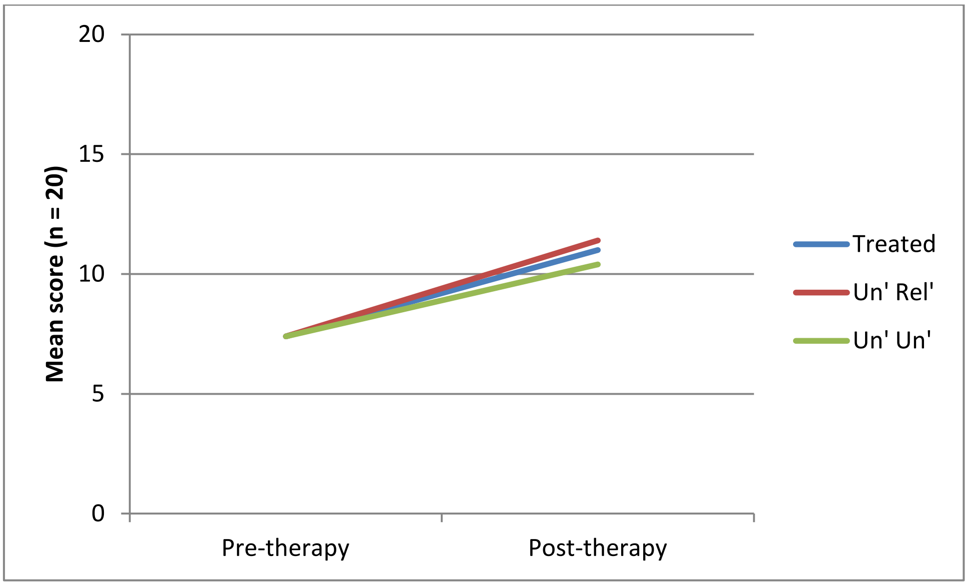

Figure 5.9 Group mean scores at pre- and post-therapy by item set - Nouns

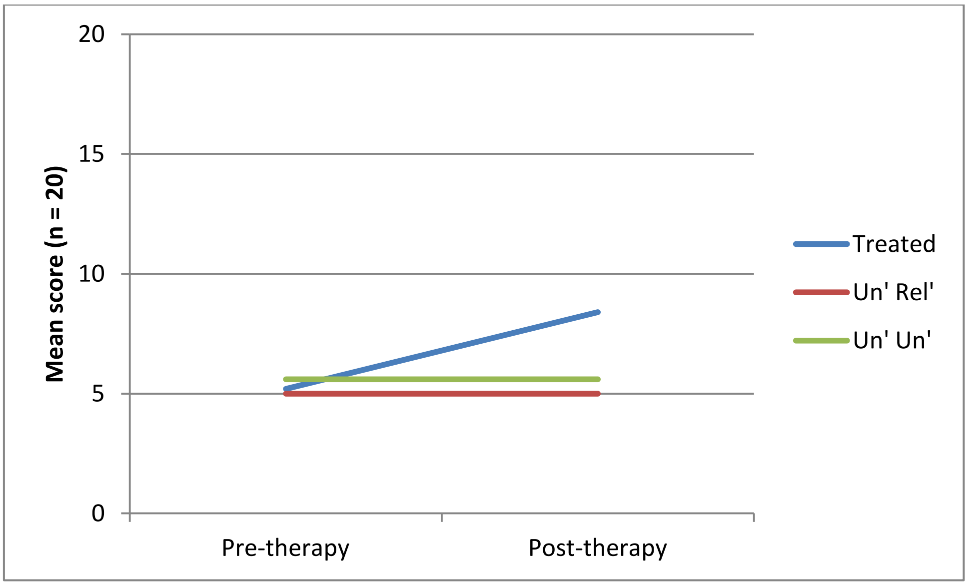

Figure 5.10 Group mean scores at pre- and post-therapy by item set - Verbs

Individual analysis

Individual analyses comparing pre-therapy 1 and post-therapy 2 scores on each item set were conducted using McNemer tests. These showed that only one comparison reached significance where RH's post-therapy naming of untreated related nouns was greater than at pre-therapy (*p* = .039). SH's post-therapy naming of treated nouns also showed a trend towards a significant increase over pre-therapy 1 performance (*p* = .07).

5.4.4. Effect of each phase of therapy

Group analysis

In addition to investigating the overall effects of therapy (i.e. the combined effect of noun-SFA therapy and verb-SFA therapy), individual analyses were conducted to investigate the effect of each phase individually. As participants had received therapy phases in different orders (i.e. noun-verb or verb-noun), naming performance was measured on assessment before and after the therapy phase of interest. For example, when noun-SFA was the first therapy phase, naming performance was considered from pre-therapy 1 to post-therapy 1. Conversely, where noun-SFA was the second therapy phase, naming performance was considered from post-therapy 1 to post-therapy 2.To aid simple interpretation, the therapy phases are renamed here as pre-therapy phase and post-therapy phase.

Figure 5.11 and Figure 5.12 present the group mean scores on both noun and verb naming preceding and following noun-SFA and verb-SFA respectively. For noun-SFA therapy, there were significant main effects of time (F (1,4) = 9.175, p = .039) and word class (F (1,4) = 12.621, p = .024) but no significant interaction (F (1,4) = 0.487, p = .524). For verb-SFA therapy, there were no significant main effects of time (F (1,4) = 2.827, p = .168) or word class (F (1,4) = 4.103, p = .133) and no significant interaction (F (1,4) = 1.584, p = .277).

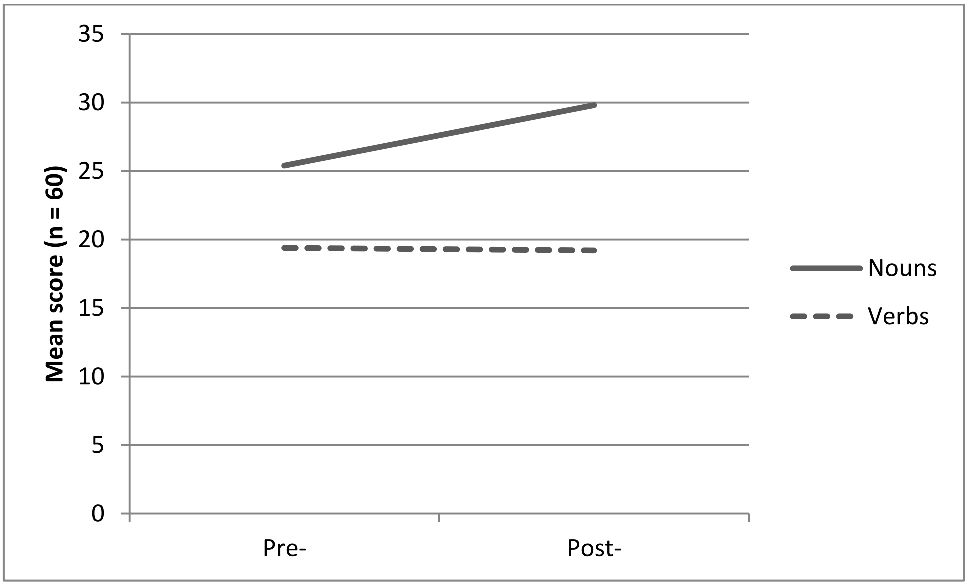

Figure 5.11 Group means on noun and verb naming pre- and post- noun-SFA therapy

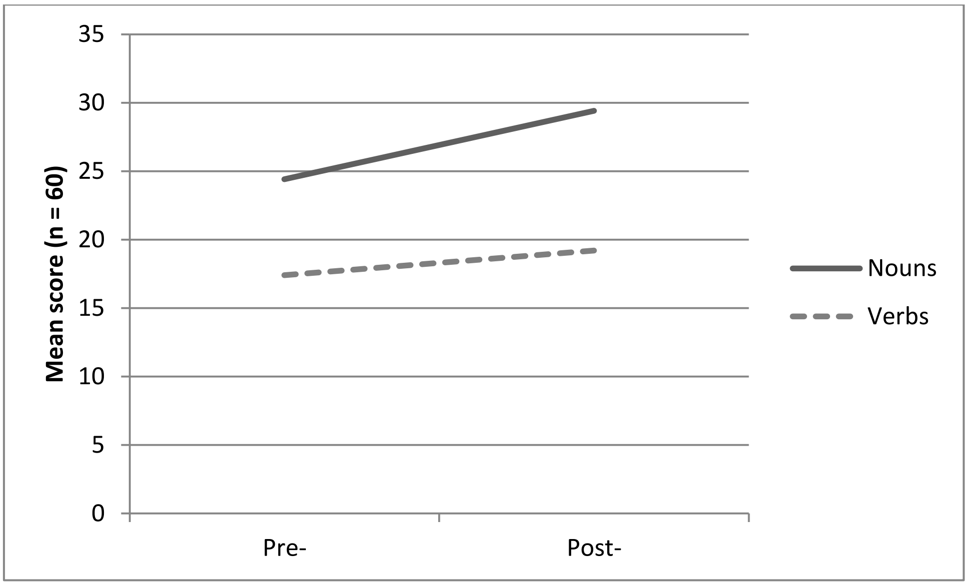

Figure 5.12 Group means on noun and verb naming pre- and post- verb-SFA therapy

Further ANOVA analysis was also carried out to investigate the effects of each therapy phase on treatment items of the congruent word class, e.g. the effects of noun-SFA on treated and untreated nouns. This employed a two-way within participant ANOVA with factors of time (2) and item set (3). The group means pre- and post-therapy phase are presented in Figure 5.13 and Figure 5.14 for noun-SFA and verb-SFA respectively. For noun SFA therapy, there was a significant effect of time (F (1,4) = 21.479, p = .01), no main effect of set (F (2,8) = 0.234, p = .797) but a significant interaction (F (2,8) = 4.872, p = .041) suggesting that the overall improvement in noun naming was attributable mostly to improvements in naming treated set nouns. For verb

SFA therapy, there was no significant main effect of time (F (1,4) = 0.434, p = .546) but there was a significant main effect of set (F (2,8) = 4.750, p = .044) and a significant interaction (F (2,8) = 9.468, p = .009) suggesting that whilst any improvements in overall naming did not reach significance, there was improvement in treated set verbs whereas there was no improvement for either untreated set.

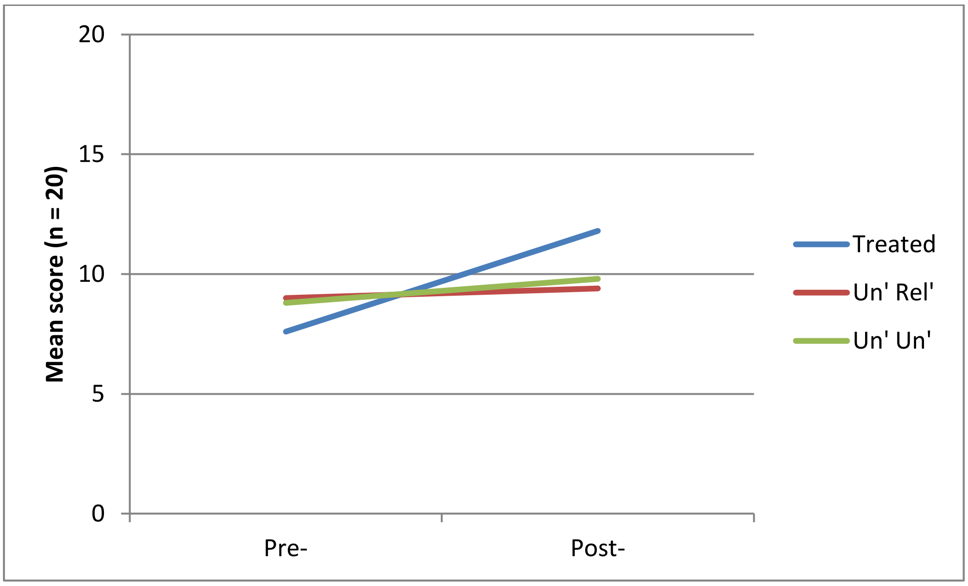

Figure 5.13 Group mean on noun item sets pre- and post- noun-SFA therapy

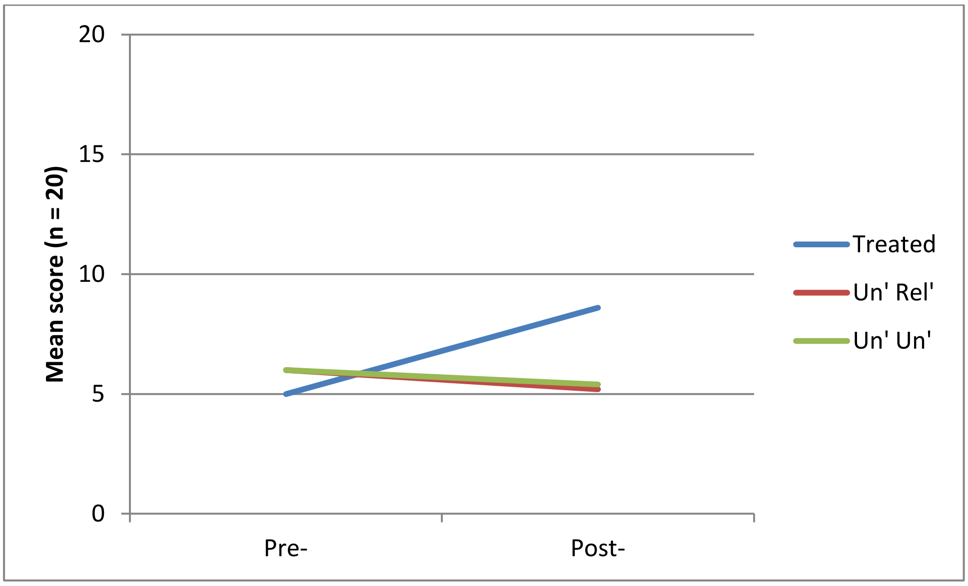

Figure 5.14 Group mean on verb item sets pre- and post- verb-SFA therapy

Individual analysis

As with group analysis, each individual participant's performance on noun and verb naming was compared before and after each separate phase of therapy (i.e. noun-

SFA and verb-SFA). This data (and group means) are presented in Figure 5.15 (noun-SFA) and Figure 5.16 (verb-SFA). In noun-SFA, no comparison of naming performance on either noun or verb naming reached significance or showed a trend towards significance. In verb-SFA, only one comparison showed a significant difference: RH's pre- and post-therapy verb naming (p = .031). Other comparisons showed trends towards significance: AB's pre- and post-therapy noun naming (p = .093); and GF's pre- and post-therapy verb naming (p = .065).

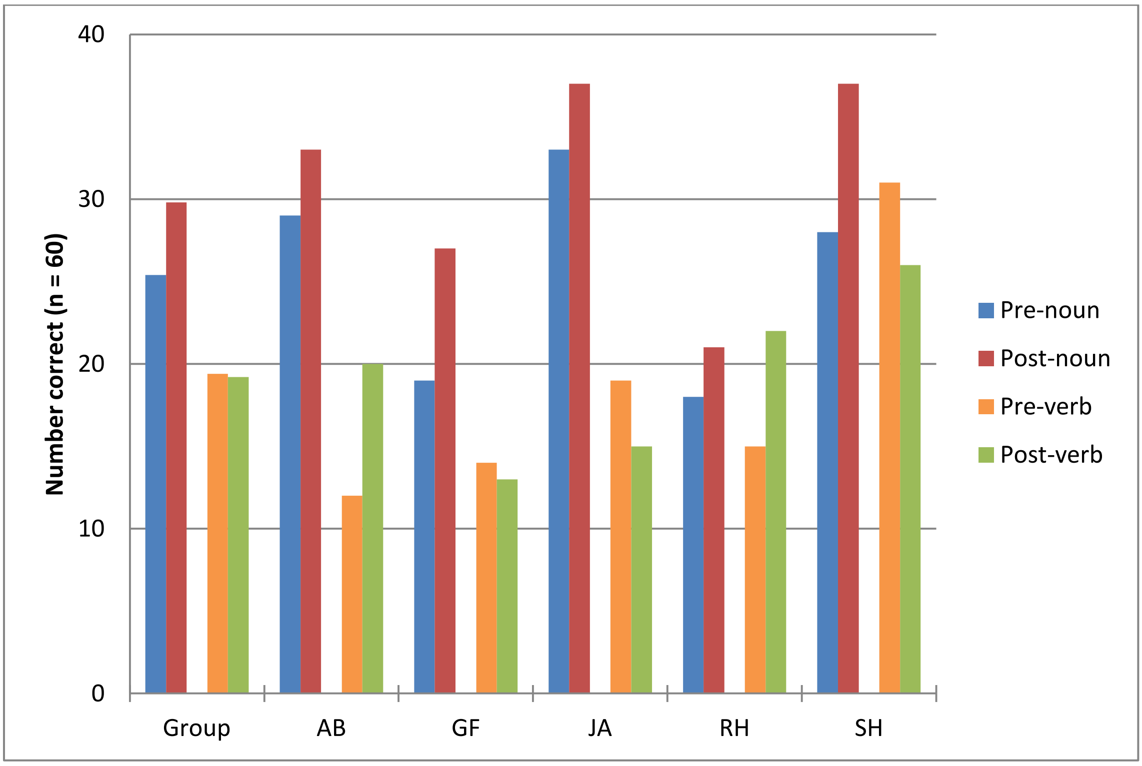

Figure 5.15 Individual (and group) noun and verb naming pre- and post- noun-SFA

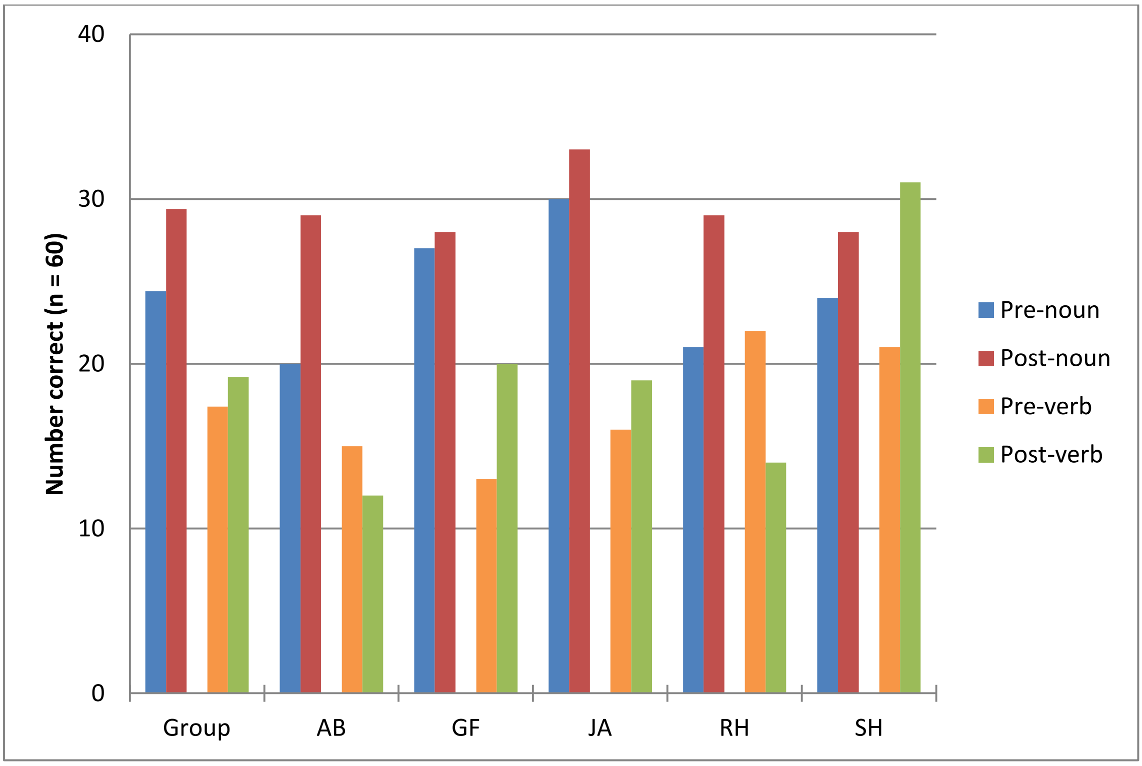

Figure 5.16 Individual (and group) noun and verb naming pre- and post- verb-SFA

5.4.5. Effect of total therapy on independent measure of object and action naming

Group and participants' performance on the OANB at pre-therapy and post-therapy were subjected to analysis in order to measure any improvements in noun and verb naming that may not be associated with increased familiarity and exposure to the same picture stimuli. In group analysis, in order to account for the different numbers of test items in the Object and Actions subtests, proportion correct scores were entered into ANOVA analysis rather than raw scores. Data were entered into a two-way within participant ANOVA with the factors of time (Pre-therapy, Post-therapy) and subtest (objects, actions). This revealed no significant effect of time (F (1,4) = 1.646, p = .269) but a significant overall effect of subtest (F (1,4) = 18.252, p = .013) and a significant interaction (F (1,4) 9.245, p = .038). Mean object naming improved from 60% correct at Pre-therapy 1 to 67% correct at Post-therapy 2, whereas action naming remained stable at 46% correct.

Individual participants' score on the OANB were compared using McNemer tests (see Table 5.4). Two participants showed significant improvement in object naming whereas no participant showed significant change in action naming.

	Object naming (*p* =)	Action naming (*p* =)
AB	.358	.735
GF	.004*	.571
JA	.228	.106
RH	.112	.635
SH	.001**	.710

Note.: * p < .05; ** p < .01

Table 5.4 Individual Pre- and Post-therapy comparisons on OANB naming)

In addition to overall naming accuracy, an analysis of each participant's error distributions was performed using chi-squared analyses (see Appendix X for breakdown of error distributions). The general error classifaction system specified by Mätzig et al (2009) was used as a guide with some slight adaptations. In general, errors that would be classified as misinterpretations of the picture stimuli were reclassified here as lexical errors (e.g. where a participant gives a noun response when the desired response was a verb and vice versa). This was done on the basis that it would be difficult to interpret whether it was a true misinterpretation or whether participants gave a lexical error in the face of being unable to retrieve the appropriate word class. Secondly, 'frank visual' errors were infrequent and were classified as a type of semantic error. Therefore, the error analysis here employed broader error categories of: semantic errors (including circumlocutions and frank visual errors), phonological errors, lexical errors, other errors (e.g. unrelated errors), and no-responses. However, in order to ensure reliable analysis, for different participants, different error classifications were combined when these error types were uncommon. Table 5.5 presents the chi-squared statistics for each participant's pre- and post-therapy comparisons of error distributions in both the object naming and action naming subtests of the OANB.

Object subtest				
	Error categories	χ^2	df	p
AB	sem, lex, phon/other, NR	10.61	3	.013*
GF	sem, lex, phon, other, NR	3.28	4	.512
JA	sem, phon, lex, other/NR	2.27	3	.518
RH	sem, lex, other/NR	6.62	2	.036*
SH	sem/lex, phon, other/NR	1.09	2	.594

Action subtest				
	Error categories	χ^2	df	p
AB	sem, lex, phon/other, NR	14.43	3	.002**
GF	sem, lex, other/NR	0.39	2	.842
JA	sem, lex, phon/other, NR	0.64	3	.887
RH	sem, lex, other/NR	0.92	2	.623
SH	sem/lex, phon/other/NR	2.96	1	.060

Note.: sem – semantic; lex – lexical; phon – phonological, NR – no-response

Table 5.5 Chi-squared analyses of OANB error patterns pre- and post-therapy

The only participant that showed widespread changes in error patterns was AB with significant changes in both object and action subtests. These changes were generally accounted for in a reduction in no-responses (i.e. from 22/56 total errors pre-therapy to 6/48 total errors post-therapy) and increases in lexical and/or semantic-type errors. RH also showed a significant change in object naming with a reduction in other/no-response errors (i.e. from 47/90 total errors pre-therapy to 25/77 total errors post-therapy) and increase in semantic and lexical errors. SH also showed a trend toward a significant change in action naming with a decrease in phonological/other/no-response errors and an increase in semantic/lexical errors.

5.4.6. Effect of total therapy on sentence processing

As participants varied on their baseline performance on sentence comprehension and production assessment, both in terms of items correct and error patterns, only individual level analyses were conducted on sentence processing assessments (i.e. SCAPA). In assessment of sentence comprehension, no participant showed significant improvement from pre- to post-therapy on either, number of sentence items correct, nor number of verbs correctly identified (McNemer test, $p < .05$). Various analyses were conducted on sentence production on the SCAPA assessment. No participant showed a significant improvement in total items correct, total number of correct verbs produced, total number of correct subject nouns produced, nor total number of correct object nouns produced (McNemer test, $p < .05$). Similarly, no participant showed any significant reduction in number or innapropriate substitutions or omissions of verbs, subject nouns, or object nouns (see Appendix Y outcome data).

5.4.7. Effect of total therapy on control measure and other language assessment

During the post-therapy phase, all participants were reassessed on their ability to repeat digits. Only GF showed any improvement in this assessment with an increased digit span of 1 item (i.e. 3 digits repeated at pre-therapy and 4 repeated at post-therapy). All other participants' post-therapy assessment was consistent with their pre-therapy assessment (i.e. AB: 2; JA: 4; RH: 3; and SH: 2).

Other language assessments were re-administered during the post-therapy phase according to each individual participant's pre-therapy performance (i.e. potential for improvement needed to have been present; no ceiling effects at pre-therapy). AB showed a significantly improved ability to read written words aloud (McNemer test, $p = .012$) but no improvement in the three-word version of the KDT ($p = .289$). GF showed no improvements in any re-assessment. JA showed a trend towards significant improvement in the PPT three-picture version ($p = .070$) but no concomitant improvements in the PPT three-word subtest ($p = .581$) nor the KDT three-picture subtest ($p = .146$). RH showed a trend towards significant improvement in the PPT three-word subtest ($p = .077$). SH showed a trend towards a significant improvement in reading written words aloud ($p = .057$) but no improvement in repeating spoken words ($p = .109$).

5.5. General Discussion

5.5.1. Summary of main findings

The intervention study reported here showed a number of interesting findings. Firstly, SFA as a therapy approach was effective in improving retrieval of nouns and verbs. Secondly, SFA therapy was generally effective at promoting generalised naming improvement in noun naming but not for verb naming, although there was individual variation observed between participants. Improvement in noun naming was also observed in an independent measure of picture naming (i.e. OANB) with no concomitant improvement observed in verb naming. Thirdly, the individual noun-SFA phase of therapy was only effective in consistently improving naming of treated nouns, although there was variation in patterns of improvement between participants. The individual verb-SFA phase of therapy was effective in improving naming of treated verbs and also untreated nouns (i.e. untreated within the verb phase) and this pattern was relatively consistent between individual participants. Finally, no participant showed any improvement in ability to understand or produce sentences as a result of the combined noun- and verb-SFA therapy phases.

5.5.2. Discussion of main findings

SFA was effective in improving noun and verb naming

The finding that SFA was effective in improving noun and verb naming is generally consistent with the majority of previous studies that have used SFA and other semantic-based therapy approaches (Boyle & Coelho, 1995; Wambaugh & Ferguson, 2007). This was expected given that the therapy approach necessitated aspects of semantic and also phonological processing (i.e. through verbal production) and would be somewhat applicable for the majority of participants with aphasia who experience word-finding difficulty as a main characteristic of aphasia. It is worth noting however, that the current results were found when participants were asked to generate fewer semantic features than is traditionally the case (i.e. four features as opposed to six).

There were two exceptions to this general finding with both JA and RH showing a reduction in their overall verb naming following the second therapy phase (i.e. JA: 16/60 pre-therapy, 15/60 post-therapy 2; RH: 15/60 pre-therapy, 14/60 post-therapy 2). What is important to point out is that both JA and RH were variable in verb naming across the duration of the study. JA also scored 19/60 at post-therapy 1 (i.e. immediately following verb-SFA) and 17/60 during the maintenance phase. This is in contrast to his noun naming performance which showed gradual improvement at each subsequent re-assessment phase. RH scored 22/60 correct at post-therapy 1 (i.e. immediately following noun-SFA) and then 16/60 correct during the maintenance phase. The improvement in verb naming following noun-SFA, when this had been the only therapy received, is particularly difficult to reconcile. Given that this followed noun-SFA, it may be hypothesised that improvement actually reflected an improved ability to retrieve nouns that were homophonous with the target verbs. While this is possible, only 13/22 correct items were produced in progressive form of the verb and, of the remaining 9 correct items that were produced as uninflected forms, few of these had target pictures which depicted an object which could be named as a homophonous noun (i.e. *push*, *cuddle*, *flush*, *rip*, *knock*). Those where noun naming may have been a stronger possibility included those where an object may have been named with a homophonous noun (e.g. *glue*, where the picture showed a boy using a tube of glue; *drink*, where the picture showed a man having a drink). Of the five participants, JA and RH arguably showed the greatest impairments to semantic processing (i.e. in CAT, PPT and KDT assessments) which may account for the observed variability as their performance may reflect poor ability to comprehend the picture stimuli and subsequent identification of corresponding semantic representations. Identifying such variable performance within re-assessment is perhaps the risk of only probing naming following the ends of phases of therapy as opposed to continual probing throughout therapy phases. JA's variable performance may also be accounted for by his performance often being affected by tiredness and frustration. The level of difficulty he experienced with the activity was exacerbated by his insight into his errors, particularly in tasks requiring language production without the opportunity for corrective feedback from the researcher (i.e. during assessment).

SFA facilitated generalised improvement of noun naming but not verb naming

The finding that SFA facilitated improved naming of untreated nouns is consistent with previous literature reporting the use of SFA therapy targeting nouns (Boyle & Coelho, 1995; Lowell et al, 1995). However, the current study suggests that the majority of this generalisation only occurred as a consequence of verb-SFA or the combined effect of noun-SFA and verb-SFA. Therefore, it may be speculated that if only noun-SFA had been conducted, then limited generalisation to naming untreated nouns would have been observed. The finding that no generalised improvement was observed with verb naming following either therapy phase is also consistent with previous reports of verb-SFA but also for other intervention studies using differing therapy approaches to remediation of verb retrieval (e.g. Faroqi-Shah & Graham, 2001; Wambaugh & Ferguson, 2007).

In considering these findings of generalised naming improvement, it is important to consider how the SFA therapy protocol may have impacted on this. As previously highlighted, the current study overcomes one of the limitations of previous studies as naming performance was only probed at the end of therapy phases and not throughout therapy. Another confound is however possible within the protocol as it is likely, through the production of semantic features, that the therapy protocol is biased towards production of nouns as opposed to verbs thereby increasing the likelihood of generalisation to noun naming. A post hoc analysis was conducted to investigate this assumption. Within this analysis, only first responses to each feature were considered (i.e. participants occasionally gave more than a single response, especially as the protocol became more familiar) as these responses were given as part of the protocol (e.g. some second responses were given after all five features were given and naming of the target had occurred). These were considered regardless of whether they were produced spontaneously or repeated following a forced-choice alternative from the researcher. These responses were either in the form of single words or phrasal responses. In single word responses, the word was coded according to its word class (i.e. noun, verb, or adjective/adverb; no other word classes were considered in this analysis). In the case of phrasal responses, all open class words were coded (e.g. a response such as *hot countries* was coded as two items: adjective/adverb and noun; similarly *mixing the cement* was coded as a verb and a noun). The only exception to this was compound nouns that were composed of two nouns (e.g. *kitchen table*, *feather duster*) which were coded as a single noun and non-specific items (e.g. *something*, *someone*, *thing*) which were excluded completely. Therefore, the minimum number of words that a participant could be expected to give as features within an individual therapy phase was 400 (i.e. four features for each of 20 target items that were exposed a total of five times).

Across all participants the mean total words given as semantic features was 450.2 (SD = 28.7) within noun-SFA and 449.2 (SD = 46.8) for verb-SFA. Interestingly, verb-SFA was significantly more effective at generating a wider variety of words (i.e. unique features) across the complete therapy phase with a mean of 133 (SD = 16.2) compared to noun-SFA where participants generated a mean of 111.6 (SD = 16.3) unique features (paired t-test: t (4) = -3.31, p = .03).

Across both therapy phases, the total proportions of words produced as nouns, verbs and adjectives was generally consistent between participants, ranging from 54-64% for nouns, 20-38% for verbs, and 8-19% for adjectives/adverbs. These proportions were also similar within each individual therapy phase (See Appendix Z), confirming the assumption that the SFA protocol as a whole was biased towards production of nouns.

A further factor that may influence the patterns of generalisation observed may be the extent to which features produced by participants overlapped with their treatment stimuli (i.e. treated and untreated items). A further analysis was therefore conducted. Nouns and verbs that were given as features five or more times throughout the entire therapy duration were compared to each participant's noun and verb treatment stimuli. The criterion of five productions for the feature was chosen as this mirrored the minimum number of times each treated item was produced within a therapy phase. Overall, across participants there was generally very little overlap of features produced and treatment stimuli and, interestingly, there was greater overlap with regard to verbs given as features and verb treatment stimuli (Nouns: M = 0.8, SD = 0.45; Verbs: M = 2.8, SD = 1.48). Therefore, while the SFA protocol used here did generate a greater number of nouns than verbs, this cannot directly account for the degree of generalisation observed within noun naming as there was little overlap between the nouns given as features and the nouns probed in naming of untreated items.

The words selected as untreated unrelated items were generally higher in lexical frequency that either the treated or untreated related items which may make it easier to affect change if such items are more frequently heard outside of the therapeutic context. However, this was the same for both noun and verb item sets and there was no significant difference in mean frequencies of nouns and verbs used in these sets across all items across all participants (noun frequency: M = 4535.1, SD = 6769.04; verb frequency: M = 4731.44, SD = 6313.25; independent t-test: t = (198) = -0.212, p =.832). It should be noted however that noun frequency was based on cumulative frequency of singular and plural forms and verb frequency was based on cumulative frequency of infinitive, past tense, progressive, and third person forms rather than base forms for both nouns and verbs. However, overall, it appears that higher lexical frequency on its own cannot account for the generalisation observed as if improvements were observed in untreated unrelated nouns it may be expected that untreated unrelated verbs may also improve to some extent.

A frequent explanation for widespread generalisation effects following SFA-types of therapy is that it gives the speakers a strategy to employ on occasions when they experience word-finding difficulty. In order to investigate whether this is likely to be a possibility, an analysis of non-target responses within the 60 item noun and verb naming outcome measures was performed. If SFA is internalised as a strategy for word-finding it might be expected that participants increase in the numbers of target responses that are given following initial non-target responses, whether they are preceded by a delay of greater than 10 seconds (i.e. the criteria for scoring a target response correct), or if they are preceded by other non-target responses (e.g. when AB responded '*fruit, oranges, apple*' for the target *apple*). Therefore, responses that could be identified as 'self-corrections' may be expected to increase as a percentage of the total number of errors as therapy progressed (given also that the total number of errors would be expected to decrease). It may also be expected that the total percentage of no-responses may decrease as the participant may attempt to self-generate semantically related features in efforts to self-cue target responses. Table 5.6 and Table 5.7 present this information for each participant at each phase of therapy (pre-therapy 1 and pre-therapy 2 errors are collated). Within both noun and verb picture naming measures, there is a great amount of variability between participants and also within participants as therapy progressed. Although there is limited evidence here that participants improved in a strategic use of SFA in the face of word-retrieval difficulties, this also needs to be considered in light of the overall improvements in noun and verb naming. However, presumably, if noun and verb naming has improved within first responses that are given within 10 seconds of picture presentation, this more likely reflects improvements in automatic semantic and/or lexical access to target representations rather than implementation of a 'silent' strategy (i.e. an internal *conscious* generation of semantically related features before successful naming).

		Pre-therapy (total n = 120)	Post-therapy 1 (n = 60)	Post-therapy 2 (n = 60)	Maintenance (n = 60)
AB	Total errors	80	31	27	31
	% self-corrected	31	26 (-5)	41 (+10)	29 (-2)
	% no-response	10	16 (+6)	8 (-2)	16 (+6)
GF	Total errors	80	33	32	42
	% self-corrected	18	24 (+6)	13 (-5)	17 (-1)
	% no-response	0	0 (+0)	0 (+0)	2 (+2)
JA	Total errors	60	27	23	17
	% self-corrected	10	7 (-3)	9 (-1)	29 (+19)
	% no-response	13	7 (-6)	17 (+4)	6 (-7)
RH	Total errors	83	39	31	33
	% self-corrected	16	10 (-6)	7 (-9)	33 (+17)
	% no-response	28	18 (-10)	32 (+4)	27 (-1)
SH	Total errors	76	32	23	18
	% self-corrected	40	28 (-12)	48 (+8)	33 (-7)
	% no-response	5	9 (+4)	0 (-5)	0 (-5)

Table 5.6 Participants' % self-corrected errors and no-responses in noun picture naming (+/- difference from pre-therapy)

		Pre-therapy (total n = 120)	Post-therapy 1 (n = 60)	Post-therapy 2 (n = 60)	Maintenance (n = 60)
AB	Total errors	90	48	40	39
	% self-corrected	9	4 (-5)	5 (-4)	8 (-1)
	% no-response	4	2 (-2)	10 (+6)	5 (+1)
GF	Total errors	90	47	40	41
	% self-corrected	10	15 (+5)	15 (+5)	2 (-8)
	% no-response	0	0 (+0)	0 (+0)	0 (+0)
JA	Total errors	88	41	45	43
	% self-corrected	1	10 (+9)	0 (-1)	5 (+4)
	% no-response	13	0 (-13)	11 (-2)	0 (-13)
RH	Total errors	90	38	46	44
	% self-corrected	9	0 (-9)	4 (-5)	5 (-4)
	% no-response	9	8 (-1)	7 (-2)	7 (-2)
SH	Total errors	76	29	34	28
	% self-corrected	18	21 (+3)	27 (+9)	18 (+0)
	% no-response	0	0 (+0)	0 (+0)	0 (+0)

Table 5.7 Participants' % self-corrected errors and no-responses in verb picture naming (+/- difference from pre-therapy)

As with overall improvements in naming performance, there were exceptions in the patterns of generalisation demonstrated by individual participants and their effect sizes (although very few comparisons were statistically significant given the small sample sizes of n = 20). Within noun picture naming, the majority of participants performed consistently in line with the group patterns (i.e. improvement in all item sets). Within noun naming, GF showed a non-significant reduction (*M pre-therapy* = 7/20 correct, *post-therapy 2* = 6/20), whereas RH also showed a non-significant decline in naming untreated unrelated nouns (*M pre-therapy* = 7/20 correct, *post-therapy 2* = 6/20). All participants showed positive effect size improvements for naming untreated related nouns. Within verb naming, all participants showed positive effects in naming treated verbs, again following the group pattern. Three out of five participants actually showed positive effects sizes (although no statistically significant improvement) in naming untreated related verbs while the remaining two showed declines. For untreated unrelated verbs, three participants showed absolutely no change while one showed positive change (SH) and one showed negative change (RH). RH's performance may again be reflective of his poor comprehension but these specific results do demonstrate that he responded positively to the items that were explicitly targeted in therapy phases, which in his case may have been to the detriment of untreated items, although with the interesting exception of untreated related nouns where his performance was consistent with the group. Some particularly noteworthy individual performances include AB's naming of untreated related verbs which improved from pre-therapy (*M* = 5/20) to post-therapy 2 (9/20) but should also be noted that this declined back to baseline level during the maintenance phase. Therefore, this may potentially be an indication, albeit a weak one, of some potential for positive improvement in verbs that are not treated in therapy. GF's patterns of improvement are also interesting given that he demonstrated equal positive effect sizes in noun naming across item sets despite being hypothesised to have a fairly intact semantic system. This may be taken as evidence that SFA is effective regardless of whether an individual suffers semantic impairment of not, although JA and RH's results may provide evidence that there is a point where severity of semantic impairment may limit the effectiveness of SFA-type therapies.

Noun-SFA and verb-SFA led to different patterns of improvement

This is potentially one of the most significant findings from the current study both from a theoretical perspective and a clinical perspective. This finding implies that conducting SFA with verb targets is more effective on improving general word-retrieval than SFA with noun targets. As highlighted above, both noun- and verb-SFA led to almost identical numbers of total words produced as semantic features but verb-SFA was significantly more effective in eliciting a greater variety of words (*M* = 133 unique features versus *M* = 111.6 unique features in verb- and noun-SFA respectively, and also with comparable standard deviations between participants). Again, given that there was little overlap in feature production and treatment items, this cannot directly account for the findings with regard to generalisation. Therefore, there must be something particular to verb-SFA that makes it more effective, especially when considering that, if target items are included within responses, noun-SFA would have an even higher proportion of noun responses relative to verb responses and verb-SFA would have an even higher proportion of verb responses relative to noun responses.

Semantic feature production with noun targets may be a relatively straightforward task as features may be more strongly associated than they are with verb targets. This would undoubtedly be the case with some features such as the Group feature which aimed to elicit superordinate category information (e.g. aiming to elicit *fruit* for *apple*, *tool* for *hammer*, and so on). This is generally supported by participant's performance within noun-SFA as these responses did tend to be consistent as the therapy phase progressed. In comparison, the purpose feature was included in verb-SFA as a hypothesised parallel to the Group feature in noun-SFA (i.e. aiming to elicit *breaking/as a way of breaking something* for *bending*, *cooking/as a way of cooking something* for *boiling*, and so on). However, as may be expected, this was not so straightforward and there was greater variability which may reflect the variety of contexts in which an action may occur and greater variety in terms of the thematic participants (i.e. objects/nouns) that can take part in the actions (e.g. for the three participants who had baking as a treated item, the Purpose was variously given as *cooking*, *making a cake/something to eat*, and *for someone's birthday*).

It may be that verb-SFA encourages greater or deeper semantic activation than noun-SFA as there is greater potential overlap and involvement with syntactic and thematic level information. While both noun-SFA and verb-SFA involve association between related nouns and verbs that could be within a thematic relation (i.e. through target noun and related action feature in noun-SFA and target verb and related object feature in verb-SFA), it may that this is more explicit within verb-SFA as the verb forms the core aim of target picture naming. This is not necessarily supported by the finding that there was limited overlap in feature production and treatment stimuli and it may be a fairly subtle distinction between the two forms of SFA used here, however, it seems intuitively plausible that it would be easier and more communicatively meaningful to generate nouns when given a target verb (i.e. verb-SFA) than verbs when given a target noun (i.e. noun-SFA). This potential for greater and/or deeper semantic activation may also be parsimonious with the view that increased complexity of therapy items leads to more effective therapy in terms of greater generalisation (e.g. Thompson, Shapiro, Kiran & Sobecks, 2003), in particular with the findings that therapy targeting atypical category members is more effective at producing within category generalisation than therapy targeting typical category members (e.g. Kiran & Thompson, 2003). Explanations for such findings generally suggest that atypical category members lead to the generation of a more diverse range of semantic features which causes more widespread semantic activation which may also be applicable in the current study as verb-SFA led to greater elicitation of unique semantic features than noun-SFA.

A further demonstration of the apparent difficulty with verb-SFA came with the analysis of spontaneous feature production which was generally lower and more variable than compared with noun-SFA. This presents a further interesting insight given that participants were generally more successful in noun-SFA yet the improvement from this was relatively restricted. Therefore, it may be hypothesised that the repeated success – and possible lack of significant challenge – of producing a fairly restricted set of semantic features led to improvement restricted to those items around which the task was based.

Whilst verb-SFA does appear to be a more difficult process than noun-SFA, there is evidence here to suggest that verb-SFA is more beneficial to improving overall word-retrieval than noun-SFA. The difficulty for the participant arises from the more abstract and variable nature of associations between target verbs and their semantic features. There are also additional difficulties that the clinician/researcher must face when administering verb-SFA as opposed to noun-SFA. These include the problems associated with identifying easily interpretable and nameable pictures of actions and also in validating the appropriateness of semantic features for target verbs, both when participants are self-generating but also when the clinician/researcher is offering forced-choice alternatives on occasions when participants are unable to self generate – a situation that arises more frequently in verb-SFA than noun-SFA.

SFA did not affect change in sentence processing abilities

A small number of previous studies have suggested some improvement in sentence production abilities following SFA although the gains have tended to be fairly modest (Boyle & Coelho, 1995; Wambaugh & Ferguson, 2007). The reason for hypothesising that SFA may lead to changes in sentence production lie with the overlap that SFA shares with sentence level therapies which involve raising meta-linguistics awareness of how predicate argument structure, and particularly thematic role assignment to arguments (e.g. Marshall, 2002; Mitchum, Greenwald & Berndt, 2000). Within noun-SFA, participants are generating actions (i.e. through the related action feature) when given a noun (i.e. a thematic role candidate) and within verb-SFA participants are generating thematic role candidates through the related object feature when given an action. The crucial difference between SFA and sentence level therapies is that SFA is not explicitly targeting meta-linguistic awareness of sentence level processing components, but rather components of individual words/concepts and there is no subsequent integration into a realised sentence frame (i.e. no verbal production of sentences using the constituent components). This may be one reason why there was an overall lack of improvement in sentence processing from pre- to post-therapy across all participants.

Another issue to consider is that a number of participants may have had additional sentence processing difficulties that were not necessarily identified within the background assessment conducted as part of the current investigations. With the exception of SH, all other participants showed severe difficulty in sentence construction from both syntactic and thematic perspectives and hence it was difficult to ascertain the precise nature of their difficulties without further assessment (e.g. grammaticality judgement, ability to construct sentences when given the individual words, and so on). However, even SH, who did show some ability to construct syntactically complete sentences, subsequently failed to show any significant improvement in sentence processing, including retrieval of appropriate verbs (her noun retrieval was already mostly accurate).

One further factor to consider may be that the assessment used to measure change in sentence processing may not have been sensitive to any changes that did occur. This is a valid argument given that the SCAPA looks at sentence comprehension and production around a restricted set of 10 different verbs and also a small set of different nouns (as the primary goal of the assessment is to differentially diagnose impairments in thematic role assignment, predicate argument structure, and thematic mapping, and not word retrieval per se). A more effective method of measuring change would likely have been in the collection of more spontaneous-type speech samples (e.g. Cinderella recall; Saffran, Berndt, Schwartz, 1989; Webster, Franklin & Howard, 2007) or other picture description tasks as in previous reports of SFA therapies. In fact, various other speech samples were also collected for each participant (but not reported here). These would have ideally included narrative retelling for all participants but this was beyond the capabilities of three of the current participants whose spontaneous output without visual stimuli (e.g. picture, written words) was extremely poor (i.e. AB, GF, and RH). Therefore, for these participants, picture descriptions samples were obtained. Across the samples that were obtained there was again no evidence of improved ability to retrieve words or ability to integrate words into syntactic frames, although in general the samples are again likely to have been too small and insensitive to adequately identify changes.

5.5.3. Limitations and further research

The suggestion that verb-SFA's greater effectiveness at promoting generalisation due to it facilitating production of a more varied set of semantic features appears to be parsimonious with the Complexity Account of Treatment Efficacy (CATE; e.g. Thompson el al, 2003). In particular, previous reports of intervention to remediate word retrieval difficulties have argued that generalisation occurs from treated atypical items to untreated typical items (within the same category) and not vice versa because atypical items enable therapy tasks to raise awareness of a greater range of semantic features (e.g. Kiran & Thompson, 2003). Typicality was not controlled for in the current intervention study as the primary concern with selecting verb stimuli was to have unambiguous picture stimuli which did not lend itself to selecting verbs according to typicality ratings previously gathered (see Chapter two). As a consequence, typicality was also not controlled for nouns. Therefore, it would be an insightful comparison as to the effectiveness of SFA therapies for nouns and verbs when typicality is controlled within treated and untreated sets and whether this does have implications for the diversity of feature production as part of the therapy protocol and patterns of improvement in picture naming. Such a design is likely to prove challenging for verb-SFA therapies given the restrictions mentioned above (i.e. selecting unambiguous pictures), although this may be easily overcome if target pictures are substituted for written words to be read aloud and with the focus of the task very much on feature generation. This again would present a problem in how to measure the outcomes, as picture naming would presumably still be the outcome measure of choice, which is then not a comparable skill to that being practiced within the therapy. However, if SFA is hypothesised to be affecting change within an impaired semantic system in the face of relative preservation of other processing components, the cross-modality nature of such a task should still be sensitive to improvements in semantic processing.

A complication with the current design and the finding that each therapy phase was associated with different effects is the fact that the two phases of therapy were continuous with no interval between (apart from reassessment of naming). These results therefore assume that there was no carry-over and no ongoing change following previous therapy phases which may impact on the outcomes of a second therapy phase. For example, if a participant undergoes noun-SFA first, improvements that may be a consequence of this phase may not be observed when naming is reassessed immediately following the end of this phase but may be present at the following assessment point (reassessment occurred generally 5 days following the end of the first therapy phase). An extended period of consolidation with no intervention (i.e. therapy or assessment) would have been insightful as it could have been assumed with more validity that participants' improvements in naming were a direct consequence of improved semantic and language processes and not any kind of extended facilitation effects associated with exposure to familiar pictures.

A further limitation, as is frequently the case with intervention studies, is the number of participants and the range of impairments and severities involved. Although the current study is more robust than many other accounts of SFA therapy (e.g. Boyle, 2005; Boyle & Coelho, 1995; Wambaugh & Ferguson, 2007), the number was still not large enough to ensure a wide spread of impairments. For example, all participants included here had hypothesised impairment in accessing phonological representations from semantics while only three out of four were confidently hypothesised to have impaired semantics. This is reflective of the fact that aphasia and language impairments have a tendency to implicate multiple processing systems and components and relatively 'pure' impairments are fairly rare. However, from the point of view of the current study, it would have been insightful to recruit participants who showed relatively pure semantic impairment, and also a wider range of severity to allow investigation of whether outcomes correlate with severity of impairment (e.g. as in Conroy et al 2009a, 2009b; Raymer et al, 2007).

5.6. Conclusions

This intervention study has demonstrated that improvement in both noun and verb retrieval can be facilitated with the use of a semantically-based therapy approach. While there are undoubtedly opportunities within the task to strengthen access links from semantic representations to phonological representations, the overall findings nevertheless suggest some level of semantic involvement in explaining the patterns of improvements across a group of five participants and within participants individually. The findings appear to suggest that overall, therapy targeting verb retrieval is more effective in promoting reactivation, reorganisation, or relearning of semantic representations and processing, however, this extends not to other verbs, but to nouns. SFA when targeting noun retrieval was successful here in reactivating, reorganising, or re-teaching semantic representations or processing of those nouns that were treated but the limited diversity that this offered in terms of raising awareness of semantic representations meant that this limited potential to spread activation to other words/concepts regardless of whether they were hypothesised to be related or not.

Chapter 6 Representation and Access to Actions/Verbs in Semantic Memory and Language Processing

6.1. Aims of Chapter

This final chapter attempts to summarise and synthesise the findings presented throughout the previous chapters and consider their implications for the principle research question motivating this thesis:

> To what extent are actions/verbs and objects/nouns represented and accessed from a unitary semantic system according to similar principles?

As the investigations reported throughout this thesis have used a diverse range of methods with both healthy speakers and speakers with language impairments, this chapter begins with a summary of the main themes and findings from each of the individual chapters. This will be followed by discussion of the evidence that this thesis has provided for and against unitary semantic representations and whether objects/nouns and actions/verbs are accessed in a similar way. Key themes and areas for further research that have emerged throughout this thesis will then be considered in more detail, particularly with respect to the issues of polysemy and typicality.

6.2. Summary of Previous Chapters

Chapter one presented an overview of semantic memory and its relationship to single-word level language processing. The majority of research in these areas has been based on observations of the processing of objects and nouns with relatively little attention paid to actions and verbs. Recent evidence has suggested that the patterns of dissociation observed between nouns and verbs in healthy speakers and speakers with language impairments may be attributable to semantic differences rather than purely grammatical class. Considering how actions/verbs are processed at a single-word level and whether their semantic representations and organisation are comparable to object/noun processing was therefore the focus of this chapter.

Chapter two presented an investigation of categorical organisation of actions/verbs, i.e. how actions/verbs cluster within broader categories of action. Theories of categorisation came to prominence in the 1970s where it was suggested that speakers organise objects in the natural world into categories based on the degree of similarity between objects. These perceptual categories are mapped onto conceptual and linguistic representations and these influence organisational principles within semantic memory. Typicality has been suggested to be an organisational principle within categories of knowledge whereby some category members are considered more representative of the category as a whole, as they possess a greater number of attributes that are common to the majority of category members. A category listing task was conducted with healthy adult participants who were asked to list verbs within categories of actions (e.g. *ways of breaking something*, *ways of cooking something*) and nouns within categories of objects (e.g. *types of bird*, *types of vegetable*). Participants listed fewer verbs and had more responses excluded within action categories than nouns within object categories. These differences were likely attributable to there being a smaller repertoire of verbs than nouns in English and also that verbs are more polysemous than nouns. Participants also listed a number of verbs in multiple categories which was not the case for nouns which showed discrete boundaries with few items overlapping into more than a single category. This was likely to be a reflection of the fact that verbs tend to be more polysemous than nouns with greater numbers of associated and related meaning senses. As with object categories, verbs that were listed in action categories by most participants also tended to be listed earlier. Subsequently, results of a typicality rating task showed typicality distributions within action categories were comparable to those within object categories. Typicality also tended to correlate significantly with production frequency measures within category listing and this was also independent of lexical frequency. Therefore, participants appeared to complete both category listing and typicality rating of verbs in action categories in a similar manner to nouns in object categories, implying somewhat similar, although not identical, principles of accessing semantic representations.

Chapter three presented an investigation into semantic similarity between verbs both within and across categories with additional analysis of semantic similarity between levels of verb specificity (i.e. superordinate/general and subordinate/specific). Analysis of a pairwise similarity rating task showed that participants could perceive distinct clusters of verbs which were consistent with data obtained in category listing, e.g. some clusters were discrete within a multidimensional semantic space whereas others blended together (i.e. *breaking* and *cutting*). Participants also perceived high-typicality category members to be closer to the centre of respective category (i.e. the category prototype) than low-typicality category members. An analysis of semantic feature composition of verbs showed a number of characteristics: (1) the majority of features were weakly associated with verbs (i.e. low production frequencies); (2) there was variation between categories of verbs in terms of mean number of features and proportion of feature types; (3) the majority of features were highly distinctive within semantic categories; (4) there was no evidence to suggest a (quantitative) featural distinction between superordinate/general and subordinate/specific verbs; and (5) there was no evidence to suggest a (quantitative) featural distinction between high-typicality and low-typicality category members and between their respective superordinate/general categories. These findings suggest that speakers perceive semantic similarity in a manner that is consistent with performance in category listing and typicality rating reported in chapter two. The reason/s why speakers have these perceptions is currently unclear as this does not appear to be directly attributable to semantic feature composition and overlap within and between categories. While further investigation of semantic feature composition is warranted, it is also plausible that speakers employ additional experiential knowledge of actions in offline tasks, such as when they are being asked to list actions and make rating judgements along particular dimensions related to semantics and meaning. Overall, these findings suggested that semantic similarity and within category organisational principles (e.g. typicality) are perhaps qualitatively different phenomena to those observed in the organisational behaviour of objects/nouns.

Chapter four reported the use of two online psycholinguistic experiments to investigate the influence of semantic representations on lexical retrieval of both verbs and nouns. Within a category verification task, participants verified whether two written verbs or nouns shared a categorical relation (e.g. *frying-cooking*; *apple-fruit*). The time taken by participants to verify that verbs did share a categorical relation was significantly influenced by verbs' typicality within the category, whereby, more typical category members were verified faster. In contrast, the time taken to verify that nouns did share a categorical relation was significantly influenced by the strength of association between the category member and the category, whereby, the stronger the association the faster the verification response. Errors in noun and verb category verification were both significantly influenced by typicality whereby the less typical a category member was, the more likely it was judged (incorrectly) not to be a category member. Within a semantically primed picture naming task, participants named pictures of actions when previously (subconsciously) exposed to a written verb and also named pictures of objects when previously exposed to a written noun. When verb or noun primes were in a coordinate relation to the target picture (e.g. *baking-FRYING*; *apple-BANANA*), participants were 27 msecs slower to name action pictures and 38 msec slower to name object pictures compared to when primes were unrelated to the target.

When verb or noun primes were in a superordinate relation to the picture target (e.g. *cooking-FRYING*; *fruit-BANANA*), participants were 53 msecs faster to name action pictures and 26 msec faster to name object pictures compared to when primes were unrelated to the target. These patterns therefore demonstrated dissociation between prime type (coordinate or superordinate) and the direction of priming effect (i.e. inhibition or facilitation) that was present for both word classes. While the results of the category verification task suggested some differences in organisational principles, or at least access principles, to actions/verbs compared to objects/nouns, the results of the semantically primed picture naming task suggested similar principles of access.

Chapter five presented an intervention study for participants with aphasia which aimed to remediate word retrieval difficulties affecting both verbs and nouns using a semantically-based therapy approach. Different patterns of improvement in verb and noun retrieval were observed in outcome measures of action and object picture naming. As a group, participants improved in their ability to retrieve verbs that were treated but showed no improvement in retrieving verbs that were not treated in therapy. In contrast, participants improved in their ability to retrieve nouns that were treated in therapy and also nouns that were not treated in therapy. This was regardless of whether or not they shared a semantic (i.e. categorical) relation to the nouns that were treated. It was subsequently found, following analysis of order effects of therapy periods, that the widespread generalisation of noun retrieval was more likely to be a consequence of the therapy phase that aimed to improve verb retrieval and was not a direct consequence of the therapy phase that aimed to improve noun retrieval. The lack of within class generalisation suggests similar semantic organisational principles within word classes (i.e. the organisational principles between nouns and between verbs). In addition, there was no evidence to suggest that improvements in word-retrieval were attributable to an internalised strategy that was employed when participants were experiencing word-finding difficulty. Patterns of improvement could be attributed to widespread semantic activation and deeper processing that was possible as a result of using a semantic-based therapy approach when promoting generation of semantic features associated with actions/verbs as opposed to objects/nouns. This was evidenced by the fact that generation of semantic features for actions/verbs proved more challenging than generating features for objects/nouns and it also promoted greater diversity in the features generated. These gains in word-retrieval were however restricted to word-retrieval at single word level with no improvements observed to word-retrieval within sentence level contexts.

This chapter will now continue with discussion of the main themes that emerge when the findings of the individual chapters are integrated within the context of semantic memory and language processing of actions/verbs and objects/nouns.

6.3. Unitary Semantics and Access Principles?

As highlighted in chapter one, recent theoretical accounts of semantics have considered that actions/verbs and objects/nouns may be represented at a featural level within a unitary semantic system (e.g. Vigliocco et al, 2004). Through the investigations presented in the current thesis, potential similarities and differences between actions/verbs and objects/nouns have been identified in terms of their semantic representations and also in accessing these representations. These similarities may be interpreted as support for unitary semantic storage and/or processing whereas differences may be interpreted and being problematic for unitary storage and/or processing.

6.3.1. Similarities between action/verb and object/noun processing

Within category listing and typicality rating (i.e. chapter two), those verbs that were listed within categories most frequently were also rated as more typical category members. This effect was also independent of lexical frequency. These results parallel the findings within noun categories both in the current thesis and in previous research (e.g. Hampton & Gardiner, 1983; Mervis et al, 1976). This therefore suggests that when participants are asked to list actions within categories, their response behaviour is influenced by similar organisational principles as when listing objects in categories. The nature of the correlation does not reveal causal explanation of response behaviour, i.e. whether typicality influences response frequency or vice versa. However, within object categories, family resemblance theories of categorisation (e.g. Rosch & Mervis, 1975) suggest that response frequency is guided by ease of access. Category members that are more central to the category prototype (i.e. high-typicality category members) are more readily accessible and hence more likely to be listed first and also more likely to be consistently listed between participants. In comparison, low-typicality category members that are stored at a greater distance from the category prototype would be listed later with more between-participant variation (i.e. less likely to be listed by all or most participants).

The results of the semantically primed picture naming task suggest that semantic relatedness of verbs to other verbs and also nouns to other nouns shows parallels and this may influence naming behaviour in different directions (i.e. facilitation or inhibition) depending on the nature of the semantic relation. For both verbs and nouns, where primes were in a superordinate semantic relation to the target (e.g. *cooking-FRYING*; *fruit-BANANA*), naming was facilitated. This therefore suggests that where primes are semantically related and are congruous in terms of semantic feature overlap (i.e. all features in the prime are assumed to be possessed by the target), there may be residual activation within the semantic network which leads to faster activation of related targets. However, where primes were coordinates (i.e. members of the same category at the same level of categorical abstraction and specificity), naming was inhibited for both verbs and nouns. This therefore suggests that where semantic relatedness presents incongruous information (i.e. the prime possesses features that are not possessed by the target and may even be contrary to the target), the residual activation in the semantic network leads to slower naming as a semantic competitor has previously been processed (e.g. Hantsch, Jescheniak & Schriefers, 2005; Levelt, Roelofs & Meyer, 1999; Vitkovitch & Tyrrell, 1999).

Within the intervention study for participants with aphasia, naming of verbs and nouns showed similar patterns of improvement and generalisation <u>within</u> word classes (i.e. noun-to-noun; verb-to-verb). While therapy was generally effective at improving naming performance of treated verbs within verb-SFA, there was no concomitant improvement in naming untreated verbs and, similarly, noun-SFA led to improved naming of treated nouns but not untreated nouns. While the lack of within word class generalisation for verbs in consistent with previous intervention studies targeting verb retrieval (e.g. Faroqi-Shah & Graham, 2011; Raymer & Ellsworth, 2002; Wambaugh & Ferguson, 2007), the lack of within word class generalisation for nouns contrasts with previous reports (e.g. Boyle & Coelho, 1995; Kiran & Thompson, 2003; Lowell et al, 1995) that suggest generalisation can be achieved both within and across semantic categories. The current findings suggest comparable organisational principles within the semantic system such that activation of semantic representations via lexical processing is not sufficient to cause co-activation of related word forms, at least not to an extent where their representations are strengthened sufficiently to reactivate impaired representations. This therefore presents an intriguing adjunct to the findings of the semantically primed picture naming task; while prior exposure and processing of a verb (or noun) facilitated or inhibited immediate subsequent processing of a related verb (or noun), presumably through co-activation within an unimpaired semantic system (i.e. in healthy adult speakers), this co-activation was not strong enough to lead to lasting change within an impaired semantic network (i.e. speakers with language impairment).

One reason why therapy did not lead to lasting change in the extended semantic network of target stimuli may be related to the intensity and duration of the therapy. The timescale of the intervention study may not have provided the required 'critical mass' of therapy to facilitate generalisation effects (e.g. see Nadeau & Kendall, 2006, for discussion). Each therapy phase constituted approximately 10 hours of therapy over five weeks with each target item being exposed five times. This is generally a shorter duration and less intense than previous reports where generalisation has been reported as a result of noun-SFA (e.g. Boyle, 2004: 12 sessions over four weeks; Stanczak, Waters & Caplan, 2005: generalisation observed after 16 sessions in one participant) although there are also reports which have seen generalisation following fewer sessions (e.g. Boyle & Coelho, 1995: observed after nine sessions). Many of these previous studies have however employed success criterion where therapy is terminated when participants achieve a pre-specified level of correct naming in treated items. This was not employed here as the priority was to ensure that all participants received an equivalent amount of therapy in both noun-SFA and verb-SFA. Therefore, this design allowed comparison of the two therapies (i.e. noun-SFA and verb-SFA) in their ability to lead to improvement over a definitive time period but does not necessarily give an answer as to their absolute potential to facilitate generalised improvement (i.e. whether, and at what point, improvement and generalisation may be observed). In general, the more intense and more prolonged an intervention period is, then the greater the positive benefit (see Basso, 2005, for a review), although there is also evidence that the same amount of therapy delivered over a longer period (i.e. less intensive) can lead to greater improvement in maintenance phases (i.e. after periods of no therapy) compared to the therapy over a shorter period (i.e. more intensive; e.g. Sage, Snell & Lambon Ralph, 2010). Therefore, this issue is one that may only be addressed through further investigation.

6.3.2. Differences between action/verb and object/noun processing

Across the investigations reported in the current thesis, there were indications that participants found tasks more difficult when the target stimuli were verbs as opposed to nouns. This was evidenced by: (1) greater numbers of errors and data exclusions in category listing, category verification and semantically primed picture naming; (2) slower response times to verbs in category verification and semantically primed picture naming; and (3) greater difficulty and more variation in semantic feature generation by both healthy speakers and within the intervention study for speakers with language impairment. Also, with speakers with language impairment, actions appeared more difficult to name from pictures compared to objects, as indicated by generally lower percentages correct. However, with most participants, this noun advantage disappeared, and in some cases even reversed, when naming performance on a subset of actions and objects that were matched for various psycholinguistic properties was compared. These findings therefore reinforce the need to understand why verbs may appear to be more difficult to process than nouns both in general (i.e. across speakers) and in specific cases (i.e. within speakers), especially in the case of speakers with language impairments. For example, it has previously been demonstrated that so-called noun-verb dissociations may be eliminated when semantic factors (i.e. imageability) are controlled for (e.g. Bird, Howard & Franklin, 2003), although grammatical class dissociations have also been observed to persist in some participants even when such factors are controlled (e.g. Berndt, Haendiges, Burton & Mitchum, 2002). Other explanations of greater difficulty with verbs, particularly in contexts where picture stimuli are eliciting verbs, concern the greater visual and inferential complexity of depictions of actions compared to depictions of objects (e.g. see Berndt, Mitchum et al, 1997, for discussion). Naming pictures of objects is perhaps a far more clear-cut task as there is one clear correct static target concept (although this may have variable associated nouns depending on individual, social, cultural variation). Naming pictures of actions involves interpretation of more complex and dynamic relations between different objects and participants. Therefore, greater numbers of errors in action picture naming tasks may be a reflection of the options that are available rather than an inability to retrieve the target verb per se; there is simply more to say about a picture of an action than there is about a picture of an object. This may be reflected in the current intervention study where participants made a greater number of 'misinterpretation' errors within the OANB subtest of action naming compared to object naming where these mostly consisted of participants naming an object shown in the picture. Some research has attempted to overcome this difficulty by eliciting verb retrieval through the use of videotaped actions where the action is seen from beginning to end rather than from one particular time point (e.g. Berndt, Mitchum et al, 1997). Findings from such studies are variable most likely as a result of individual differences in the participants under investigation. Some have found that little difference between video and picture stimuli in terms of their effectiveness in eliciting target verbs (i.e. Berndt, Mitchum et al, 1997) while other have shown that dynamic depictions leads to improved verb production compared to static depictions (in the same participants; e.g. Pashek & Tompkins, 2002). Also, verb production elicited through video stimuli has been shown to recruit greater and more widespread neural activation including areas associated with manipulation of objects, compared to elicitation through picture stimuli (e.g. den Ouden, Fix, Parrish & Thompson, 2009). Such findings therefore raise further questions as to the ecological validity of eliciting verb production in single-word contexts, not least making clinical decisions regarding diagnosis and interventions based on such observations.

Within category verification, responses to verbs and nouns were influenced by different variables. For verbs, response time was influenced by typicality within the respective category whereas for nouns, response time was influenced by association strength. With both verbs and nouns however, the chance of making an error was influenced by typicality alone. This is perhaps one of the more difficult findings to reconcile as previous research has found that category verification with nouns is also predicted, at least partially, by typicality (e.g. Casey, 1992; Hampton, 1997; Larochelle & Pineau, 1994). Therefore it was unexpected that this did not feature as an influential predictor variable of response time within group or individual analyses.

While the results of the intervention study with participants with aphasia showed similarities in patterns of naming improvement <u>within</u> word classes, they also showed different patterns of improvement <u>between</u> word classes. There was evidence to suggest that verb-SFA facilitated improved naming of treated verbs and untreated nouns whereas noun-SFA only facilitated naming of treated nouns. There was no evidence that these results could be accounted for in terms of overlap between semantic features elicited during the therapy task or in terms of strategic use of SFA when faced with pictures that could not be named spontaneously during assessment phases. Precisely why these patterns were observed is difficult to explain but they may give insight into the types of information that become activated within the semantic network of stimuli when they are targeted (i.e. treated) using SFA-type therapy tasks. Across participants and in both therapy approaches used here (i.e. verb-SFA and noun-SFA), the majority of semantic features produced were lexically realised as nouns (54-64%) but there may be qualitative differences in terms of the types of information that these represent and how this information maps onto the semantic representations of the verbs and nouns being targeted. It could be hypothesised that the types of features elicited in verb-SFA overlapped with objects that may be consistent with thematic role information about the target verb (i.e. Related object – THEME/PATIENT; Tool – INSTRUMENT) which may help to reinforce semantic reactivation. In comparison, the balance of features elicited by noun-SFA were perhaps more likely to be semantic features which were 'parts' as opposed to 'wholes' (e.g. see Tversky & Hemenway, 1984, for discussion of objects, parts, and categories) and which may therefore be concepts that were not likely to be tapped within an object (i.e. whole object) naming task.

Support for the suggestion that the semantic networks of verbs contain thematic role information, in conjunction with (or in preference to) more perceptual-type features comes from priming studies such as Ferretti, McRae & Hatherell's (2001). They found that prior exposure to verbs facilitates lexical decision response times to typical agents, patients and instruments compared to unrelated thematic participants (e.g. *scrubbing-JANITOR*, *arresting-CROOK*, *stirred-SPOON*;) although there was no such effect of facilitation for typical locations (e.g. *swam-OCEAN*). In order to investigate this hypothesis, a further outcome measure could be incorporated into future intervention studies where feature generation could be elicited under conditions of no feedback from the clinician/researcher (e.g. *give a description of a tiger/apple/chair;* where richness of descriptions could be compared before and after therapy in terms of number and appropriateness of features produced). This may also be explored in greater depth in healthy speakers. One possibility may be to conduct a detailed qualitative analysis of feature types from features given within feature listing tasks, and in the case of verbs to compare elicited features with frequency of occurrence of the features as syntactic arguments and thematic roles in sentence contexts (e.g. obtained through analysis of corpus data).

While verbs are generally accepted to be conceptually and linguistically more complex than nouns (e.g. Black & Chiat, 2003; Druks, 2002; Marshall, 2003), the precise reason for why verbs are more difficult to process within specific tasks is often overlooked in favour of a general explanation of overall complexity. This is especially important to consider in light of the conclusions reached in chapter five where the increased difficulty associated with carrying out the semantic feature analysis therapy with verbs actually led to more widespread improvement in word retrieval. The findings from the current investigations suggest that semantic complexity, and more specifically the role of polysemy, may be a potentially important factor to consider when investigating the processing of verbs and storage and access to semantic representations of actions.

6.3.3. Conclusions on unitary semantics and access

Although it was previously stated that any differences observed between verbs and nouns throughout the current thesis could be interpreted as potentially being problematic for a unitary view of semantics (see section 6.3.1), the specific differences observed in the current investigations are not incompatible with a unitary view. The greater number of errors within verb tasks which has been interpreted as a possible indication of the difficulty of eliciting verbs and the fact that verbs tend to be more polysemous than nouns, may indicate differential connections between verbs within a semantic system but this may still be based on the same fundamental representational principles of conceptual featural representation (i.e. within the FUSS model of Vigliocco et al, 2004). The current investigations have not however been able to elucidate the significance of the featural representation of verbs in governing some of the dimensions on which verbs have been shown to vary in the current thesis (i.e. typicality) and from previous research (i.e. the distinction between general and specific verbs). Slower responses to verb stimuli may again reflect the difficulty with reliably eliciting verbs from experimental stimuli and/or a difference in the speed of accessing and retrieving appropriate semantic representations from picture stimuli of actions. Again, slower responses cannot directly be attributed to a fundamental difference in featural representation within a unitary semantic system.

The greater variation in feature generation in both healthy speakers and speakers with language impairment is perhaps the most insightful observed difference with regard to drawing conclusions about a unitary semantic system that is responsible for verb and noun processing. Again, this doesn't provide evidence against conceptual featural representation within a unitary semantic system although it does suggest some qualitative differences in the nature of this featural representation, i.e. the type of features that are relevant to the conceptual representation of objects and actions are different. This may in itself be a contributory factor for the other observed differences as featural representation is assumed to be significant in dictating typicality (e.g. Rosch & Mervis, 1975) and semantic similarity in general (e.g. Maki et al, 2006; Mirman & Magnuson, 2009).

Throughout the experiments reported in the current thesis, where verb and noun production has been directly elicited on a one-off basis (i.e. category listing, semantically primed picture naming), participants have shown comparable response behaviour (e.g. the same direction of priming effects). Therefore, these types of task, provide the strongest evidence in the current investigations for unitary semantics. These tasks suggest that when participants are able to access semantic representations via lexical representations, then their subsequent behaviour (i.e. generation and retrieval of related and unrelated lexical items) will be influenced according to similar access principles which are presumably influenced by prior activation within the semantic system. Such behaviour is presumably insensitive to whether the featural representation of verbs and nouns is qualitatively different as the featural representations appropriate for each word class are already activated by prior exposure to related items in the same word class (e.g. through the category verb in category listing, or via the prime stimuli in semantically primed picture naming). It is perhaps only when the tasks necessitate conscious retrieval of featural information that participants demonstrate differences between word classes (e.g. semantic feature listing and the intervention study) as these do not explicitly aim to exploit the activation between different lexical items via their semantic feature representations. For example, although the intervention study aimed to do this implicitly through observation of generalisation effects, this was not done explicitly within the therapy task itself. A possible adaptation of the intervention study which may aim to exploit activation between lexical items may have been to ask participants to think of 'related actions' thereby reinforcing semantic activation between verbs via their relevant featural representations.

More thorough investigation of priming effects between word classes may help to elucidate the strength of association between word classes. For example, if verbs' semantic features are more consistently representing arguments (i.e. nouns) while nouns' semantic features more consistently represent attribute information (i.e. other nouns), then there may be dissociations (or at least differential magnitudes of priming effects) that can be observed when conducting priming tasks where nouns and verbs are systematically presented as either primes or target stimuli (e.g. *[noun]-[NOUN]*; *[noun]-[VERB]*; *[verb]-[VERB]*; *[verb]:[NOUN]*) and when the relationship between prime and target is systematically varied in terms of whether they are in a featural or other relationship or not.. There has already been some research which has highlighted some investigations in these areas. Mahon et al (2007) used a picture-word interference task to investigate the effect related verbs on the naming of object pictures. Vigliocco et al (2005) similarly used a picture-word interference task and investigated the effect of related and unrelated nouns and verbs on the naming of object and action pictures both within word class and across word class conditions. Tyler & Moss (1997) have investigated the effects of object/noun primes on the speed at which different types of semantic features were recognised within a lexical decision task. However, such studies naturally report the use of different tasks, different stimuli, and different participants so it is currently difficult if not impossible to draw conclusions as to how this supports a view of unitary semantics.

6.4. Further Research

In addition to the suggestions for further research that have been raised throughout individual chapters, there are two broad themes that have arisen as potentially important, yet under-researched, aspects of semantic organisation between actions/verbs: (1) polysemy, i.e. overlapping meaning representations; and (2) typicality, i.e. representativeness of a particular category, or class of actions.

6.4.1. Polysemy as a psycholinguistic variable

The fact that verbs are more polysemous than nouns has previously been highlighted as one of the potential reasons why verbs are more complex than nouns and are consequently found to be more difficult in a variety of contexts from language acquisition, performance of healthy speakers, and performance of speakers with language impairment (e.g. Black & Chiat, 2003; Druks, 2002; Marshall, 2003). This thesis highlights that this has potentially been a variable that has received little attention but which could be very influential on speakers' performance in a number of tasks. Studies occasionally take account for when verbs have a homophonous noun form but few consider the number of different meaning senses that particular verbs possess.

The polysemous nature of verbs was demonstrated in category listing with some categories of verbs showing overlapping category members (e.g. *breaking* and *cutting* categories). The category listing task and the subsequent typicality rating task also demonstrated that where category members overlap they can hold differential association strengths and differential representativeness (i.e. typicality) within the different categories they are associated with, For example, *ripping* and *tearing* received production frequencies of 15 and 14 respectively (from a total of 35 participants) and mean typicalities of 2.24 and 2.51 as ways of *breaking* something. In comparison, *ripping* and *tearing* received production frequencies of 9 and 7 and mean typicalities of 3.09 (for both) respectively, as ways of *cutting* something. While verbs are generally more polysemous than nouns, some verbs are also more polysemous than other verbs and have preferential, or perhaps default, meaning senses. It may be plausible that such variables may influence performance in psycholinguistic tasks such as category verification, so that the greater number of meaning senses a verb has, the longer it takes to verify one particular meaning (i.e. to associate it with one particular category), especially if this is not the preferred meaning. Number of meaning senses may also impact on the assessment and intervention of language impairments. For example, stimuli may need to be carefully selected so that intervention tasks (e.g. semantic feature generation) is concordant with preferred meaning senses of the lexical stimuli and does not cause conflict (e.g. in some situations it may be plausible that *tearing* is a way of *breaking* something and therefore *breaking* would be an acceptable semantic feature for generation; in other contexts *tearing* may also be a plausible means of *making* something such as in papier mache where paper is torn into strips).

The greater polysemy of verbs may potentially restrict the validity of methods such as semantic feature listing. When collecting semantic features for objects/nouns, the experimenter can be fairly confident that there exists between-participant agreement on the concept that the lexical item denotes. However, given that verbs can represent subtle shades of meaning and can be used even when discussing a varied range of situations (e.g. *washing* may infer different actions, processes, instruments, and so on, in the context of *washing your face*, *washing the car*, *washing the dishes*, and so on), the same confidence may not be guaranteed when listing features for actions/verbs. This may be especially problematic when features are elicited by a single word (i.e. the verb in question) which is devoid of context. In such cases, participants may generate semantic features relevant to a particular context that they imagine and associate with the verb but this may be different to a context that another participant generates. This may be one reason why there is an apparent lack of consistency between participants in feature listing (i.e. why the majority of features have low production frequencies and are highly distinctive; see chapter three).

In considering further how polysemy of verbs may be investigated, especially with respect to categorisation and overlap between categories, there are a number of potential opportunities. Further category listing experiments may compare speakers' performance in listing actions at varying levels of specificity (i.e. levels of categorisation) to see if, and to what extent, overlapping category membership is prevalent at differing levels. For example, speakers may be asked to list *ways of moving* in addition to *ways of running/walking/jumping* and so on. This may be problematic in that it may not make sense to ask speakers to list *ways of <u>doing</u> something* and compare this to more specific actions (e.g. *ways of breaking/cleaning/making something*) but this may be overcome with careful experimental design, such as with specific task instructions and examples to ensure participants know what is required of them. Such data may then be informative when interpreting the results of parallel investigations, such as semantic feature listing, where it may be clearer to identify (or refute) whether there is a featural basis for explaining speakers' perceptions of how verbs may form categories, or clusters, and also in terms of how verbs are rated for typicality and semantic similarity.

The issue of semantic complexity as a result of polysemy may also be demonstrated in studies that have investigated the effects of verbs' argument structure properties. In a similar spirit to classification systems such as Levin's (1993) system whereby it is assumed that verbs' underlying semantic representations are reflected in their distributions in sentence contexts, Shapiro and colleagues (e.g. Shapiro & Levine, 1990; Shapiro, Zurif & Grimshaw, 1987, 1989; Thompson, Lange, Schneider & Shapiro, 1997) have demonstrated that verbs are more difficult to retrieve as argument structure properties become more complex. This complexity is reflected in the canonical number of arguments that a verb is associated with (e.g. one, two, or three arguments) and also as the number of different possible argument arrangements increases. For example, transitive verbs, such as *solved* which are expressed with two arguments, as in *the teacher solved the equation*, are easier to retrieve than dative verbs such as *donated*, which may be expressed with two arguments as in *Renoir donated the painting*, or with three arguments as in *Renoir donated the painting to the government*. This added difficulty has been demonstrated in both healthy speakers and speakers described as having Broca's type aphasia. In Shapiro and colleagues' investigations, this was evidenced by slower response times and greater errors within a series of lexical decision tasks. Participants listened to sentences and were simultaneously asked to make lexical decisions to visually presented words/nonwords. Effects on response time and errors tended to be largest when lexical decision was presented immediately following the auditory presentation of the main verb in the sentences. This was interpreted as showing that on immediate processing of a verb, all possible argument arrangements become activated based on the semantic representations of the verb. Those verbs with greater argument possibilities then slowed performance in lexical decision tasks to a greater extent due to the reduced processing capacity available. In addition, no such effects in interference were observed in speakers described as having fluent-type aphasia. These participants' semantic impairments were assumed to render them insensitive to widespread semantic activation and consequently the argument complexity of the verb. Hence, while this thesis has earlier claimed (see chapter two; section 2.2.4.) that resources such as Levin's (1993) classification system may not be adequate for some psycholinguistic investigations due to them being developed in a post hoc manner, they may nevertheless be an important adjunct and consideration alongside other indices of semantic complexity.

6.4.2. Typicality of actions/verbs

Even though actions/verbs may not be organised into discrete semantic categories in the same manner as objects/nouns, it does appear as though speakers can perceive that some actions are more or less typical, or representative, than others, when considered in relation to general categories of actions. Furthermore, there is evidence that these perceptions, obtained via typicality rating, are a reflection of a psycholinguistic variable that affects performance in online tasks (i.e. category verification). Within object categories, typicality is assumed to reflect the degree to which a concept overlaps in terms of semantic features with other objects within the same category (e.g. Rosch & Mervis, 1975). Therefore, highly typical category members share a greater number of features with other high-typicality category members and low-typicality category members share fewer features with fellow category members and share more features with objects in other categories than high-typicality category members do.

When speakers rate typicality of objects within categories, they do this without being influenced by the frequency with which the lexical items appears in the language (e.g. Mervis, Catlin & Rosch, 1976) and typicality has also been argued to be independent of other variables such as familiarity (e.g. Boster, 1988; Hampton & Gardiner, 1983). Typicality has also been shown to influence performance in various psycholinguistic tasks with healthy speakers (e.g. Casey, 1992; Larochelle & Pineau, 1994) and also speakers with language impairments (e.g. Kiran & Thompson, 2003). Within the current thesis it has been found, in relation to actions/verbs, that: (1) speakers' ratings of typicality are correlated with association strength but independent of lexical frequency; (2) speakers may perceive high-typicality actions to be closer to the centre of the hypothesised category, or cluster of actions centred around more general actions, than low-typicality actions; (3) there is little quantitative difference in the featural composition of high-typicality actions compared to low-typicality actions; and (4) typicality predicts healthy speakers' performance in category verification both in terms of response time and also the likelihood of making an error. Therefore, the current thesis suggests that the notion of typicality is relevant for semantic processing of actions but as yet does not provide an answer to what determines typicality of actions.

Categories of actions may be considered more comparable to so-called ad hoc categories than categories of naturally occurring objects. Where featural overlap determines typicality in object categories (e.g. Rosch & Mervis, 1975), Barsalou (1983) has argued that ad hoc categories present "fundamentally different forms of graded structure" (p225) where only properties relevant to the goal which dictates the category members (e.g. *things not to eat on a diet*) are considered by speakers when they come to rate typicality (e.g. *<edible>*, *<high in calories>*). Therefore, further research would be insightful in investigating further the effects that typicality has on performance in tasks with actions/verbs, in order to validate the notion of typicality as a psycholinguistic variable that is important to consider when working with actions/verbs. In accordance with this it will be important to differentiate any effects that typicality may have from other variables which have not been investigated in depth in the current thesis (e.g. familiarity, number of senses, concreteness, imageability, and so on).

6.5. Concluding Remarks

This thesis has attempted to fill a need in areas relevant to linguistics, psychology and aphasiology. A frequent complaint in such areas, specifically when conducting work with verbs, is that far too little is known about the semantic representations of verbs. Such comments have often come when attempting to explain 'null' results, or when participants' performance differs with verbs compared to their performance with nouns:

> A theory of the mental organisation of verb meanings that is well developed and accepted by most researchers is still unavailable. This state of affairs has consequences for the empirical work conducted on verb production. The selection of the experimental materials is necessarily based on intuition, which ... is often unclear. Thus, on the one hand, materials are typically less controlled in verb than in noun research. On the other hand, it is very difficult to establish their adequacy across studies. (Collina & Tabossi, 2007:75)

Verbs are undoubtedly more complex than nouns across a number of different linguistic and psycholinguistic dimensions and this additional complexity does have an impact of aspects of language behaviour (see Black & Chiat, 2003; Druks, 2002; Marshall, 2003; Mätzig et al, 2009, for reviews). However, a non-specific or all-encompassing description of verbs of being more complex than nouns is not useful to address the concerns of a researcher who holds views similar to Collina & Tabossi (2007). The investigations reported in the current thesis have tended to focus on the semantic properties of verbs at a single-word level in a number of tasks. This has been informative in revealing where and how the representation and processing of actions/verbs and objects/nouns show similarities and where they show differences. Therefore, while investigating verb processing at the level of single words may exclude the more natural sentential context, as verbs naturally denote dynamic actions and relations between different entities which are more readily expressed within sentence contexts, it does nevertheless appear to be an informative place to (re)start the investigation of the semantic representations of verbs.

Appendix A Category Listing - Quantitative Summary of Excluded Responses

	PP		Adv/Adj		Noun		Repetition		Other		Total	
											N	%
Breaking	4		6		6		1		6		23	3.2
Cleaning	9		2		9		1		0		21	2.9
Cooking	3		7		12		0		0		22	3.1
Cutting	17		48		40		0		0		105	14.7
Hitting	16		27		13		2		0		58	8.1
Jumping	38		41		10		3		1		93	13.0
Making	4		1		9		0		9		23	3.2
Running	32		69		30		1		12		144	10.2
Talking	9		93		31		0		0		133	18.6
Walking	20		87		7		1		1		116	16.2
Total	N	%	N	%	N	%	N	%	N	%	N	
	152	21.3	381	53.3	167	23.4	9	1.3	29	4.1	738	

Note.: PP – Preposition phrase; Adv/Adj – Adverbs or adjective

Table A1. Exclusion data from verb category listing

	PP		Adv/Adj		Verb		Repetition		Other		Total	
											N	%
Animal	0		0		0		1		0		1	2.4
Bird	0		0		0		0		0		0	0.0
Clothes	0		0		0		1		0		1	2.4
Fruit	0		0		0		1		0		1	2.4
Furniture	0		0		0		3		0		3	7.1
Music	0		0		0		0		0		0	0.0
Sport	0		0		0		3		0		3	7.1
Tool	0		0		0		2		0		2	4.8
Transport	0		2		29		0		0		31	73.8
Vegetable	0		0		0		0		0		0	0
Total	N	%	N	%	N	%	N	%	N	%	N	
	0	0.0	2	4.8	29	69.0	11	26.2	0	0.0	42	

Note.: PP – Preposition phrase; Adv/Adj – Adverbs or adjective

Table A2. Exclusion data from noun category listing

Appendix B Category Listing - Gender Quantitative Comparisons

Category	Male		Female		t-test		
	M	SD	M	SD	t	df	p
Breaking	7.23	3.4	10.46	3.3	2.467	24	0.021*
Cleaning	10.00	2.2	10.15	3.1	0.147	24	0.885
Cooking	10.62	3.8	10.31	3.8	0.209	24	0.836
Cutting	6.00	3.1	5.00	2.9	0.850	24	0.404
Hitting	7.31	2.6	7.69	4.3	0.278	19.85	0.784
Jumping	5.23	2.1	6.62	3.0	1.361	24	0.186
Making	8.23	2.8	9.00	3.4	0.622	24	0.540
Running	3.46	1.8	3.15	1.6	0.456	24	0.652
Talking	6.08	3.3	8.77	4.6	1.722	24	0.098
Walking	4.69	3.3	7.62	3.7	2.137	24	0.043*

Table B1. Within verb category gender quantitative comparisons

Category	Male		Female		t-test		
	M	SD	M	SD	t	df	p
Animals	22.85	6.7	22.85	3.9	0.000	24	1
Birds	17.92	6.5	19.77	4.6	0.839	24	0.410
Clothes	21.77	5.0	20.23	2.4	0.995	17.13	0.333
Fruit	17.92	4.2	20.69	3.9	1.734	24	0.096
Furniture	13.23	3.0	15.85	3.5	2.042	24	0.052
Musical instruments	18.54	5.0	20.08	3.3	0.920	24	0.367
Sport	20.62	4.9	19.69	1.9	0.637	15.52	0.923
Tools	13.85	3.7	13.08	2.8	0.595	22.18	0.558
Transport	15.77	4.7	16.15	3.7	0.232	24	0.818
Vegetables	14.69	4.1	17.54	3.4	1.917	24	0.067

Table B1. Within noun category gender quantitative comparisons

Appendix C Category Listing – Presentation List Quantitative Comparisons

Category	List A		List B		t-test		
	M	SD	M	SD	t	df	p
Breaking	8.76	2.7	9.65	4.3	0.718	32	0.478
Cleaning	9.94	3.0	10.76	2.6	0.856	32	0.399
Cooking	9.76	2.1	11.24	4.2	1.296	23.89	0.207
Cutting	5.29	2.9	6.06	3.3	0.718	32	0.478
Hitting	7.00	3.5	7.88	4.0	0.682	32	0.500
Jumping	6.35	2.8	5.82	2.6	0.573	32	0.571
Making	8.53	3.3	9.65	3.2	1.005	32	0.322
Running	3.24	1.9	3.06	1.5	0.303	32	0.764
Talking	7.47	3.7	6.76	4.9	0.469	32	0.643
Walking	7.82	3.3	3.88	2.9	3.644	32	0.001*

Note.: * $p < .05$

Table C1. Within verb category presentation list quantitative comparisons

Category	List A		List B		t-test		
	M	SD	M	SD	t	df	p
Animals	22.71	5.9	23.18	4.8	0.255	32	0.800
Birds	19.82	5.2	18.12	4.7	1.010	32	0.320
Clothes	21.47	3.9	21.35	3.6	0.092	32	0.928
Fruit	20.29	4.2	18.88	3.4	1.079	32	0.288
Furniture	15.65	2.8	13.65	3.6	1.816	32	0.079
Musical instruments	19.53	3.8	18.76	4.4	0.538	32	0.594
Sport	20.24	3.5	20.06	3.5	0.148	32	0.883
Tools	13.82	3.4	12.59	3.3	1.079	32	0.289
Transport	16.88	4.3	15.06	4.6	1.194	32	0.241
Vegetables	17.41	3.6	16.41	3.5	0.822	32	0.417

Table C2. Within verb category presentation list quantitative comparisons

Appendix D Category Listing – Verb Responses and Quantitative Data

Response	Production freq’	Mean rank	1st ranked
Dropping	28	3.57	3
Smashing	28	1.96	22
Snapping	20	4.00	3
Crushing	15	4.67	
Hitting	15	5.87	2
Ripping	15	6.14	1
Tearing	14	6.00	
Throwing	14	5.29	1
Bending	13	6.38	
Cracking	10	3.40	
Cutting	10	6.60	
Destroying	7	6.29	
Shattering	6	4.83	
Slicing	6	8.33	
Kicking	5	5.40	
Squashing	5	9.60	
Stamping on	5	8.60	
Chipping	4	4.75	
Bashing	3	7.67	
Burning	3	5.33	
Crashing	3	3.00	
Denting	3	10.67	
Falling	3	8.00	
Forcing	3	8.67	
Jumping on	3	4.33	1
Pulling apart	3	6.00	
Pulling	3	8.00	
Shredding	3	9.33	
Splintering	3	7.33	
Splitting	3	7.67	

Table D1. Responses in category listing task – Breaking

Response	Production freq’	Mean rank	1st ranked
Washing	30	3.97	4
Scrubbing	30	2.53	15
Wiping	22	4.00	4
Brushing	21	5.76	
Polishing	21	4.00	5
Hoovering	20	6.00	
Mopping	19	7.16	
Dusting	17	5.47	1
Rubbing	14	6.29	2
Rinsing	11	5.55	
Soaking	10	6.30	
Sweeping	10	7.90	
Spraying	9	7.33	1
Bleaching	7	5.86	
Cleansing	6	6.17	
Scouring	6	6.17	
Shining	6	6.00	2
Bathing	5	8.60	
Hosing	5	8.60	
Shampoo	5	7.40	
Steaming	5	7.40	
Drying	4	7.25	
Scraping	4	10.25	
Showering	4	9.75	
Sterilising	4	8.25	
Vacuuming	4	9.50	
Dabbing	3	12.00	
Disinfecting	3	5.00	
Dry cleaning	3	6.00	
Lathering	3	5.33	

Table D2. Responses in category listing task – Cleaning

Response	Production freq’	Mean rank	1st ranked
Frying	32	2.63	11
Boiling	31	3.58	7
Grilling	31	5.42	1
Baking	30	3.23	11
Roasting	23	5.57	
Steaming	21	6.10	
Microwaving	20	7.55	
Barbequing	15	7.67	
Simmering	14	6.00	
Poaching	12	6.58	
Toasting	12	8.50	
Heating	9	7.11	
Sautéing	8	6.63	
Stir frying	8	5.50	1
Burning	7	8.71	
Blanching	5	6.40	1
Melting	5	8.60	1
Mixing	5	11.40	
Stewing	5	7.00	
Griddling	4	6.75	
Searing	4	7.75	
Slow cooking	4	7.00	1
Braising	3	8.67	
Char-grilling	3	7.33	
Chopping	3	12.67	
Deep frying	3	5.00	
Freezing	3	10.00	
Scrambling	3	14.67	

Table D3. Responses in category listing task – Cooking

Response	Production freq’	Mean rank	1st ranked
Slicing	22	1.91	11
Chopping	21	2.29	7
Sawing	20	4.10	3
Dicing	13	4.23	
Stabbing	10	4.00	
Tearing	9	3.33	1
Hacking	8	3.38	1
Ripping	7	3.57	2
Snipping	5	4.00	
Scissoring	4	6.00	1
Slashing	4	2.50	2
Snapping	4	5.75	1
Splitting	4	6.00	
Trimming	4	7.00	
Carving	3	3.33	1
Grating	3	5.00	
Piercing	3	5.33	
Severing	3	4.33	

Table D4. Responses in category listing task – Cutting

Response	Production freq'	Mean rank	1st ranked
Punching	31	2.00	16
Slapping	24	3.67	4
Kicking	16	3.81	
Smacking	12	3.50	3
Bashing	8	4.50	2
Crashing	8	6.25	
Hammering	8	3.88	3
Smashing	8	5.88	
Banging	7	4.29	1
Thumping	7	5.00	
Whacking	7	6.71	1
Elbowing	6	7.17	
Head butting	6	5.33	
Walloping	6	6.00	
Colliding	5	7.20	
Jabbing	5	4.00	
Beating	4	6.50	1
Tapping	4	4.25	1
Batting	3	1.33	2
Flicking	3	6.33	
Knocking	3	7.33	
Nudging	3	6.33	
Poking	3	6.00	
Pushing	3	8.00	
Striking	3	3.00	
Throwing	3	8.33	

Table D5. Responses in category listing task – Hitting

Response	Production freq’	Mean rank	1st ranked
Hopping	22	2.64	7
Leaping	17	3.18	8
Long jump	16	3.31	1
Skipping	16	3.50	
High jumping	15	3.33	6
Bouncing	14	4.00	2
Star jumping	11	3.45	2
Bounding	10	4.70	1
Diving	9	4.33	1
Springing	9	4.33	1
Triple jump	7	4.57	
Bungee jump	5	2.40	2
Trampolining	5	5.20	
Vaulting	5	4.40	2
Leapfrog	3	6.67	
Pike jump	3	6.33	
Show jumping	3	6.67	
Skydiving	3	5.00	

Table D6. Responses in category listing task – Jumping

Response	Production freq’	Mean rank	1st ranked
Baking	23	3.30	6
Building	21	4.00	3
Cooking	19	4.42	6
Drawing	17	6.35	
Gluing	17	3.47	6
Painting	16	5.94	
Creating	12	5.17	1
Sewing	9	3.78	1
Sticking	9	4.00	2
Moulding	8	5.00	1
Constructing	8	4.88	1
Carving	6	6.17	
Designing	6	5.83	
Mixing	6	7.50	1
Sculpting	6	6.33	1
Assembling	5	3.80	
Hammering	5	8.80	
Knitting	5	3.80	1
Writing	5	11.00	
Cutting	4	4.75	1
Printing	4	4.25	
Sawing	4	7.75	
Crafting	3	5.00	1
Fixing	3	6.00	
Folding	3	7.00	
Inventing	3	7.67	
Joining	3	2.67	1
Manufacturing	3	8.67	
Recording	3	5.00	
Typing	3	9.00	
Welding	3	7.33	1

Table D7. Responses in category listing task – Making

Response	Production freq’	Mean rank	1st ranked
Jogging	31	1.58	18
Sprinting	31	1.71	13
Racing	8	3.38	
Marathoning	6	3.00	
Dashing	4	2.75	
Striding	3	3.67	

Table D8. Responses in category listing task – Running

Response	Production freq'	Mean rank	1st ranked
Shouting	26	2.85	5
Whispering	26	2.96	8
Chatting	14	2.79	8
Speaking	10	4.30	4
Conversing	9	6.00	
Discussing	9	6.33	
Gossiping	8	4.75	
Yelling	8	6.63	
Screaming	7	5.14	
Arguing	7	3.71	
Signing	7	6.71	1
Mumbling	5	4.80	
Answering	4	2.25	
Chattering	4	9.25	3
Debating	4	9.50	
Lecturing	4	5.25	
Questioning	4	5.50	
Nattering	3	4.00	
Saying	3	6.67	
Stating	3	6.33	
Stuttering	3	4.67	

Table D9. Responses in category listing task – Talking

Response	Production freq’	Mean rank	1st ranked
Strolling	15	2.93	7
Ambling	12	2.58	1
Striding	10	2.80	6
Power-walking	9	2.89	2
Hiking	8	5.75	1
Sauntering	8	3.75	2
Wandering	8	3.63	
Limping	7	4.29	1
Meandering	7	4.86	
Dawdling	6	3.50	
Marching	6	4.17	1
Shuffling	6	4.17	
Jogging	5	5.40	1
Pacing	5	4.80	1
Running	5	4.60	
Staggering	5	6.00	
Tip toeing	5	7.00	
Rambling	4	7.00	
Stepping	4	7.75	
Fell walking	3	3.67	1
Hopping	3	5.67	1
Moonwalking	3	3.00	1
Speed walking	3	1.00	3
Trekking	3	6.00	1
Trotting	3	3.33	

Table D10. Responses in category listing task – Walking

Appendix E Category Listing – Noun Responses and Quantitative Data

Response	Production freq’	Mean rank	1st ranked
Dog	33	4.00	8
Cats	32	3.75	10
Lion	27	7.89	3
Tiger	25	8.84	1
Elephant	23	9.09	1
Giraffe	23	9.00	1
Cow	22	10.95	1
Horse	22	10.68	2
Fish	20	13.40	
Pig	18	10.83	1
Hamster	17	9.65	
Whale	17	18.47	
Monkey	16	12.38	
Mouse	16	9.56	
Rabbit	16	9.56	2
Sheep	14	10.43	
Bear	13	9.46	1
Bird	13	10.23	
Dolphin	13	17.31	
Snake	13	16.62	
Deer	12	15.75	
Goat	12	17.33	
Leopard	12	13.92	
Rat	11	12.00	
Shark	11	14.45	
Zebra	11	11.36	
Hippo	10	11.50	
Kangaroo	10	18.30	
Antelope	9	15.00	
Donkey	9	13.56	
Guinea pig	9	10.33	
Rhino	9	14.44	

Table E1. Responses in category listing task – Animals

Response	Production freq’	Mean rank	1st ranked
Seal	9	16.89	
Chicken	8	12.25	
Duck	7	10.43	
Gerbil	7	14.71	
Lizard	7	16.29	
Bat	6	16.33	
Cheetah	6	11.83	
Crocodile	6	16.83	
Gorilla	6	12.67	
Hyena	6	14.17	
Panda	6	9.67	1
Polar bear	6	12.83	
Badger	5	15.60	
Koala	5	19.80	
Sea lion	5	20.60	
Squirrel	5	17.40	
Camel	4	6.00	
Fox	4	14.25	
Octopus	4	14.75	
Penguin	4	13.00	1
Spider	4	11.75	
Alligator	3	19.00	
Armadillo	3	18.67	
Boar	3	19.67	
Bull	3	21.00	
Chinchilla	3	12.33	
Crab	3	17.33	
Frog	3	16.00	
Hedgehog	3	26.00	
Meerkat	3	17.00	
Owl	3	17.00	
Walrus	3	18.67	
Wolf	3	8.67	

Table E1. Responses in category listing task – Animals (continued)

Response	Production freq’	Mean rank	1st ranked
Robin	30	6.23	4
Blackbird	23	4.91	7
Sparrow	23	6.04	3
Blue tit	22	8.41	
Penguin	21	11.48	
Duck	20	11.80	
Eagle	20	6.55	
Ostrich	20	11.70	1
Pigeon	20	6.85	2
Seagull	18	10.83	
Crow	16	6.38	
Flamingo	16	12.06	
Parrot	16	9.38	
Chicken	15	13.27	
Pheasant	13	15.85	
Swan	13	11.69	1
Owl	12	8.25	2
Raven	12	8.50	
Hawk	11	8.18	1
Goose	10	15.20	
Peacock	10	14.40	
Starling	10	7.30	
Finch	9	11.89	
Magpie	9	7.78	
Budgie	8	7.88	
Chaffinch	8	12.13	1
Thrush	8	9.13	
Canary	7	11.71	
Heron	7	11.86	
Puffin	7	12.29	
Swallow	7	10.00	1
Turkey	7	17.14	

Table E2. Responses in category listing task – Birds

Response	Production freq'	Mean rank	1st ranked
Dodo	6	12.17	1
Dove	6	8.50	
Emu	6	12.50	
Kestrel	6	10.00	1
Pelican	6	15.67	1
Wren	6	12.00	
Albatross	5	16.20	
Guinea fowl	5	17.00	
Hen	5	15.00	
Partridge	5	16.60	
Vulture	5	14.80	
Bluebird	4	7.25	
Buzzard	4	14.25	
Cockerel	4	16.50	
Cuckoo	4	9.75	
Falcon	4	10.00	
Jay	4	10.25	
Woodpecker	4	9.25	
Cormorant	3	16.00	
Humming bird	3	11.00	
Kingfisher	3	9.33	
Kite	3	11.00	
Kiwi	3	18.67	
Moorhen	3	18.67	
Nightingale	3	7.67	
Osprey	3	9.00	
Red kite	3	8.33	
Rook	3	11.00	
Stork	3	12.00	
Toucan	3	6.67	
Warbler	3	9.33	

Table E2. Responses in category listing task – Birds (continued)

Response	Production freq’	Mean rank	1st ranked
T shirt	33	6.70	4
Shirt	32	7.47	
Trousers	32	5.63	6
Coat	31	11.48	2
Socks	31	9.68	
Jacket	26	11.27	2
Jumper	26	6.81	5
Skirt	26	10.35	2
Hat	25	14.64	3
Shorts	23	9.96	
Jeans	22	7.68	2
Tie	20	11.80	
Dress	19	10.95	1
Scarf	19	11.58	1
Vest	19	12.58	
Blouse	18	12.83	
Bra	18	13.33	
Cardigan	17	9.76	
Hoodie	16	11.50	1
Shoes	15	8.20	3
Gloves	14	13.29	
Tights	13	14.31	
Waistcoat	13	16.62	
Boxer shorts	11	13.09	
Pants	10	8.45	
Knickers	10	13.80	
Suit	10	14.60	
Leggings	9	15.33	
Sweater	9	13.22	
Underwear	8	9.75	
Polo shirt	7	10.86	
Belt	6	12.67	

Table E3. Responses in category listing task – Clothes

Response	Production freq’	Mean rank	1st ranked
Pyjamas	6	9.83	1
Underpants	6	16.00	
Boots	5	12.60	
Anorak	4	14.50	
Cravat	4	14.25	
Dressing gown	4	13.50	
Leg warmers	4	20.50	
Swimming costume	4	15.75	
Tank top	4	9.50	
Blazer	3	15.67	
Boob tube	3	15.00	
Calottes	3	17.33	
Dungarees	3	14.67	
Nightdress	3	12.33	
Pullover	3	11.00	
Stockings	3	22.00	
Strappy top	3	12.00	
Top	3	12.00	
Tracksuit bottoms	3	11.00	

Table E3. Responses in category listing task – Clothes (continued)

Response	Production freq'	Mean rank	1st ranked
Apple	35	1.57	25
Banana	34	3.97	3
Pear	34	6.85	
Orange	33	4.97	1
Grapes	31	9.32	
Strawberry	28	10.43	1
Pineapple	24	8.75	
Raspberry	24	12.54	
Melon	23	14.48	
Peach	23	10.70	
Plum	21	11.71	
Blueberry	20	14.90	
Kiwi fruit	19	9.84	1
Mango	18	12.06	1
Nectarine	18	11.61	
Passionfruit	18	12.50	
Tomato	18	9.33	2
Lemons	17	12.12	
Gooseberry	16	14.88	
Blackberry	14	13.79	
Grapefruit	13	12.69	
Lime	13	12.38	
Cherry	12	11.17	
Clementine	11	13.36	
Satsuma	11	7.73	
Star fruit	11	15.27	
Blackcurrant	10	17.20	
Pomegranate	10	16.20	
Tangerine	9	10.22	
Apricots	8	11.88	
Dragon fruit	7	11.86	
Lychee	7	10.29	

Table E4. Responses in category listing task – Fruit

Response	Production freq’	Mean rank	1st ranked
Watermelon	7	16.57	
Coconut	5	12.20	
Cranberry	5	16.40	
Guava	5	8.00	
Papaya	5	9.80	
Prune	5	10.40	
Redcurrant	5	16.80	
Dates	4	12.50	
Fig	4	16.25	
Loganberry	4	19.25	
Rhubarb	4	17.75	
Avocado	3	9.00	
Mandarin	3	9.33	

Table E4. Responses in category listing task – Fruit (continued)

Response	Production freq’	Mean rank	1st ranked
Chair	34	1.59	26
Table	32	3.31	4
Sofa	29	4.97	2
Bed	27	6.78	1
Desk	27	9.26	
Wardrobe	27	8.22	
Armchair	20	10.00	1
Cupboard	18	7.72	
Chest of drawers	17	9.41	
Stool	17	8.35	
Shelf	14	8.64	
Coffee table	13	7.15	
Bedside table	12	9.75	
Lamp	12	8.42	
Bench	11	9.09	
Cabinet	10	7.60	
Drawers	10	8.50	
Sideboard	9	7.78	
Bookshelf	8	8.75	
Futon	8	6.38	
Bean bag	7	11.14	
Dresser	7	10.29	
Foot stool	6	8.83	
Bookcase	5	9.80	
Chaise longue	5	7.80	
Chest	5	8.20	
Dining table	5	9.00	
Dressing table	5	9.00	
Tv cabinet	5	12.80	
Fireplace	4	12.50	
Rocking chair	4	8.00	
Rug	4	11.00	

Table E5. Responses in category listing task – Furniture

Response	Production freq’	Mean rank	1st ranked
Settee	4	6.75	
Bath	3	15.33	
Carpet	3	7.00	
Cushion	3	10.00	
Filing cabinet	3	17.00	
Hat stand	3	12.00	
Nest of tables	3	9.67	
Pouf	3	10.00	
Sink	3	14.67	

Table E5. Responses in category listing task – Furniture (continued)

Response	Production freq’	Mean rank	1st ranked
Violin	34	6.82	6
Piano	33	5.67	9
Drums	32	8.16	1
Flute	32	7.81	1
Guitar	32	6.22	6
Trumpet	28	8.43	1
Cello	26	7.92	1
Clarinet	25	10.56	2
Triangle	25	13.12	
Oboe	23	7.17	3
Trombone	23	9.61	1
Keyboards	22	12.55	
Recorder	21	11.95	
Saxophone	21	8.57	1
Viola	21	9.81	
Double bass	20	10.15	
Harp	20	9.65	1
Xylophone	15	14.13	
Accordion	13	18.46	
Bassoon	12	10.58	
Organ	11	14.45	
Banjo	10	9.70	1
Cymbal	10	16.00	
Harmonica	10	14.60	
Piccolo	10	14.20	
Tuba	10	10.90	
Bass	9	9.78	
Bass guitar	9	10.67	
Tambourine	9	14.22	
Harpsichord	7	13.29	
Horn	7	12.14	
Bagpipes	6	15.33	

Table E6. Responses in category listing task – Musical instruments

Response	Production freq’	Mean rank	1st ranked
Bongo	6	9.83	
French horn	5	10.20	
Voice	5	12.00	
Cornet	4	11.00	
Mandolin	4	10.75	
Maracas	4	15.75	
Ukulele	4	13.25	
Electric guitar	3	18.00	
Euphonium	3	16.00	
Glockenspiel	3	21.67	
Gong	3	13.67	
Percussion	3	15.33	
Synthesizer	3	13.67	
Tin whistle	3	8.67	1

Table E6. Responses in category listing task – Musical instruments (continued)

Response	Production freq’	Mean rank	1st ranked
Football	35	2.51	16
Rugby	29	4.66	3
Hockey	28	6.43	6
Tennis	28	7.11	2
Swimming	27	9.44	
Basketball	25	9.36	
Cricket	25	8.92	2
Netball	24	8.17	
Badminton	23	10.52	
Volleyball	23	11.65	1
Cycling	17	9.88	1
Athletics	16	9.31	
Baseball	16	13.88	
Ice hockey	15	14.67	
Running	15	9.00	2
Squash	14	10.43	
Lacrosse	13	10.85	1
Rounders	13	11.54	
Horse riding	11	12.27	
American football	11	12.91	
Rowing	11	15.64	
Skiing	11	15.09	
Golf	10	6.90	
Table tennis	10	14.10	
Javelin	9	16.78	
Water polo	8	14.13	
Darts	8	13.38	
Diving	8	12.75	
Wrestling	8	14.00	
Boxing	7	9.14	
Gymnastics	7	12.57	
Snooker	7	11.71	

Table E7. Responses in category listing task – Sports

Response	Production freq'	Mean rank	1st ranked
Snowboarding	7	14.29	
Polo	6	16.17	
Sailing	6	15.83	
Surfing	6	15.33	
Archery	5	16.20	
Aussie rules football	5	11.60	
Handball	5	15.00	
Judo	5	13.00	
Rugby league	5	13.20	
Shooting	5	17.00	
Shot put	5	12.80	
Trampolining	5	12.60	
Ultimate frisbee	5	17.20	
Abseiling	4	18.75	
Hurdles	4	11.50	
Karate	4	13.75	
Long jump	4	16.00	
Rugby union	4	3.00	
Bowling	3	18.00	
Climbing	3	14.00	
Cross-country	3	8.33	
Curling	3	18.67	
Dancing	3	8.67	1
Fencing	3	17.67	
Ice skating	3	17.33	
Pool	3	12.33	

Table E7. Responses in category listing task – Sports (continued)

Response	Production freq’	Mean rank	1st ranked
Hammer	34	2.44	21
Screwdriver	34	4.26	5
Drill	29	5.79	1
Saw	27	4.30	3
Spanner	22	5.82	3
Pliers	17	9.65	
Wrench	16	5.56	
Chisel	14	6.71	1
Knife	12	7.83	1
Plane	12	7.00	
Spirit level	12	8.67	
Sander	10	8.60	
Chainsaw	9	8.44	
Fork	9	8.22	
Nails	9	7.56	
Spade	9	8.33	
Tape measure	9	8.78	
Mallet	7	10.00	
Allen key	6	8.83	
Hacksaw	6	8.33	
Axe	5	7.20	
Rake	5	11.60	
Ruler	5	6.40	
Scissors	5	7.20	
Spoon	5	10.60	
File	4	7.50	
Jigsaw	4	10.75	
Ratchet	4	4.75	
Screw	4	8.75	
Trowel	4	12.75	
Clamp	3	9.67	
Glue gun	3	9.33	

Table E8. Responses in category listing task – Tools

Response	Production freq’	Mean rank	1st ranked
Lawn mower	3	9.00	
Shifter	3	8.33	
Soldering iron	3	13.00	
Tweezers	3	11.33	

Table E8. Responses in category listing task – Tools (continued)

Response	Production freq’	Mean rank	1st ranked
Bus	34	2.18	9
Car	34	1.88	21
Plane	32	7.03	1
Bicycle	31	5.74	
Train	29	5.69	
Motorbike	27	8.07	
Boat	24	10.17	
Helicopter	20	10.40	
Coach	17	8.29	
Metro	17	8.18	1
Scooter	16	11.06	
Ferry	13	11.92	
Taxi	12	9.67	
Lorry	11	9.64	
Skateboard	10	11.40	
Tram	10	9.00	
Underground	10	11.20	
Hovercraft	9	10.00	
Ship	9	10.56	
Van	8	11.38	
Foot	7	7.29	
Horse	7	13.14	
Roller skates	6	13.00	
Jet ski	6	17.17	
Moped	6	11.00	
Submarine	6	14.67	
Rollerblades	6	12.33	
Minibus	5	10.00	
Monorail	5	12.00	
Skis	5	12.80	
Truck	5	11.60	
Cable car	4	13.25	

Table E9. Responses in category listing task – Transport

Response	Production freq'	Mean rank	1st ranked
Hot air balloon	4	14.50	
Rickshaw	4	13.25	
Tricycle	4	12.00	
Yacht	4	14.25	
Ambulance	3	11.33	
Carriage	3	11.33	
Glider	3	14.67	
Hang glider	3	15.33	
Horse and carriage	3	13.00	
Jet	3	10.67	
Pogo stick	3	19.00	
Rocket	3	12.33	
Subway	3	8.00	
Tandem	3	10.67	
Tube	3	6.67	
Unicycle	3	8.67	

Table E9. Responses in category listing task – Transport (continued)

Response	Production freq’	Mean rank	1st ranked
Carrot	34	2.47	20
Potato	33	4.24	6
Broccoli	30	6.43	3
Peas	28	9.00	1
Onion	25	9.60	
Cauliflower	24	9.21	
Turnip	24	8.75	
Pepper	23	9.17	
Cabbage	22	7.77	
Parsnip	20	8.50	1
Lettuce	19	9.00	1
Swede	19	9.32	
Aubergine	18	11.06	
Sweet potato	18	10.89	
Sweetcorn	18	10.61	
Courgette	17	9.24	2
Brussel sprouts	16	10.19	
Cucumber	16	10.06	1
Mushroom	14	9.57	
Leek	12	12.08	
Spinach	10	9.90	
Broad beans	9	10.44	
Radish	9	11.89	
Beetroot	8	11.25	
Celery	8	12.38	
Green beans	8	8.50	
Mange tout	8	11.00	
Beans	7	10.43	
Tomato	7	8.71	
Marrow	6	11.83	
Pumpkin	6	12.83	
Runner beans	6	8.83	

Table E10. Responses in category listing task – Vegetables

Response	Production freq’	Mean rank	1st ranked
Butternut squash	4	14.50	
Rocket	4	13.50	
Spring onion	4	15.75	
Asparagus	3	12.33	
Celeriac	3	15.67	
Red onion	3	10.67	
Shallots	3	13.00	

Table E10. Responses in category listing task – Vegetables (continued)

Appendix F Typicality Rating – Verb Data

Response	Mean typicality	SD	Not in category	Frequency
Snapping	1.47	0.853	0	2013
Smashing	1.72	1.093	0	754
Dropping	1.94	1.171	2	5786
Destroying	2.20	1.654	7	3154
Ripping	2.24	1.335	1	653
Shattering	2.25	1.398	0	278
Cracking	2.28	1.262	0	777
Crushing	2.41	1.282	1	497
Bashing	2.45	1.353	1	158
Pulling apart	2.46	1.325	0	-
Tearing	2.51	1.322	2	1614
Cutting	2.69	1.619	2	7309
Hitting	2.73	1.359	4	6842
Stamping on	2.82	1.633	1	556
Burning	3.09	1.756	3	2004
Squashing	3.10	1.528	2	114
Splitting	3.18	1.373	2	1447
Crashing	3.24	1.682	5	1575
Jumping on	3.30	1.801	2	3725
Kicking	3.47	1.716	3	2293
Chipping	3.53	1.574	4	193
Throwing	3.56	1.767	7	7010
Bending	3.70	1.854	3	1874
Falling	3.74	1.833	15	18902
Slicing	3.75	1.766	2	298
Denting	3.79	1.651	4	80
Shredding	4.00	1.627	2	51
Forcing	4.05	1.834	8	4807
Splintering	4.21	1.594	5	49
Pulling	4.37	1.805	5	9336

Table F1. Mean typicality ratings and BNC lexical frequency values – Breaking

Response	Mean typicality	SD	Not in category	Frequency
Washing	1.17	0.509	0	2320
Wiping	1.17	0.375	0	1608
Hoovering	1.40	0.664	0	66
Vacuuming	1.41	0.722	0	30
Sweeping	1.48	0.685	0	1415
Mopping	1.56	0.803	0	251
Dusting	1.59	0.979	0	319
Scrubbing	1.63	0.954	0	253
Showering	1.81	1.255	1	213
Bathing	1.93	1.283	1	316
Disinfecting	2.02	1.186	0	34
Polishing	2.06	1.225	0	258
Shampoo	2.07	1.290	1	30
Rinsing	2.13	1.325	0	165
Scouring	2.27	1.415	0	219
Cleansing	2.37	1.568	2	129
Bleaching	2.40	1.437	0	53
Brushing	2.49	1.609	6	1193
Dry cleaning	2.53	1.591	1	-
Hosing	2.56	1.317	0	26
Soaking	2.61	1.496	1	499
Sterilising	2.75	1.663	0	31
Dabbing	2.97	1.323	0	173
Rubbing	3.04	1.551	3	1567
Shining	3.06	1.448	2	840
Lathering	3.13	1.647	4	11
Spraying	3.17	1.564	2	310
Scraping	3.48	1.418	2	456
Steaming	3.68	1.677	3	155
Drying	4.24	1.789	20	1325

Table F2. Mean typicality ratings and BNC lexical frequency values – Cleaning

Response	Mean typicality	SD	Not in category	Frequency
Frying	1.26	0.612	0	87
Grilling	1.33	0.665	0	88
Roasting	1.38	0.809	0	86
Baking	1.45	0.840	0	231
Boiling	1.59	0.948	0	555
Toasting	1.62	0.912	0	154
Stewing	1.69	0.944	0	56
Microwaving	1.73	1.085	1	19
Barbequing	1.76	1.170	0	-
Stir-frying	1.78	1.050	0	-
Simmering	1.88	1.151	1	74
Slow cooking	2.06	1.296	0	-
Heating	2.15	1.452	0	440
Steaming	2.21	1.315	0	155
Poaching	2.22	1.412	0	141
Scrambling	2.29	1.339	0	497
Deep frying	2.37	1.495	0	-
Sautéing	2.72	1.834	1	-
Char-grilling	2.75	1.793	0	-
Searing	2.79	1.734	3	46
Braising	2.81	1.793	1	4
Chopping	2.96	1.809	17	412
Mixing	3.07	1.657	13	1211
Griddling	3.11	1.863	2	1
Melting	3.26	1.793	3	747
Blanching	3.38	1.970	0	44
Burning	4.01	1.964	10	2004
Freezing	4.96	1.962	28	827

Table F3. Mean typicality ratings and BNC lexical frequency values – Cooking

Response	Mean typicality	SD	Not in category	Frequency
Slicing	1.24	0.530	0	298
Chopping	1.29	0.669	0	412
Dicing	1.87	1.200	0	55
Sawing	1.92	1.096	0	150
Trimming	1.93	1.101	0	312
Scissoring	2.11	1.568	1	11
Snipping	2.14	1.178	0	78
Carving	2.18	1.458	0	416
Slashing	2.44	1.411	0	341
Severing	2.60	1.582	1	268
Hacking	3.07	1.531	1	292
Tearing	3.09	1.847	8	653
Ripping	3.09	1.704	11	1614
Splitting	3.63	1.810	5	1447
Grating	3.64	1.801	2	268
Piercing	3.94	1.780	9	237
Stabbing	3.97	1.883	6	454
Snapping	4.11	1.735	11	2013

Table F4. Mean typicality ratings and BNC lexical frequency values – Cutting

Response	Mean typicality	SD	Not in category	Frequency
Punching	1.21	0.722	0	591
Slapping	1.72	0.999	0	619
Smacking	1.75	1.132	0	393
Thumping	1.83	1.192	1	503
Whacking	2.05	1.410	0	86
Striking	2.07	1.381	0	3882
Beating	2.09	1.326	2	4960
Bashing	2.26	1.411	2	158
Whalloping	2.32	1.663	2	-
Hammering	2.59	1.649	0	485
Banging	2.69	1.542	2	719
Colliding	2.73	1.636	1	247
Elbowing	2.74	1.262	1	85
Smashing	2.80	1.455	5	754
Kicking	2.91	1.936	6	2293
Knocking	2.99	1.735	1	2656
Batting	3.04	1.795	3	323
Jabbing	3.09	1.530	1	194
Pushing	3.29	1.753	6	4194
Head-butting	3.37	2.009	1	13
Tapping	3.47	1.834	2	1234
Crashing	3.76	1.913	9	1575
Flicking	4.17	1.567	4	734
Nudging	4.30	1.748	4	331
Poking	4.31	1.743	6	596
Throwing	4.68	2.008	26	7010

Table F5. Mean typicality ratings and BNC lexical frequency values – Hitting

Response	Mean typicality	SD	Not in category	Frequency
Assembling	1.56	0.907	0	674
Constructing	1.60	1.017	0	1680
Cooking	1.63	1.033	0	1655
Creating	1.64	1.092	1	12816
Building	1.66	1.085	0	9223
Manufacturing	1.71	1.095	0	615
Baking	1.73	0.966	0	231
Moulding	2.05	1.323	0	222
Sculpting	2.12	1.588	0	31
Crafting	2.21	1.431	1	76
Painting	2.39	1.346	3	1631
Knitting	2.44	1.638	0	912
Sewing	2.49	1.419	0	268
Joining	2.52	1.621	0	12071
Drawing	2.55	1.459	3	1162
Inventing	2.63	1.668	2	914
Designing	2.64	1.480	4	1957
Carving	2.78	1.635	1	416
Sticking	2.79	1.591	2	3083
Writing	2.89	1.647	3	23173
Welding	2.95	1.815	0	95
Gluing	3.01	1.662	0	111
Recording	3.14	1.708	3	4052
Mixing	3.17	1.697	1	1211
Printing	3.19	1.629	4	988
Typing	3.26	1.728	5	474
Hammering	3.29	1.654	6	485
Cutting	3.41	1.728	4	7309
Fixing	3.49	1.762	14	1551
Sawing	3.64	1.597	6	150
Folding	3.69	1.673	5	547

Table F6. Mean typicality ratings and BNC lexical frequency values – Making

Response	Mean typicality	SD	Not in category	Frequency
Speaking	1.10	0.520	1	18311
Chatting	1.20	0.508	0	1055
Saying	1.31	0.706	2	272787
Discussing	1.51	0.743	1	7913
Conversing	1.65	1.132	2	153
Chattering	1.81	1.132	0	207
Gossiping	1.97	1.096	2	167
Answering	2.05	1.146	3	6993
Nattering	2.09	1.187	0	18
Questioning	2.23	1.136	2	2003
Whispering	2.25	1.094	0	2631
Mumbling	2.48	1.167	0	538
Debating	2.49	1.063	3	668
Stating	2.55	1.402	3	5508
Arguing	2.68	1.254	2	9356
Lecturing	2.95	1.396	1	384
Shouting	3.04	1.538	3	4455
Yelling	3.42	1.577	2	1040
Stuttering	4.01	1.686	0	90
Screaming	4.17	1.745	8	1878
Signing	4.55	1.848	3	4360

Table F7. Mean typicality ratings and BNC lexical frequency values – Talking

Response	Mean typicality	SD	Not in category	Frequency
Strolling	1.56	0.981	0	637
Wandering	2.02	1.233	1	1703
Striding	2.21	1.300	0	978
Hiking	2.37	1.334	0	55
Dawdling	2.43	1.322	1	65
Speed walking	2.52	1.572	0	-
Power-walking	2.59	1.685	0	-
Stepping	2.59	1.652	2	3678
Trekking	2.61	1.358	0	95
Pacing	2.70	1.488	0	446
Ambling	2.74	1.547	0	147
Sauntering	2.75	1.438	0	164
Meandering	3.02	1.653	0	111
Running	3.11	2.120	20	26170
Staggering	3.11	1.414	0	405
Marching	3.15	1.788	0	1127
Fell-walking	3.19	1.701	1	-
Jogging	3.38	1.878	11	264
Rambling	3.50	1.689	1	88
Limping	3.58	1.710	1	232
Tip-toeing	3.75	1.777	0	134
Shuffling	3.78	1.689	1	503
Trotting	4.45	1.633	5	331
Hopping	5.18	1.497	13	334
Moonwalking	5.47	1.689	8	-

Table F8. Mean typicality ratings and BNC lexical frequency values – Walking

Appendix G Typicality Rating – Noun Data

Response	Mean typicality	Lexical frequency
Blackbird	1.000	238
Sparrow	1.047	204
Robin	1.093	227
Starlings	1.182	182
Thrush	1.186	350
Pigeon	1.250	856
Crow	1.256	337
Seagull	1.364	134
Swallow	1.419	217
Wren	1.465	67
Dove	1.477	309
Cuckoo	1.535	491
Hawk	1.698	348
Woodpecker	1.727	131
Raven	1.744	103
Nightingale	1.773	75
Owl	1.773	1621
Eagle	1.791	1757
Parrot	1.837	493
Pheasant	1.930	317
Canary	1.953	103
Budgie	1.977	133
Swan	2.000	582
Chicken	2.070	1992
Duck	2.159	714
Falcon	2.182	406
Hen	2.182	845
Albatross	2.205	55
Peacock	2.295	182
Vulture	2.295	287
Goose	2.302	864

Table G1. Mean typicality ratings and BNC lexical frequency values – Birds

Response	Mean typicality	Lexical frequency
Warbler	2.310	146
Heron	2.326	84
Osprey	2.326	20
Turkey	2.360	887
Stork	2.476	60
Buzzard	2.477	209
Flamingo	2.651	66
Pelican	2.721	143
Puffin	2.905	80
Ostrich	3.047	105
Toucan	3.143	23
Penguin	3.227	212
Emu	3.512	291

Note.: Typicality ratings from Hampton & Gardiner (1983)

Table G1. Mean typicality ratings and BNC lexical frequency values – Birds (continued)

Response	Mean typicality	Lexical frequency
Dress	1.000	3458
Skirt	1.022	1757
Trousers	1.022	2428
Shirt	1.044	3381
Jeans	1.067	1233
Jumper	1.178	722
Jacket	1.244	3347
Suit	1.267	2425
Blouse	1.289	578
Coat	1.289	3696
Cardigan	1.422	319
Socks	1.600	1123
Anorak	1.822	250
Pants	1.822	556
Dungarees	1.844	78
Tights	1.955	370
Vest	1.956	346
Shorts	2.000	883
Stockings	2.044	725
Pyjamas	2.205	438
Waistcoat	2.333	310
Scarf	2.644	685
Swimming costume	2.756	-
Ties	2.800	2344
Gloves	2.844	1345
Hat	2.844	3697
Belt	3.133	2553
Cravat	3.289	43

Note.: Typicality ratings from Hampton & Gardiner (1983)

Table G2. Mean typicality ratings and BNC lexical frequency values – Clothes

Response	Mean typicality	Lexical frequency
Apple	1.023	3444
Orange	1.023	1148
Pear	1.163	445
Banana	1.233	968
Grapefruit	1.256	5
Strawberry	1.256	612
Grapes	1.279	795
Plum	1.302	399
Cherry	1.419	695
Peach	1.419	512
Pineapple	1.419	326
Lemons	1.512	1334
Tangerine	1.512	23
Mandarin	1.605	322
Satsuma	1.643	24
Raspberry	1.651	273
Blackberry	1.721	188
Apricots	1.814	139
Melon	1.814	268
Blackcurrant	1.881	127
Gooseberry	2.047	104
Lime	2.093	619
Watermelon	2.140	26
Redcurrant	2.429	23
Nectarine	2.615	25
Mango	2.791	98
Blueberry	2.814	17
Cranberry	2.814	94
Fig	2.837	1930
Pomegranate	2.837	42
Prune	2.884	77

Table G3. Mean typicality ratings and BNC lexical frequency values – Fruit

Response	Mean typicality	Lexical frequency
Dates	2.929	12742
Guava	3.485	26
Coconut	3.581	364

Note.: Typicality ratings from Hampton & Gardiner (1983)

Table G3. Mean typicality ratings and BNC lexical frequency values – Fruit (continued)

Response	Mean typicality	Lexical frequency
Chair	1.000	8491
Armchair	1.039	891
Table	1.039	21594
Settee	1.098	341
Sofa	1.098	1044
Bed	1.176	16664
Wardrobe	1.216	1072
Dresser	1.510	291
Desk	1.529	4414
Sideboard	1.569	215
Cupboard	1.647	1840
Stool	1.706	1087
Cabinet	1.765	6759
Bookcase	1.824	214
Chest	2.216	3745
Bench	2.235	2522
Shelf	2.627	2530
Sink	3.588	887

Note.: Typicality ratings from Hampton & Gardiner (1983)

Table G4. Mean typicality ratings and BNC lexical frequency values – Furniture

Response	Mean typicality	Lexical frequency
Football	1.000	6674
Rugby	1.000	2745
Tennis	1.022	2522
Badminton	1.133	189
Basketball	1.178	209
Hockey	1.200	597
Squash	1.267	314
Swimming	1.400	1397
Baseball	1.523	414
Running	1.556	1428
Golf	1.733	3393
Volleyball	1.756	87
Table tennis	1.844	-
Boxing	1.956	1187
Sailing	1.956	668
Javelin	1.978	79
Lacrosse	2.089	19
Skiing	2.111	327
Gymnastics	2.178	93
Rowing	2.182	316
Polo	2.356	594
Horse riding	2.378	-
Fencing	2.400	361
Handball	2.409	32
Archery	2.444	105
Wrestling	2.489	224
Judo	2.545	133
Diving	2.556	102
Snooker	2.689	325
Shooting	2.756	873
Karate	2.867	308

Table G5. Mean typicality ratings and BNC lexical frequency values – Sport

Response	Mean typicality	Lexical frequency
Trampolining	2.978	-
Pool	3.244	5478
Surfing	3.267	12
Dancing	5.156	589

Note.: Typicality ratings from Hampton & Gardiner (1983)

Table G5. Mean typicality ratings and BNC lexical frequency values – Sport (continued)

Response	Mean typicality	Lexical frequency
Car	1.000	33944
Bus	1.109	6124
Taxi	1.174	2041
Van	1.196	2341
Lorry	1.370	1977
Motorbike	1.522	374
Train	1.696	6547
Scooter	1.957	102
Tube	1.978	2939
Ambulance	2.089	1788
Bicycle	2.109	1031
Tram	2.435	765
Plane	2.630	4396
Carriage	2.848	2306
Ferry	2.957	1053
Hovercraft	2.978	-
Boat	3.043	7173
Helicopter	3.130	1531
Tricycle	3.196	50
Ship	3.239	6294
Cable car	3.696	-
Rickshaw	3.773	37
Submarine	4.022	740
Glider	4.109	602
Hot air balloon	4.239	-
Hang glider	4.565	-
Roller skates	4.848	5
Skateboard	4.891	60
Hot air balloon	4.239	-

Note.: Typicality ratings from Hampton & Gardiner (1983)

Table G6. Mean typicality ratings and BNC lexical frequency values – Transport

Response	Mean typicality	Lexical frequency
Carrot	1.000	854
Cabbage	1.021	479
Cauliflower	1.104	78
Beans	1.125	1736
Peas	1.146	776
Potato	1.146	2458
Onion	1.375	1211
Lettuce	1.447	432
Swede	1.543	309
Turnip	1.604	102
Sweetcorn	1.622	29
Broccoli	1.638	132
Leek	1.667	187
Spinach	1.681	197
Parsnip	1.702	78
Beetroot	1.766	59
Cucumber	1.936	248
Celery	1.957	197
Asparagus	1.958	100
Courgette	1.977	148
Mushroom	2.021	783
Radish	2.125	58
Marrow	2.170	332
Aubergine	2.417	70
Shallots	2.689	212
Tomato	2.771	1461
Pepper	3.063	1082

Note.: Typicality ratings from Hampton & Gardiner (1983)

Table G7. Mean typicality ratings and BNC lexical frequency values – Vegetables

Appendix H Similarity Rating - Stimuli

	High-typicality		Low-typicality	
	Item	*M* typicality	Item	*M* typicality
Break	Drop	1.94	Bend	3.70
	Smash	1.72	Dent	3.79
	Snap	1.47	Shred	4.00
	Rip	2.24	Splinter	4.21
Cook	Bake	1.45	Blanch	3.38
	Fry	1.26	Braise	2.81
	Grill	1.33	Griddle	3.11
	Roast	1.38	Melt	3.26
Cut	Chop	1.29	Grate	3.64
	Dice	1.87	Hack	3.07
	Saw	1.92	Split	3.63
	Slice	1.24	Stab	3.97
Make	Assemble	1.56	Fold	3.69
	Build	1.66	Mix	3.17
	Construct	1.60	Print	3.19
	Create	1.64	Type	3.26

Table H1. Stimuli for verb similarity rating task

Appendix I Verb Semantic Feature Analysis – Across and Within Category Analysis Stimuli

Break	Clean	Cook	Cut
Bend	Brush	Bake	Chop
Burn	Shine	Boil	Saw
Crash	Spray	Burn	Snap
Cut	Steam	Chop	Stab
Destroy	Wash	Fry	
Drop		Grill	
Fall		Mix	
Hit		Roast	
Kick		Steam	
Pull			
Smash			
Snap			
Throw			

Hit	Make	Talk	Walk
Crash	Bake	Argue	Hop
Hammer	Build	Chat	Jog
Kick	Construct	Say	Limp
Knock	Cook	Scream	March
Punch	Cut	Shout	Run
Push	Draw	Speak	Stagger
Slap	Fix	Whisper	Step
Smash	Hammer	Yell	Wander
Throw	Mix		
	Paint		

Table I1. Verbs taken from Vinson & Vigliocco (2008) feature norms for feature analysis

Appendix J Verb Semantic Feature Analysis – Feature Distinctiveness by Category Analysis

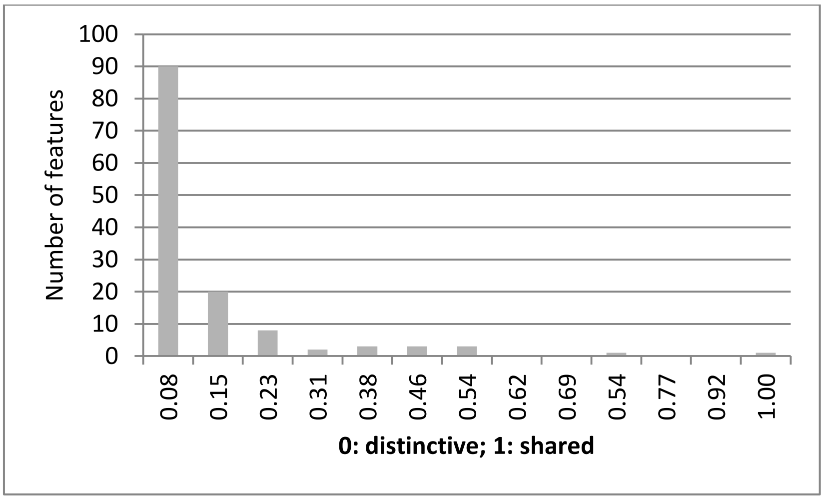

Figure J1 Distribution of feature distinctiveness – breaking

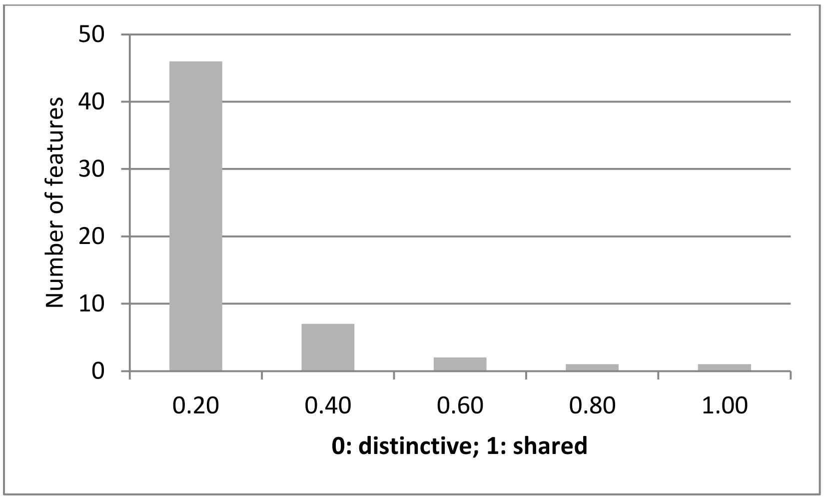

Figure J2 Distribution of feature distinctiveness – cleaning

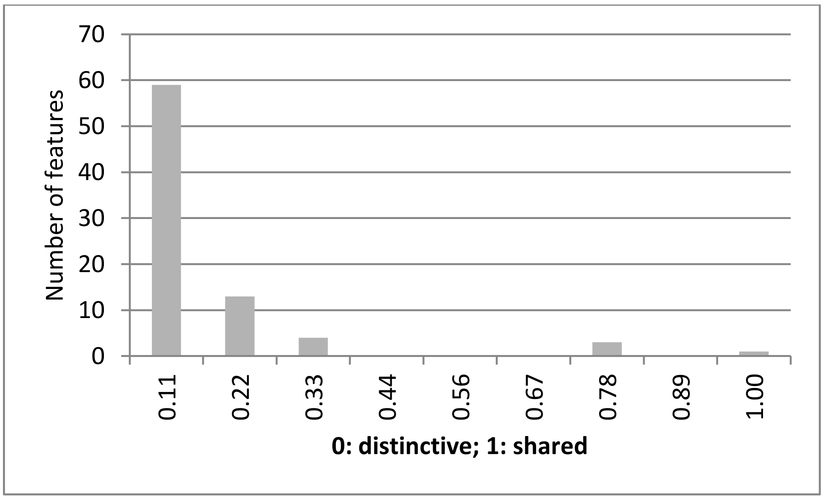

Figure J3 Distribution of feature distinctiveness – cooking

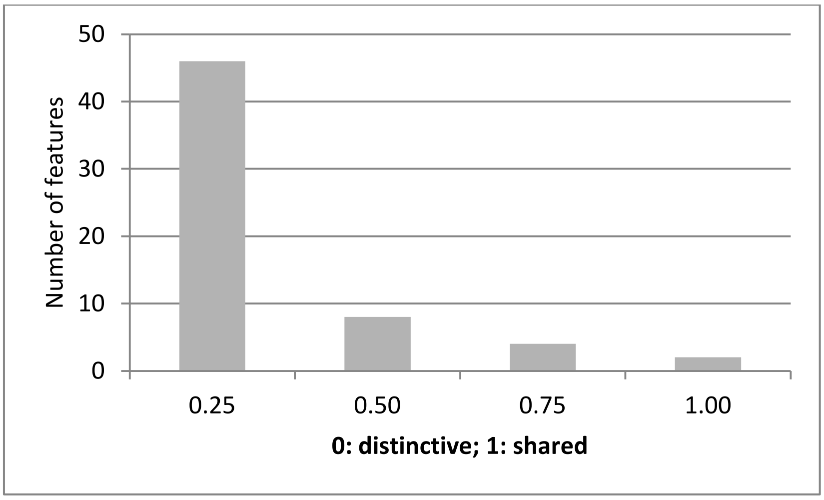

Figure J3 Distribution of feature distinctiveness – cutting

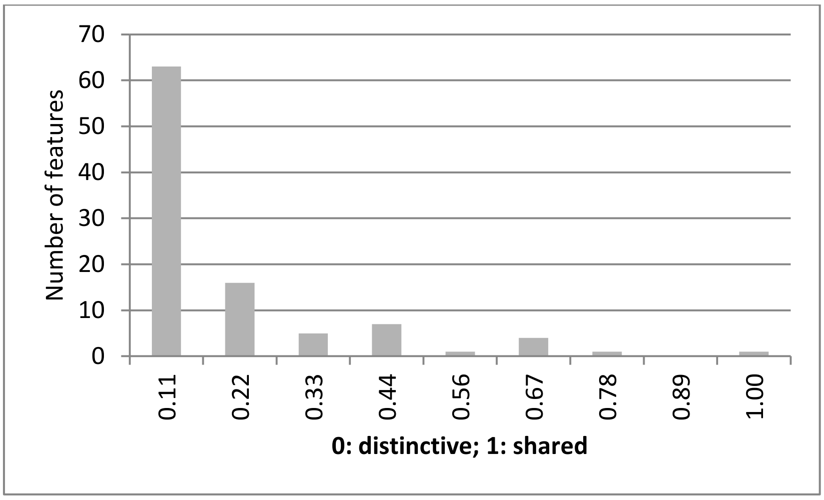

Figure J5 Distribution of feature distinctiveness – hitting

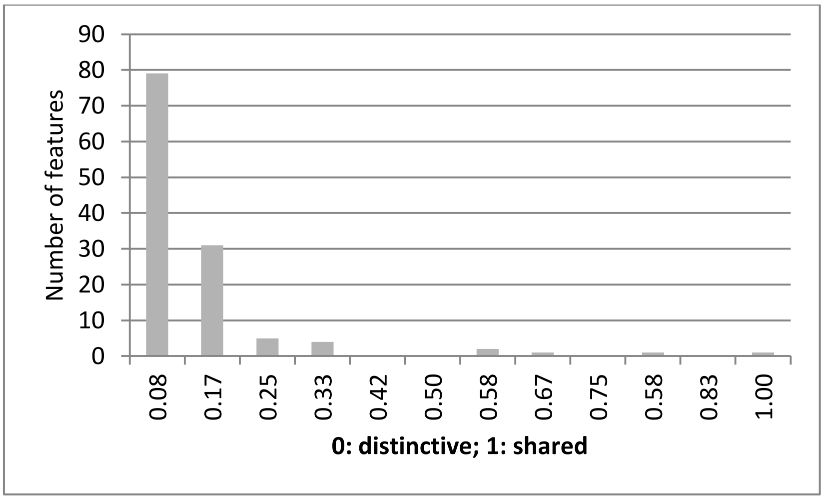

Figure J6 Distribution of feature distinctiveness – making

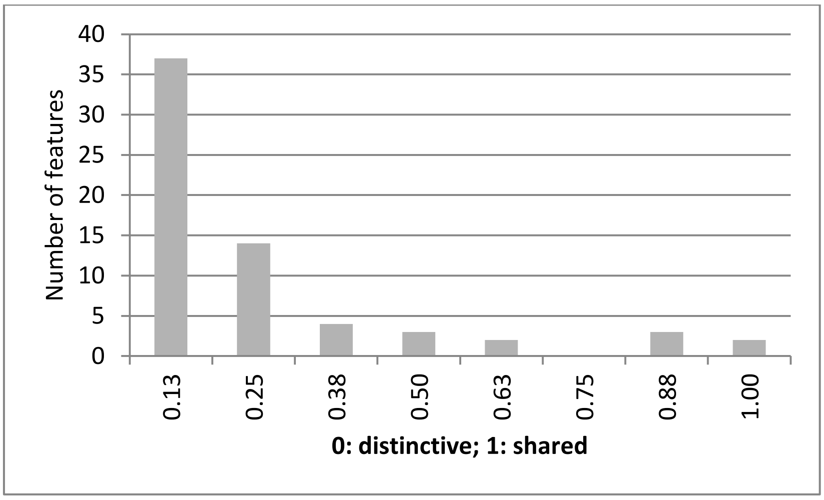

Figure J7 Distribution of feature distinctiveness – talking

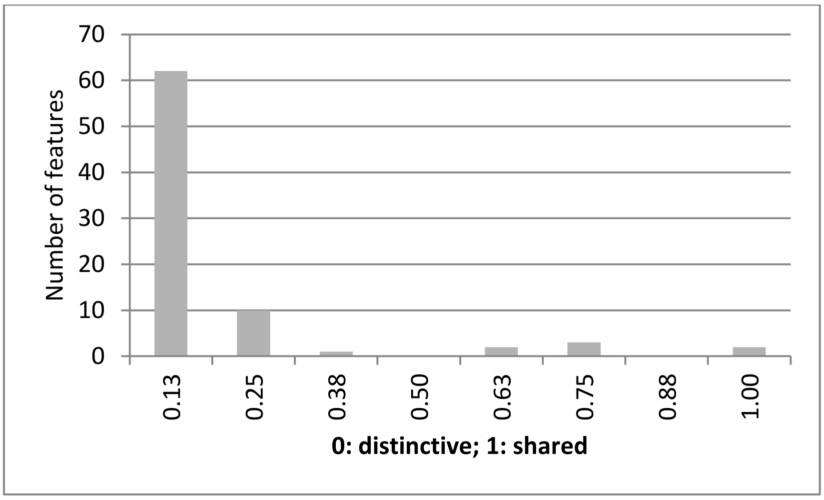

Figure J8 Distribution of feature distinctiveness – walking

Appendix K Verb Semantic Feature Analysis – Feature Type by Level of Feature Distinctiveness Analysis

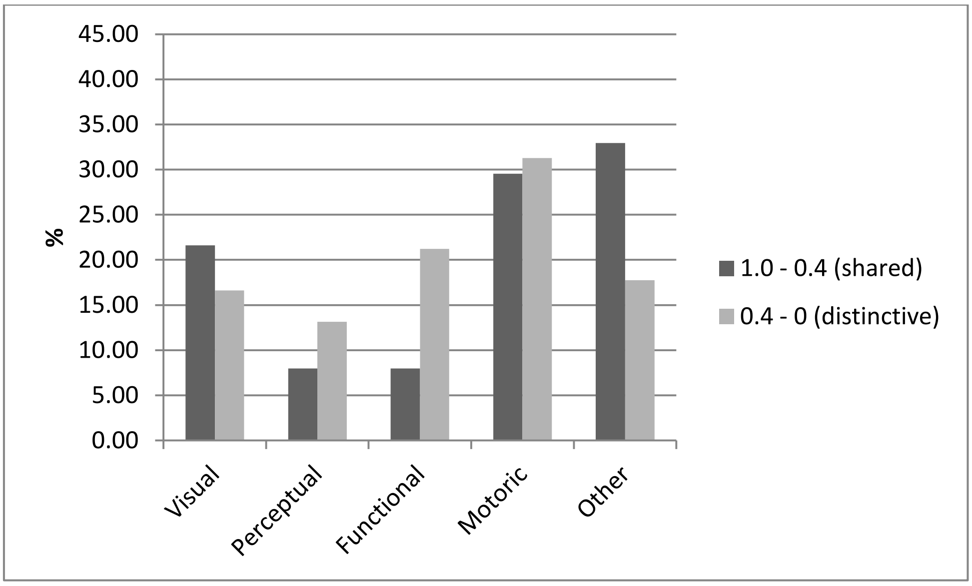

Figure K1 Feature type distribution by feature distinctiveness - break

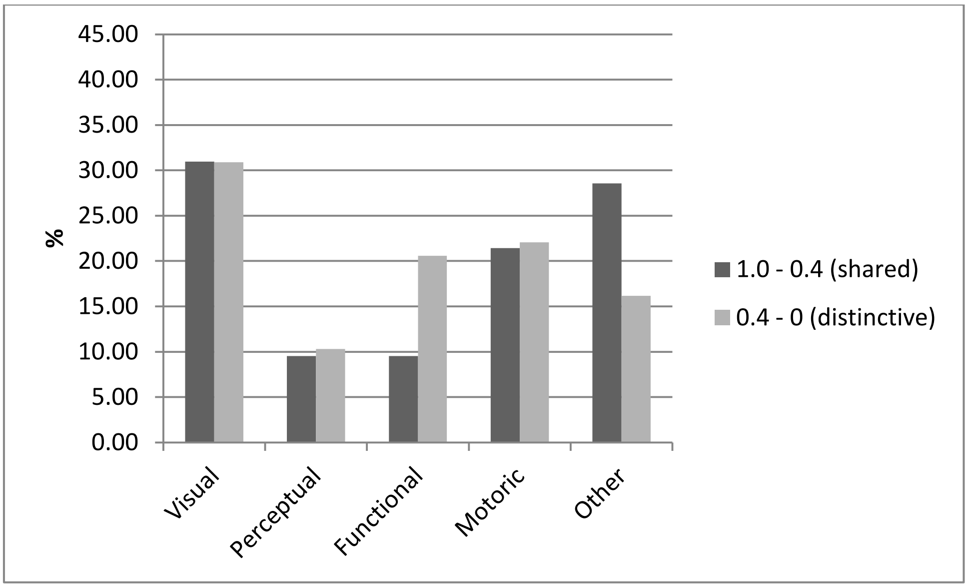

Figure K2 Feature type distribution by feature distinctiveness - clean

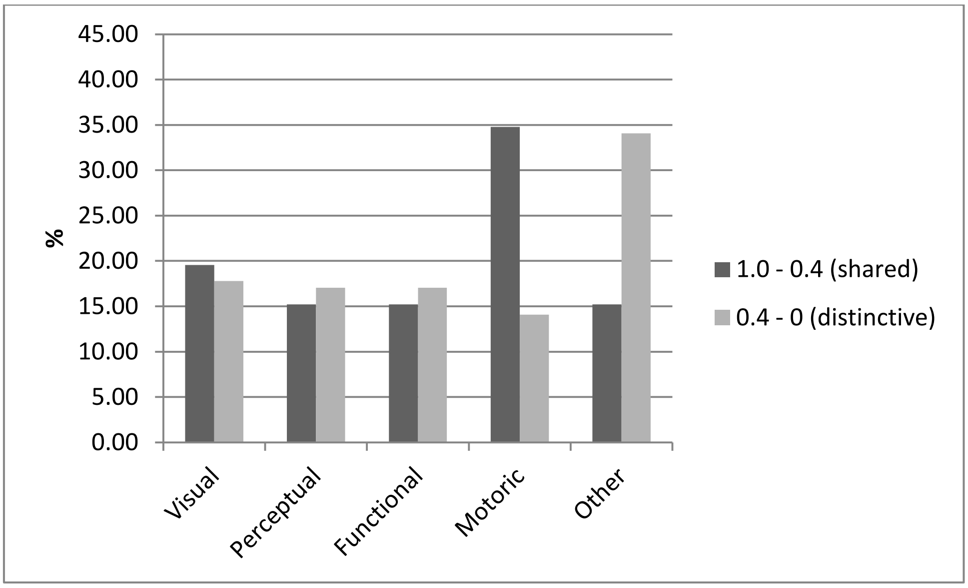

Figure K3 Feature type distribution by feature distinctiveness - cook

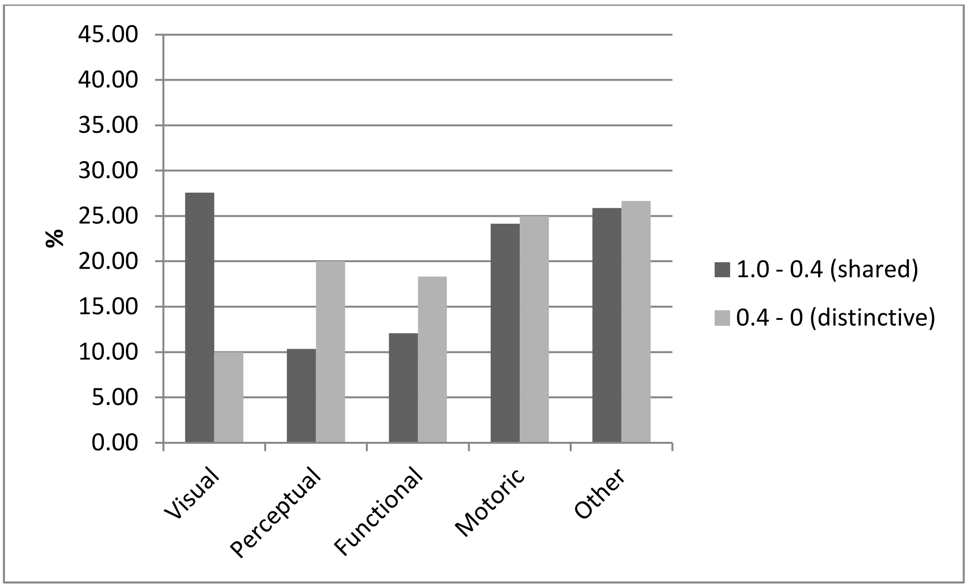

Figure K4 Feature type distribution by feature distinctiveness – cut

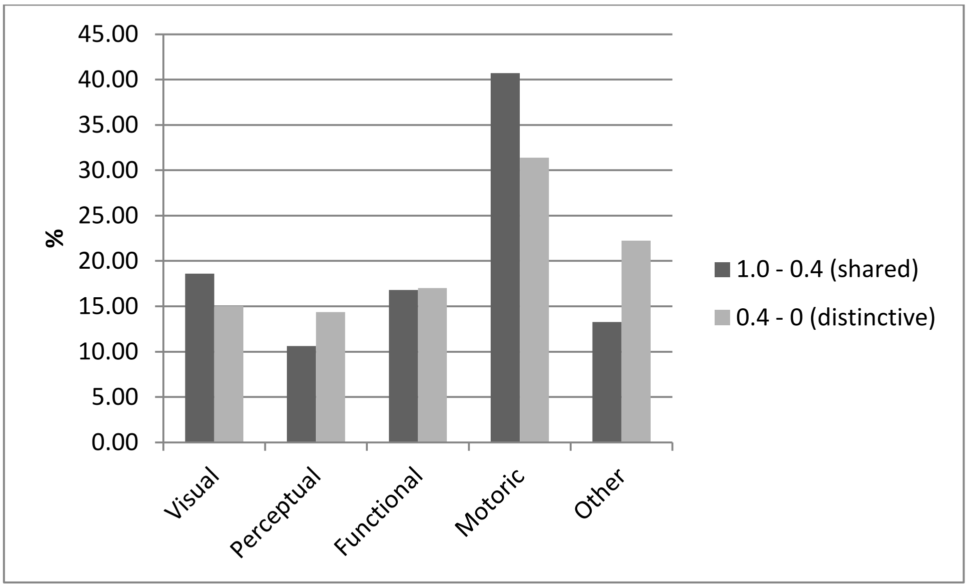

Figure K5 Feature type distribution by feature distinctiveness - hit

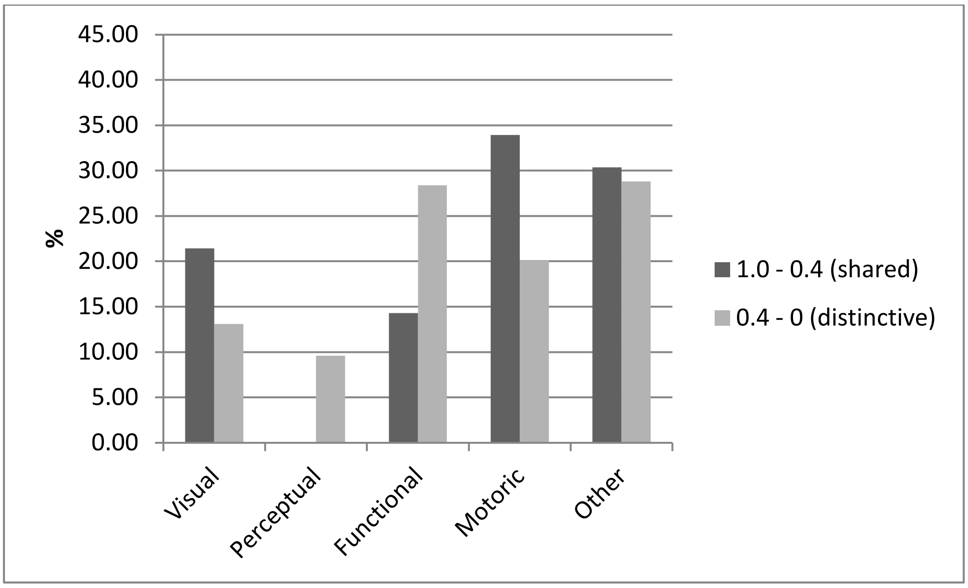

Figure K6 Feature type distribution by feature distinctiveness – make

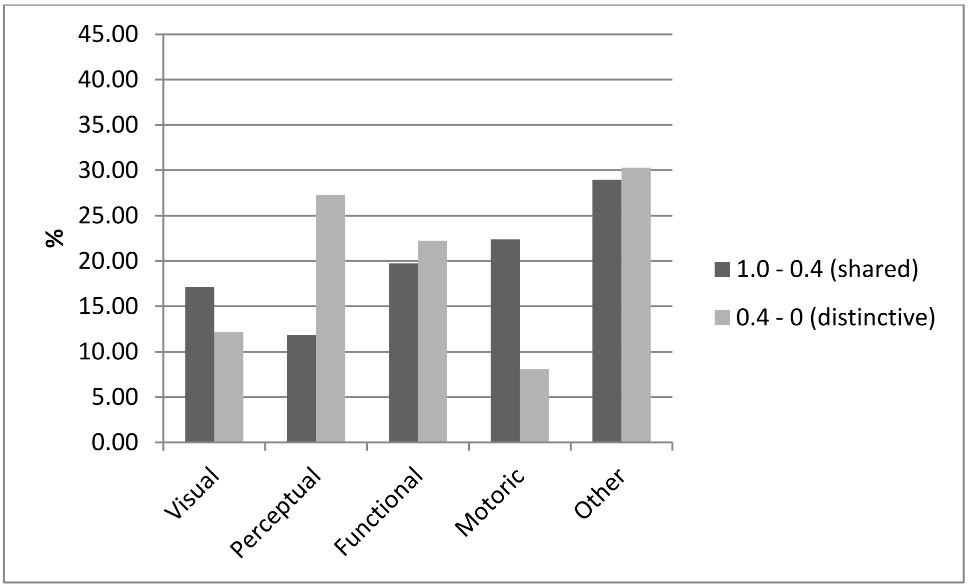

Figure K7 Feature type distribution by feature distinctiveness - talk

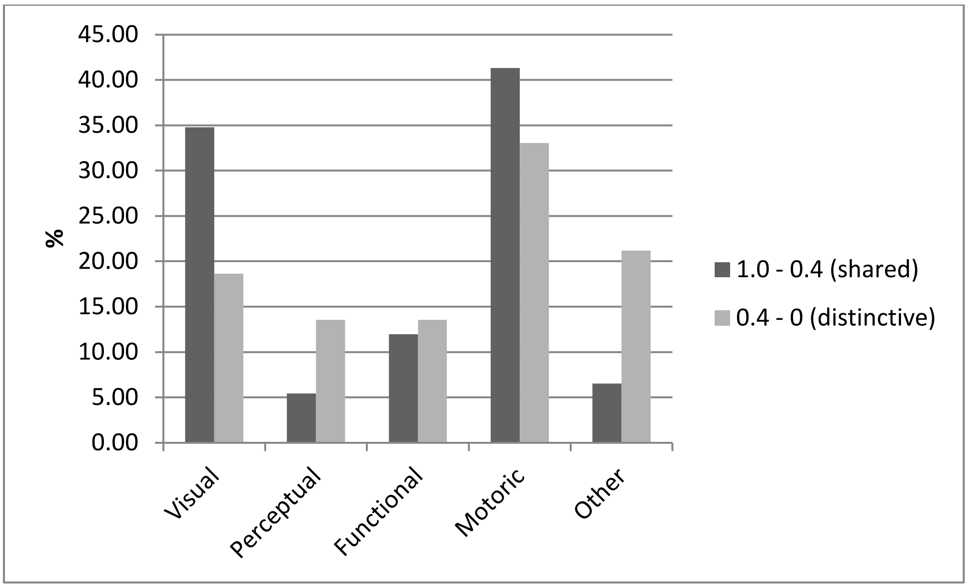

Figure K8 Feature type distribution by feature distinctiveness - walk

Appendix L Verb Semantic Feature Analysis – Low Distinctiveness Features by Category

Category	Feature	Feature type	Distinctiveness
Break	Action	Visual, motoric	1.00
	Object	Other	0.77
	Break	Functional, motoric	0.54
	Humans	Other	0.54
	Hurt	Perceptual	0.54
	Intentional	Other	0.46
	Move	Visual, motoric	0.46
	Something	Other	0.46

Category	Feature	Feature type	Distinctiveness
Clean	Action	Visual, motoric	1.00
	Clean	Visual, functional	0.80
	Liquid	Other	0.60
	Water	Other	0.60
	Air	Other	0.40
	Emit	Visual, perceptual, motoric	0.40
	Hand	Motoric	0.40
	Humans	Other	0.40
	Mist	Visual	0.40
	Object	Other	0.40
	Wet	Perceptual	0.40

Category	Feature	Feature type	Distinctiveness
Cook	Action	Visual, motoric	1.00
	Cook	Functional, motoric	0.78
	Food	Other	0.78
	Hot	Perceptual	0.78

Table L1. Features with distinctiveness from 1.00 to 0.40 (inclusive) within categories (i.e. shared features)

Category	Feature	Feature type	Distinctiveness
Cut	Action	Visual, motoric	1.00
	Humans	Other	1.00
	Cut	Functional, motoric	0.75
	Fast	Visual, perceptual, motoric	0.75
	Object	Other	0.75
	Sharp	Visual, perceptual, motoric	0.75
	Arm	Visual, motoric	0.50
	Intentional	Other	0.50
	Knife	Other	0.50
	Move	Visual, motoric	0.50
	Separate	Functional	0.50
	Split	Visual, functional	0.50
	Tool	Other	0.50
	Wood	Other	0.50

Category	Feature	Feature type	Distinctiveness
Hit	Action	Visual, motoric	1.00
	Hit	Functional, motoric	0.78
	Contact	Functional, motoric	0.67
	Force	Functional, motoric	0.67
	Hand	Motoric	0.67
	Humans	Other	0.67
	Object	Other	0.56
	Anger	Other	0.44
	Hurt	Perceptual	0.44
	Loud	Perceptual	0.44
	Move	Visual, motoric	0.44
	Strike	Motoric	0.44
	Surface	Visual, perceptual	0.44
	Swing	Visual, motoric	0.44

Table L1. Features with distinctiveness from 1.00 to 0.40 (inclusive) within categories (i.e. shared features) cont.

Category	Feature	Feature type	Distinctiveness
Make	Action	Visual, motoric	1.00
	Humans	Other	0.83
	Make	Functional	0.67
	Hand	Motoric	0.58
	Tool	Other	0.58

Category	Feature	Feature type	Distinctiveness
Talk	Action	Visual, motoric	1.00
	Humans	Other	1.00
	Communicate	Functional	0.88
	Voice	Other	0.88
	Word	Other	0.88
	Mouth	Visual, motoric	0.63
	Noise	Perceptual	0.63
	Loud	Perceptual	0.50
	Speak	Functional, motoric	0.50
	Talk	Functional	0.50

Category	Feature	Feature type	Distinctiveness
Walk	Action	Visual, motoric	1.00
	Move	Visual, motoric	1.00
	Humans	Other	0.75
	Leg	Visual, motoric	0.75
	Walk	Functional, motoric	0.75
	Foot	Visual, functional, motoric	0.63
	Slow	Visual, perceptual, motoric	0.63

Table L1. Features with distinctiveness from 1.00 to 0.40 (inclusive) within categories (i.e. shared features) cont.

Appendix M Verb Semantic Feature Analysis – General/Specific and High-/Low-Typicality Analyses Stimuli

General	Specific			
	High-typicality	*M* typicality	Low-typicality	*M* typicality
Break	Drop	1.94	Bend	3.70
	Smash	1.72	Throw	3.56
Cook	Fry	1.26	Burn	4.01
	Grill	1.33	Mix	3.07
Cut	Chop	1.29	Snap	4.11
	Saw	1.92	Stab	3.97
Hit	Hammer	2.59	Crash	3.76
	Punch	1.21	Push	3.29
Make	Build	1.66	Fix	3.49
	Construct	1.60	Write	2.89
Talk	Chat	1.20	Scream	4.17
	Whisper	2.25	Yell	3.42
Walk	Step	2.59	Jog	3.38
	Wander	2.02	Limp	3.58

Table M1. Stimuli for General/Specific and High-/Low-typicality verb semantic feature analyses

Appendix N Category Verification – Verb Stimuli

Target	#letters	Prodfreq	Meanrank	Typicality	Lexfreq	Familiarity
Bending	7	0.37	6.38	3.7	1894	510
Burning	7	0.09	5.33	3.09	2004	548
Cracking	8	0.29	3.40	2.28	-	-
Crushing	8	0.43	4.67	2.41	497	480
Cutting	7	0.29	6.6	2.69	7309	581
Denting	7	0.09	10.67	3.79	80	480
Destroying	10	0.2	6.29	2.2	3154	551
Dropping	8	0.8	3.57	1.94	5786	577
Forcing	7	0.09	8.67	4.05	4807	552
Kicking	7	0.14	5.4	3.47	2293	564
Pulling	7	0.09	8	4.37	9336	565
Ripping	7	0.43	6.14	2.24	-	-
Slicing	7	0.17	8.33	3.75	298	540
Smashing	8	0.8	1.96	1.72	754	536
Snapping	8	0.57	4	1.47	2013	526
Splitting	9	0.09	7.67	3.18	1447	514
Tearing	7	0.4	6	2.51	1614	555
Throwing	8	0.4	5.29	3.56	7010	548

Table N1. Category verification (positive response items) and predictor variable values – Breaking

Target	#letters	Prodfreq	Meanrank	Typicality	Lexfreq	Familiarity
Bathing	7	0.14	8.60	1.93	316	599
Bleaching	9	0.20	5.86	2.4	53	549
Brushing	8	0.60	5.76	2.49	1193	579
Drying	6	0.11	7.25	4.24	1325	615
Dusting	7	0.49	5.47	1.59	319	558
Hoovering	9	0.57	6.00	1.4	-	-
Hosing	6	0.14	8.60	2.56	26	449
Polishing	9	0.60	4.00	2.06	258	485
Rinsing	7	0.31	5.55	2.13	165	480
Rubbing	7	0.40	6.29	3.04	-	-
Shining	7	0.17	6.00	3.06	840	558
Showering	9	0.11	9.75	2.07	213	593
Soaking	7	0.29	6.30	2.61	-	-
Spraying	8	0.26	7.33	3.17	310	521
Steaming	8	0.14	7.40	3.68	155	545
Sweeping	8	0.29	7.90	1.48	1415	495
Washing	7	0.86	3.97	1.17	2320	632
Wiping	6	0.63	4.00	1.17	-	-

Table N2. Category verification (positive response items) and predictor variable values – Cleaning

Target	#letters	Prodfreq	Meanrank	Typicality	Lexfreq	Familiarity
Assembling	10	0.14	3.80	1.56	674	482
Baking	6	0.66	3.30	1.73	231	549
Building	8	0.60	4.00	1.66	9223	544
Cooking	7	0.54	4.42	1.63	1655	568
Crafting	8	0.09	5.00	2.21	76	487
Designing	9	0.17	5.83	2.64	1957	538
Drawing	7	0.49	6.35	2.55	1162	542
Fixing	6	0.09	6.00	3.49	1551	573
Hammering	9	0.14	8.80	3.29	485	515
Joining	7	0.09	2.67	2.52	12071	544
Knitting	8	0.14	3.80	2.44	912	501
Painting	8	0.46	5.94	2.39	1631	551
Recording	9	0.09	5.00	3.14	4052	609
Sawing	6	0.11	7.75	3.64	150	552
Sewing	6	0.26	3.78	2.49	268	517
Sticking	8	0.26	4.00	2.79	3083	528
Typing	6	0.09	9.00	3.26	474	567
Writing	7	0.14	11.00	2.89	23173	560

Table N3. Category verification (positive response items) and predictor variable values – Making

Target	#letters	Prodfreq	Meanrank	Typicality	Lexfreq	Familiarity
Answering	9	0.11	2.25	2.05	6993	605
Arguing	7	0.20	3.71	2.68	9356	564
Chatting	8	0.40	2.79	1.2	-	-
Debating	8	0.11	9.50	2.49	668	459
Discussing	10	0.26	6.33	1.51	-	-
Gossiping	9	0.23	4.75	1.97	-	-
Lecturing	9	0.11	5.25	2.95	384	624
Mumbling	8	0.14	4.80	2.48	-	-
Questioning	11	0.11	5.50	2.23	2003	588
Saying	6	0.09	6.67	1.31	-	-
Screaming	9	0.20	5.14	4.17	1878	522
Shouting	8	0.74	2.85	3.04	4455	557
Signing	7	0.20	6.71	4.55	4360	543
Speaking	8	0.29	4.30	1.1	18311	600
Stating	7	0.09	6.33	2.55	5508	560
Stuttering	10	0.09	4.67	4.01	-	-
Whispering	10	0.74	2.96	2.25	2631	550
Yelling	7	0.23	6.63	3.42	1040	509

Table N4. Category verification (positive response items) and predictor variable values – Breaking

Appendix O Category Verification – Noun Stimuli

Target	#letters	Prodfreq	Meanrank	Typicality	Lexfreq	Familiarity
Budgie	6	0.23	7.88	1.977	133	1.42
Canary	6	0.20	11.71	1.953	103	1.26
Crow	4	0.46	6.38	1.256	377	1.32
Dove	4	0.17	8.50	1.477	309	1.55
Heron	5	0.20	11.86	2.326	84	2.00
Osprey	6	0.09	9.00	2.326	20	2.29
Owl	3	0.34	8.25	1.773	1621	1.16
Penguin	7	0.60	11.48	3.227	212	1.32
Pigeon	6	0.57	6.85	1.250	856	1.10
Raven	5	0.34	8.50	1.744	103	1.65
Robin	5	0.86	6.23	1.093	227	1.13
Seagull	7	0.51	10.83	1.364	134	1.23
Sparrow	7	0.66	6.04	1.047	204	1.03
Starling	8	0.29	7.30	1.182	182	1.48
Swan	4	0.37	11.69	2.000	582	1.19
Toucan	6	0.09	6.67	3.143	23	3.16
Turkey	6	0.20	17.14	2.360	887	1.26
Wren	4	0.17	12.00	1.465	67	1.61

Table O1. Category verification (positive response items) and predictor variable values – Birds

Target	#letters	Prodfreq	Meanrank	Typicality	Lexfreq	Familiarity
Belt	4	0.17	12.67	3.133	2553	1.30
Blouse	6	0.51	12.83	1.289	578	1.13
Cardigan	8	0.49	9.76	1.422	319	1.50
Coat	4	0.89	11.48	1.289	3696	1.07
Dress	5	0.54	10.95	1.000	3458	1.10
Dungarees	9	0.09	14.67	1.844	78	1.83
Gloves	6	0.40	13.29	2.844	1345	1.33
Hat	3	0.71	14.64	2.844	3697	1.47
Jeans	5	0.63	7.68	1.067	1233	1.00
Pants	5	0.29	8.45	1.822	556	1.23
Pyjamas	7	0.17	9.83	2.205	438	1.30
Shirt	5	0.91	7.47	1.044	3381	1.03
Shorts	6	0.66	9.96	2.000	883	1.23
Sock	4	0.89	9.68	1.600	1123	1.07
Stockings	9	0.09	22.00	2.044	725	1.37
Suit	4	0.29	14.60	1.267	2425	1.17
Trousers	8	0.91	5.63	1.022	2428	1.00
Waistcoat	9	0.37	16.62	2.333	310	1.60

Table O1. Category verification (positive response items) and predictor variable values – Clothes

Target	#letters	Prodfreq	Meanrank	Typicality	Lexfreq	Familiarity
Armchair	8	0.57	10.00	1.039	891	1.10
Bed	3	0.77	6.78	1.176	16664	1.03
Bench	5	0.31	9.09	2.235	2522	1.48
Bookcase	8	0.14	9.80	1.824	214	1.42
Cabinet	7	0.29	7.60	1.765	6759	1.77
Chair	5	0.97	1.59	1.000	8491	1.07
Chest	5	0.14	8.20	2.216	3745	1.74
Cupboard	8	0.51	7.72	1.647	1840	1.26
Desk	4	0.77	9.26	1.529	4414	1.32
Dresser	7	0.20	10.29	1.510	291	1.84
Settee	6	0.11	6.75	1.098	341	1.52
Shelf	5	0.40	8.64	2.627	2530	1.36
Sideboard	9	0.26	7.78	1.569	215	1.68
Sink	4	0.09	14.67	3.588	887	1.65
Sofa	4	0.83	4.97	1.098	1044	1.32
Stool	5	0.49	8.35	1.706	1087	1.36
Table	5	0.91	3.31	1.039	21594	1.03
Wardrobe	8	0.77	8.22	1.216	1072	1.26

Table O1. Category verification (positive response items) and predictor variable values – Furniture

Target	#letters	Prodfreq	Meanrank	Typicality	Lexfreq	Familiarity
Apple	5	1.00	1.57	1.023	3444	1.06
Blackcurrant	12	0.29	17.20	1.881	127	1.59
Blueberry	9	0.57	14.90	2.814	17	2.84
Coconut	7	0.14	12.20	3.581	364	1.69
Cranberry	9	0.14	16.40	2.814	94	2.78
Date	4	0.11	12.50	2.929	12742	1.63
Fig	3	0.11	16.25	2.837	1930	2.03
Gooseberry	10	0.46	14.88	2.047	104	1.69
Grapefruit	10	0.37	12.69	1.256	5	1.28
Lemon	5	0.49	12.12	1.512	1334	1.19
Lime	4	0.37	12.38	2.093	619	1.91
Mango	5	0.51	12.06	2.791	98	2.75
Orange	6	0.94	4.97	1.023	1148	1.03
Peach	5	0.66	10.70	1.419	512	1.47
Pear	4	0.97	6.85	1.163	445	1.19
Plum	4	0.60	11.71	1.302	399	1.28
Raspberry	9	0.69	12.54	1.651	273	1.44
Satsuma	7	0.31	7.73	1.643	24	2.09

Table O1. Category verification (positive response items) and predictor variable values – Fruit

Appendix P Category Verification – (Individual) Regression Model Statistics for Error Production

	# letters		Prod' freq		Mean rank		Typicality		Lexical freq		Familiarity	
	Wald	p =	Wald	p =	Wald	p =	Wald	p =	Wald	p =	Wald	p =
1	0.062	.803	0.077	.781	0.161	.688	0.225	.635	0.248	.618	1.648	.199
2	0.344	.557	0.159	.690	0.010	.921	4.917	.027*	0.893	.345	0.227	.634
3	0.214	.644	0.803	.370	0.004	.947	5.792	.016*	0.142	.706	1.027	.311
4	0.992	.319	5.916	.015*	0.017	.895	10.173	.001**	0.696	.404	4.898	.027*
5	0.036	.849	1.993	.158	0.300	.584	0.734	.391	0.015	.904	0.149	.699
6	1.225	.268	0.988	.320	0.003	.958	1.087	.297	0.091	.763	0.201	.654
7	0.877	.349	0.352	.553	0.771	.380	7.318	.007**	2.036	.154	0.760	.383
8	1.342	.247	0.511	.475	0.511	.475	6.058	.014*	0.546	.460	0.435	.510
9	1.244	.265	0.847	.357	0.323	.570	2.964	.085	0.000	.987	0.616	.432
10	0.354	.552	0.425	.514	2.781	.095	1.007	.316	0.057	.812	1.250	.264

Table P1. Logistic regression predictor variable statistics (enter method) for individual errors in verb category verification (without dummy variables)

	# errors	Omnibus tests		R^2 values		H & L test	% correctly predicted		
	(n / 60)	χ^2	p =	Min	Max	p =	Correct	Errors	Total
1	7	2.17	.903	.036	.069	.751	100	0.0	88.3
2	8	14.60	.024	.216	.397	.991	98.1	25.0	88.3
3	13	16.99	.009	.247	.380	.142	93.6	38.5	91.7
4	23	50.74	<.001	.571	.776	.890	89.2	91.3	90.0
5	12	5.69	.459	.090	.143	.901	100	0.0	80.0
6	14	8.14	.228	.127	.191	.395	95.7	7.1	75.0
7	8	18.96	.004	.271	.498	.732	98.1	37.5	90.0
8	10	15.69	.016	.230	.387	.610	98.0	20.0	85.0
9	3	10.08	.121	.155	.472	1.000	98.2	0.0	93.3
10	13	14.01	.030	.208	.321	.539	93.6	38.5	81.7

Table P2. Logistic regression model statistics (enter method) for individual errors in verb category verification (without dummy variables)

	# letters		Prod' freq		Mean rank		Typicality		Lexical freq		Familiarity	
	Wald	p =	Wald	p =	Wald	p =	Wald	p =	Wald	p =	Wald	p =
1	0.510	.475	0.612	.434	0.512	.474	0.490	.484	0.551	.458	2.201	.138
2	0.009	.925	0.154	.694	0.100	.752	3.853	.050	0.623	.430	0.001	.978
3	0.084	.772	0.675	.411	0.010	.921	3.870	.049	0.427	.513	0.770	.380
4	0.439	.508	6.118	.013	0.013	.911	9.204	.002	0.471	.493	3.819	.051
5	0.099	.753	1.418	.234	0.140	.708	0.609	.435	0.183	.669	0.918	.338
6	0.303	.582	0.106	.744	0.187	.666	2.497	.114	0.293	.588	0.286	.593
7	1.456	.228	1.056	.304	0.562	.454	5.609	.018	2.434	.119	1.253	.263
8	2.758	.097	0.324	.569	2.070	.150	5.467	.019	1.249	.264	0.224	.636
9	0.000	.996	0.933	.334	0.153	.696	1.223	.269	0.666	.414	0.155	.694
10	0.089	.766	1.206	.272	0.649	.421	0.559	.455	0.056	.814	0.918	.338

Table P3. Logistic regression predictor variable statistics (enter method) for individual errors in verb category verification (with dummy variables)

	# errors	Omnibus tests		R^2 values		H & L test	% correctly predicted		
	(n / 60)	χ^2	p =	Min	Max	p =	Correct	Errors	Total
1	7	6.705	.668	.106	.206	.419	100.0	14.3	90.0
2	8	20.787	.014	.293	.5387	.665	98.1	62.5	93.3
3	13	23.727	.005	.327	.504	.676	95.7	69.2	90.0
4	23	57.468	<.001	.616	.837	.973	91.9	91.3	91.7
5	12	11.318	.255	.172	.272	.127	100.0	33.3	86.7
6	14	16.596	.055	.242	.365	.787	95.7	28.6	80.0
7	8	21.123	.012	.297	.545	.980	98.1	37.5	90.0
8	10	21.377	.011	.300	.505	.513	92.0	40.0	83.3
9	3	17.442	.042	.252	.770	.999	100.0	66.7	98.3
10	13	21.347	.011	.299	.462	.394	91.5	46.2	81.7

Table P4. Logistic regression model statistics (enter method) for individual errors in verb category verification (without dummy variables)

	# letters		Prod' freq		Mean rank		Typicality		Lexical freq		Familiarity	
	Wald	p =	Wald	p =	Wald	p =	Wald	p =	Wald	p =	Wald	p =
1											1.235	.266
2							8.372	.004				
3							10.098	.001				
4			5.462	.019			11.876	.001			5.871	.015
5			2.922	.087								
6			2.937	.087								
7							8.524	.004				
8							8.925	.003				
9							4.010	.045				
10					7.441	.006						

Table P5. Logistic regression predictor variable statistics (backwards stepwise method) for individual errors in verb category verification (without dummy variables)

	Model	Omnibus tests		R^2 values		H & L test	% correctly predicted		
		χ^2	p =	Min	Max	p =	Correct	Errors	Total
1	6 / 7	1.309	.253	.022	.042	.725	100.0	0.0	88.3
2	6 / 6	12.359	<.001	.186	.342	.099	94.2	12.5	83.3
3	6 / 6	14.043	<.001	.209	.322	.112	93.6	30.8	80.0
4	4 / 4	49.011	<.001	.558	.759	.906	89.2	73.9	83.3
5	6 / 6	4.359	.037	.070	.111	.623	100.0	0.0	80.0
6	6 / 6	4.023	.045	.065	.098	.809	100.0	0.0	76.7
7	6 / 6	12.820	<.001	.192	.354	.646	96.2	25.0	86.7
8	6 / 6	12.460	<.001	.188	.316	.751	94.0	20.0	81.7
9	6 / 6	5.905	.015	.094	.286	.795	100.0	0.0	95.0
10	6 / 6	8.987	.003	.139	.215	.366	93.6	7.7	75.0

Table P6. Logistic regression model statistics (backwards stepwise method) for individual errors in verb category verification (without dummy variables)

	# letters		Prod’ freq		Mean rank		Typicality		Lexical freq		Familiarity	
	Wald	p =	Wald	p =	Wald	p =	Wald	p =	Wald	p =	Wald	p =
1											1.805	.179
2							8.327	.004				
3							9.024	.003				
4			6.369	.012			12.553	<.001			4.686	.030
5			2.922	.087								
6							6.630	.010				
7							8.524	.004				
8							8.925	.003				
9							4.010	.045				
10			4.840	.028								

Table P7. Logistic regression predictor variable statistics (backwards stepwise method) for individual errors in verb category verification (with dummy variables)

	Model	Omnibus tests		R^2 values		H & L test	% correctly predicted		
		χ^2	p =	Min	Max	p =	Correct	Errors	Total
1	8 / 10	4.542	.103	.073	.142	.806	100.0	0.0	88.3
2	8 / 8	18.943	<.0014	.271	.498	.925	96.2	62.5	91.7
3	7 / 7	20.621	<.001	.291	.449	.234	95.7	61.5	88.3
4	6 / 6	55.67	<.001	.601	.817	.798	91.9	87.0	90.0
5	9 / 9	4.359	.037	.070	.111	.623	100.0	0.0	80.0
6	8 / 8	14.180	.001	.210	.318	.951	95.7	42.9	83.3
7	9 / 9	12.820	<.001	.192	.354	.646	96.2	25.0	86.7
8	9 / 9	12.460	<.001	.188	.316	.751	94.0	20.0	81.7
9	9 / 9	5.905	.015	.094	.286	.795	100.0	0.0	95.0
10	7 / 7	14.305	.003	.212	.327	.713	95.7	53.8	86.7

Table P8. Logistic regression model statistics (backwards stepwise method) for individual errors in verb category verification (without dummy variables)

Appendix Q Category Verification – (Individual and Group) Regression Model Statistics for Response Time Analysis

	Model reported	*F*-statistic	Adjusted R^2
Group mean	6 / 6	*F* (1,59) = 5.16, *p* = .027	.066
1	5 / 5	*F* (2,59) = 4.86, *p* = .011	.116
2	6 / 7	*F* (1,59) = 2.39, *p* = .127	.023
3	6 / 7	*F* (1,59) = 0.63, *p* = .429	.006
4	6 / 7	*F* (1,59) = 1.51, *p* = .225	.009
5	6 / 7	*F* (1,59) = 2.69, *p* = .106	.028
6	6 / 7	*F* (1,59) = 2.58, *p* = .114	.026
7	6 / 7	*F* (1,59) = 1.83, *p* = .182	.014
8	6 / 6	*F* (1,59) = 8.39, *p* = .005	.111
9	5 / 5	*F* (2,59) = 4.11, *p* = .022	.095
10	6 / 7	*F* (1,59) = 2.50, *p* = .119	.025

Table Q1. Regression model statistics for group and individual response times in verb category verification (across categories; without dummy variables)

	Model reported	*F*-statistic	Adjusted R^2
Group mean	6 /6	*F* (4,59) = 7.72, *p* < .001	.313
1	8 / 8	*F* (2,59) = 4.86, *p* = .011	.116
2	7 / 7	*F* (3,59) = 4.07, *p* = .011	.135
3	7 / 9	*F* (3,59) = 2.70, *p* = .054	.080
4	6 / 7	*F* (4,59) = 3.94, *p* = .007	.166
5	9 / 10	*F* (1,59) = 2.69, *p* = .106	.028
6	8 / 8	*F* (2,59) = 7.28, *p* = .002	.175
7	8 / 9	*F* (2,59) = 3.02, *p* = .057	.064
8	6 / 6	*F* (4,59) = 5.21, *p* = .001	.222
9	7 / 7	*F* (3,59) = 4.70, *p* = .005	.158
10	9 / 10	*F* (1,59) = 2.50, *p* = .119	.025

Table Q2. Regression model statistics for group and individual response times in verb category verification (across categories; with dummy variables)

	Model reported	*F*-statistic	Adjusted R^2
Group mean	5 / 5	*F* (2,71) = 14.27, *p* < .001	.272
1	6 / 6	*F* (1,71) = 9.26, *p* = .003	.104
2	6 / 7	*F* (1,71) = 1.13, *p* = .291	.002
3	6 / 6	*F* (1,71) = 12.14, *p* = .001	.136
4	5 / 5	*F* (2,71) = 7.81, *p* < .001	.178
5	6 / 6	*F* (1,71) = 3.31, *p* = .073	.031
6	5 / 5	*F* (2,71) = 10.23, *p* < .001	.206
7	6 / 6	*F* (1,71) = 9.46, *p* = .003	.106
8	5 / 5	*F* (2,71) = 2.64, *p* = .078	.044
9	5 / 5	*F* (2,71) = 2.31, *p* = .107	.036
10	5/ 5	*F* (2,71) = 9.39, *p* < .001	.191

Table Q3. Regression model statistics for group and individual response times in noun category verification (across categories; without dummy variables)

	Model reported	*F*-statistic	Adjusted R^2
Group mean	8 / 8	*F* (2,71) = 15.67, *p* < .001	.293
1	9 / 9	*F* (1,71) = 9.26, *p* = .003	.104
2	8 / 9	*F* (2,71) = 6.48, *p* = .003	.134
3	9 / 9	*F* (1,71) = 12.14, *p* = .001	.136
4	5 / 5	*F* (5,71) = 8.15, *p* < .001	.335
5	9 / 9	*F* (1,71) = 3.31, *p* = .073	.031
6	7 / 7	*F* (3,71) = 8.32, *p* < .001	.236
7	8 / 8	*F* (2,71) = 6.26, *p* = .003	.129
8	6 / 9	*F* (4,71) = 2.24, *p* = .074	.065
9	8 / 8	*F* (2,71) = 2.31, *p* = .107	.036
10	8 / 8	*F* (2,71) = 14.33, *p* < .001	.273

Table Q4. Regression model statistics for group and individual response times in noun category verification (across categories; with dummy variables)

	Model reported	F-statistic	Adjusted R^2
breaking	6 / 6	F (1,15) = 3.99, p = .065	.166
cleaning	6 / 6	F (1,13) = 4.03, p = .068	.189
making	5 / 5	F (2,17) = 5.18, p = .020	.329
talking	4 / 4	F (3,11) = 6.99, p = .013	.620

Table Q5. Regression model statistics (backwards stepwise method) for group response times in verb category verification (within categories)

	Model reported	F-statistic	Adjusted R^2
birds	6 / 6	F (1,17) = 12.15, p = .003	.396
clothes	6 / 7	F (1,17) = 1.24, p = .282	.014
fruit	6 / 6	F (1,17) = 7.75, p = .013	.284
furniture	6 / 6	F (1,17) = 6.48, p = .022	.244

Table Q6. Regression model statistics (backwards stepwise method) for group response times in noun category verification (within categories)

Appendix R Semantically Primed Picture Naming – Coordinate and Superordinate Prime Stimuli

Verbs			Nouns		
Target	Related prime	Unrelated prime	Target	Related prime	Unrelated prime
Wash	Mop	Draw	Shirt	Dress	Banana
Brush	Polish	Paint	Trousers	Coat	Pear
Polish	Brush	Sew	Coat	Trousers	Strawberry
Mop	Wash	Build	Dress	Shirt	Apple
Drop	Smash	Boil	Apple	Banana	Dress
Smash	Drop	Fry	Banana	Apple	Shirt
Snap	Crush	Roast	Pear	Strawberry	Trousers
Crush	Snap	Bake	Strawberry	Pear	Coat
Build	Sew	Brush	Chair	Bed	Saw
Draw	Paint	Polish	Table	Sofa	Drill
Paint	Draw	Mop	Sofa	Table	Screwdriver
Sew	Build	Wash	Bed	Chair	Hammer
Punch	Hammer	Chop	Hammer	Saw	Bed
Slap	Kick	Saw	Screwdriver	Drill	Sofa
Kick	Slap	Tear	Drill	Screwdriver	Table
Hammer	Punch	Slice	Saw	Hammer	Chair
Slice	Tear	Punch	Bus	Train	Pepper
Chop	Saw	Kick	Car	Plane	Onion
Saw	Chop	Hammer	Plane	Car	Potato
Tear	Slice	Slap	Train	Bus	Carrot
Fry	Roast	Crush	Carrot	Pepper	Train
Boil	Fry	Drop	Potato	Carrot	Bus
Bake	Boil	Smash	Onion	Potato	Car
Roast	Bake	Snap	Pepper	Onion	Plane

Table F1.Verb and noun stimuli for semantically primed picture naming (coordinate condition)

Verbs			Nouns		
Target	Related prime	Unrelated prime	Target	Related prime	Unrelated prime
Wash	Clean	Put	Shirt	Clothes	Bird
Brush	Clean	Put	Trousers	Clothes	Bird
Polish	Clean	Move	Coat	Clothes	Sport
Mop	Clean	Move	Dress	Clothes	Sport
Drop	Break	Talk	Apple	Fruit	Fish
Smash	Break	Talk	Banana	Fruit	Fish
Snap	Break	Run	Pear	Fruit	Weapon
Crush	Break	Run	Strawberry	Fruit	Weapon
Build	Make	Walk	Chair	Furniture	Flower
Draw	Make	Walk	Table	Furniture	Flower
Paint	Make	Jump	Sofa	Furniture	Music
Sew	Make	Jump	Bed	Furniture	Music
Punch	Hit	Jump	Hammer	Tool	Music
Slap	Hit	Jump	Screwdriver	Tool	Music
Kick	Hit	Walk	Drill	Tool	Flower
Hammer	Hit	Walk	Saw	Tool	Flower
Slice	Cut	Run	Bus	Transport	Weapon
Chop	Cut	Run	Car	Transport	Weapon
Saw	Cut	Talk	Plane	Transport	Fish
Tear	Cut	Talk	Train	Transport	Fish
Fry	Cook	Move	Carrot	Vegetable	Sport
Boil	Cook	Move	Potato	Vegetable	Sport
Bake	Cook	Put	Onion	Vegetable	Bird
Roast	Cook	Put	Pepper	Vegetable	Bird

Table F2.Verb and noun stimuli for semantically primed picture naming (superordinate condition)

Appendix S Intervention Study – Verb Treatment Stimuli

Target	Category	#Syllables	Lexical frequency
Bake	Cook	1	231
Barbeque	Cook	3	0
Beat	Hit	1	4960
Bend	Break	1	1894
Blow	Other	1	3091
Boil	Cook	2	555
Bounce	Other	1	872
Brush	Clean	1	1193
Build	Make	1	9223
Bury	Other	2	724
Carry	Other	2	20583
Catch	Other	1	7760
Chew	Other	1	634
Chop	Cut	1	412
Count	Other	1	2884
Crack	Break	1	777
Crash	Hit	1	1575
Crush	Break	1	497
Deliver	Other	3	3668
Dig	Other	1	1676
Draw	Make	1	11602
Drink	Other	1	4651
Drop	Break	1	5786
Dry	Clean	1	1325
Dust	Clean	1	319
Eat	Other	1	9158
Fasten	Other	2	338
Feed	Other	1	3438
Flush	Other	1	524
Fold	Make	1	547
Fry	Cook	1	87

Table S1. Verb stimuli used as treatment items across the five participants

Target	Category	#Syllables	Lexical frequency
Glue (Stick)	Make	1	111 (3083)
Grill	Cook	1	88
Hammer	Hit	2	485
Hang	Other	1	4438
Hoover	Clean	2	66
Hug	Other	1	639
Juggle	Other	2	190
Kick	Hit	1	2293
Kiss	Other	1	2561
Knit	Make	1	912
Knock	Hit	1	2656
Lick	Other	1	712
Lift	Other	1	4264
Light	Other	1	1132
Measure	Other	2	2807
Microwave	Cook	3	19
Mix	Make	1	1211
Mop	Clean	1	251
Order	Other	2	3917
Pack	Other	1	1173
Paint	Make	1	1633
Pick	Other	1	10129
Play	Other	1	27001
Poach	Cook	1	141
Polish	Clean	2	258
Post	Other	1	316
Pour	Other	1	1997
Punch	Hit	1	591
Push	Hit	1	6985
Read	Other	1	13288
Ride	Other	1	3936

Table S1. Verb stimuli used as treatment items across the five participants (continued)

Target	Category	#Syllables	Lexical frequency
Rip (Tear)	Break / Cut	1	653 (4960)
Roast	Cook	1	86
Saw	Cut	1	150
Scrub	Clean	1	253
Sew	Make	1	268
Sharpen	Other	2	304
Slap	Hit	1	619
Smash	Break	1	754
Smell	Other	1	1532
Snap	Break	1	2013
Spray	Clean	1	310
Spread	Other	1	2445
Stamp	Break	1	556
Sweep	Clean	1	1415
Teach	Other	1	5237
Tie	Other	1	1476
Toast	Cook	1	154
Toss	Other	1	837
Wash	Clean	1	2320
Watch	Other	1	15121
Water	Other	2	255
Weigh	Other	1	1485
Wipe	Clean	1	1608
Write	Make	1	23173

Note.: Items in parentheses indicate picture stimuli used with different targets between participants

Table S1. Verb stimuli used as treatment items across the five participants (continued)

Appendix T Intervention Study – Noun Treatment Stimuli

Target	Category	#Syllables	Lexical frequency
Anchor	Other	2	471
Apple	Fruit	2	3444
Ashtray	Other	2	287
Asparagus	Vegetable	4	100
Axe	Tool	1	866
Ball	Other	1	8542
Banana	Fruit	3	968
Barn	Other	1	1514
Basket	Other	2	1680
Bath	Furniture	1	2989
Beans	Vegetable	1	1736
Bed	Furniture	1	16664
Belt	Clothing	1	2553
Bench	Furniture	1	2522
Book	Other	1	36284
Boots	Clothing	1	3489
Bottle	Other	2	5441
Bowl	Other	1	2693
Box	Other	1	10285
Broccoli	Vegetable	3	132
Button	Other	2	2136
Camel	Animal	2	499
Candle	Other	2	1524
Cannon	Other	2	114
Car	Transport	1	33944
Carrot	Vegetable	2	854
Cat	Animal	1	5385
Cauliflower	Vegetable	4	78
Celery	Vegetable	3	197
Chair	Furniture	1	8491
Cherry	Fruit	2	695

Table T1. Noun stimuli used as treatment items across the five participants

Target	Category	#Syllables	Lexical frequency
Chisel	Tool	2	127
Church	Other	1	23453
Cloud	Other	1	3398
Coat	Clothing	1	3696
Cow	Animal	1	2508
Desk	Furniture	1	4414
Dog	Animal	1	12015
Donkey	Animal	2	615
Drawers	Furniture	1	1499
Dress	Clothing	1	3458
Drill	Tool	1	777
Elephant	Animal	3	1455
Envelope	Other	3	1688
Eye	Other	1	36162
Fence	Other	1	2089
Finger	Other	2	8407
Flag	Other	1	1934
Flower	Other	2	6879
Fork	Tool	1	991
Giraffe	Animal	2	89
Glasses	Other	2	2334
Gorilla	Animal	3	210
Grapes	Fruit	1	795
Hammer	Tool	2	1034
Hanger	Other	2	254
Hat	Clothing	1	3697
Helicopter	Transport	4	1531
Horse	Animal	1	12167
Kangaroo	Animal	3	160
Kettle	Other	2	894
Key	Other	1	4254

Table T1. Noun stimuli used as treatment items across the five participants (continued)

Target	Category	#Syllables	Lexical frequency
Kite	Other	1	935
Knife	Tool	1	3209
Ladder	Other	2	1492
Lamp	Furniture	1	1998
Leg	Other	1	11108
Lemon	Fruit	2	1334
Lettuce	Vegetable	2	432
Lion	Animal	2	2081
Monkey	Animal	2	1051
Moon	Other	1	2735
Motorbike	Transport	3	374
Mountain	Other	2	6031
Mushroom	Vegetable	2	783
Nail	Tool	2	1813
Necklace	Other	2	379
Onion	Vegetable	2	1211
Orange	Fruit	2	1148
Peach	Fruit	1	512
Pear	Fruit	1	445
Peas	Vegetable	1	776
Pepper	Vegetable	2	1082
Pineapple	Fruit	3	326
Plane	Transport	1	4396
Pliers	Tool	2	69
Potato	Vegetable	3	2458
Pumpkin	Vegetable	2	123
Rabbit	Animal	2	2393
Rocking chair	Furniture	3	0
Ruler	Tool	2	1561
Sandwich	Other	2	1635
Saw	Tool	1	625
Scarf	Clothing	1	685

Table T1. Noun stimuli used as treatment items across the five participants (continued)

Target	Category	#Syllables	Lexical frequency
Scissors	Tool	2	425
Screwdriver	Tool	3	269
Sheep	Animal	1	2971
Shirt	Clothing	1	3381
Shoes	Clothing	1	4452
Sink	Furniture	1	887
Skateboard	Transport	2	60
Skirt	Clothing	1	1757
Sofa	Furniture	2	1044
Squirrel	Animal	2	378
Strawberry	Fruit	3	612
Submarine	Transport	3	740
Suitcase	Other	2	732
Sweetcorn	Vegetable	2	29
Swing	Other	1	1157
Table	Furniture	2	21594
Telephone	Other	3	5529
Television	Other	4	9917
Tiger	Animal	2	1302
Tomato	Fruit / vegetable	3	1461
Train	Transport	1	6547
Trousers	Clothing	2	2428
Tweezers	Tool	2	59
Umbrella	Other	3	914
Wardrobe	Furniture	2	1072
Watch	Other	1	2205
Windmill	Other	2	246
Window	Other	2	18299
Wrench	Tool	1	68
Yacht	Transport	1	1403
Zebra	Animal	2	225

Table T1. Noun stimuli used as treatment items across the five participants (continued)

Appendix U Intervention Study – Frequency matching Statistics for Treatment Stimuli by Participant

	AB		GF		JA		RH		SH	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Verbs										
Total	3063.38	5331.72	3139.3	5261.23	2704.53	4089.44	2964.08	5151.00	3338.37	5641.73
Treated	2474.75	5129.2	2762.7	5559.06	3223.6	5479.64	2656.45	5507.83	1484.75	2046.07
Untreated (Rel)	1732.3	2798.63	1751.85	2357.05	1311.45	2264.72	1954.15	2522.63	2619.8	5237.21
Untreated (Un)	4983.1	6929.9	4903.35	6636.65	3578.55	3693.90	4281.65	6533.83	5910.55	7489.88
Nouns										
Total	4378	7528.13	2023.32	2148.11	3253.05	5264.56	1964.78	2841.64	3380.45	6409.57
Treated	2665.45	4217.36	1333.65	1317.27	3534.75	5994.52	1344.2	1178.00	3624.95	5853.00
Untreated (Rel)	2780.95	4833.12	1597.5	1259.35	1228.6	1995.91	1561.9	1229.99	2651.55	7417.44
Untreated (Un)	7687.6	10841.43	3138.8	3001.06	4995.8	6209.01	2988.25	4527.16	3864.85	6117.76

Table U1. Descriptive data for lexical frequency of treatment sets between participants

Participant	Word class	Sets	t =	p =
AB	Verbs	Treated vs. Un-Rel	0.568	.573
		Treated vs. Un-Un	-1.301	.201
		Un-Rel vs. Un-Un	-1.945	.059
	Nouns	Treated vs. Un-Rel	-0.081	.936
		Treated vs. Un-Un	-1.931	.061
		Un-Rel vs. Un-Un	-1.849	.072
GF	Verbs	Treated vs. Un-Rel	0.749	.459
		Treated vs. Un-Un	-1.106	.276
		Un-Rel vs. Un-Un	-2.001	.053
	Nouns	Treated vs. Un-Rel	-0.647	.521
		Treated vs. Un-Un	-2.463	.018*
		Un-Rel vs. Un-Un	-2.118	.041*
JA	Verbs	Treated vs. Un-Rel	1.442	.157
		Treated vs. Un-Un	-0.240	.811
		Un-Rel vs. Un-Un	-2.340	.025*
	Nouns	Treated vs. Un-Rel	1.632	.111
		Treated vs. Un-Un	-0.757	.454
		Un-Rel vs. Un-Un	-2.583	.014*
RH	Verbs	Treated vs. Un-Rel	0.518	.607
		Treated vs. Un-Un	-0.851	.400
		Un-Rel vs. Un-Un	-1.486	.145
	Nouns	Treated vs. Un-Rel	-0.572	.571
		Treated vs. Un-Un	-1.572	.124
		Un-Rel vs. Un-Un	-1.360	.182
SH	Verbs	Treated vs. Un-Rel	-0.903	.372
		Treated vs. Un-Un	-2.549	.015*
		Un-Rel vs. Un-Un	-1.610	.116
	Nouns	Treated vs. Un-Rel	0.461	.648
		Treated vs. Un-Un	-0.127	.900
		Un-Rel vs. Un-Un	-0.564	.576

Table U2. Lexical frequency comparisons between treatment sets (independent t-test)

Appendix V Intervention Study – Example SFA Worksheets

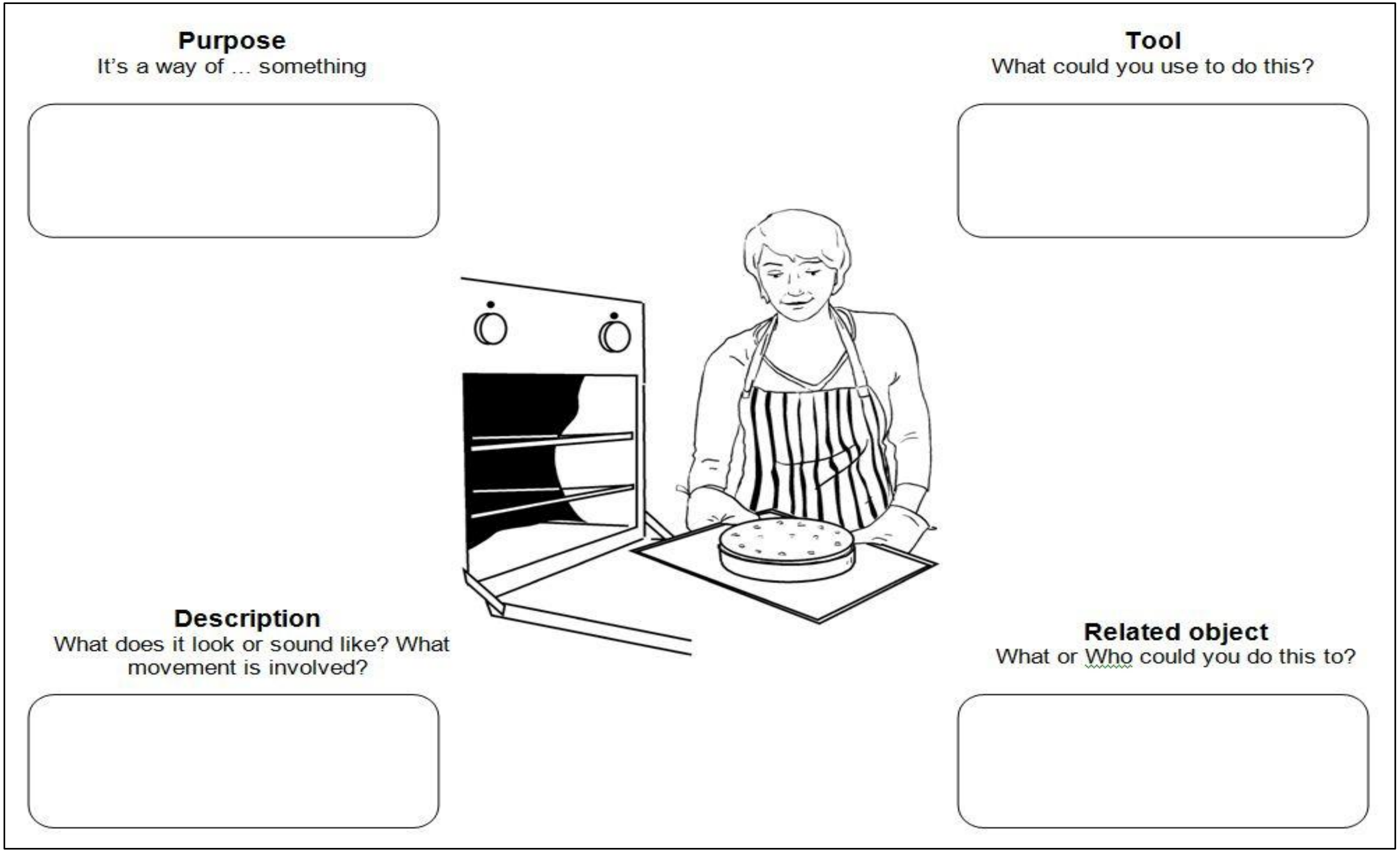

Figure V1. Example of verb-SFA worksheet (baking)

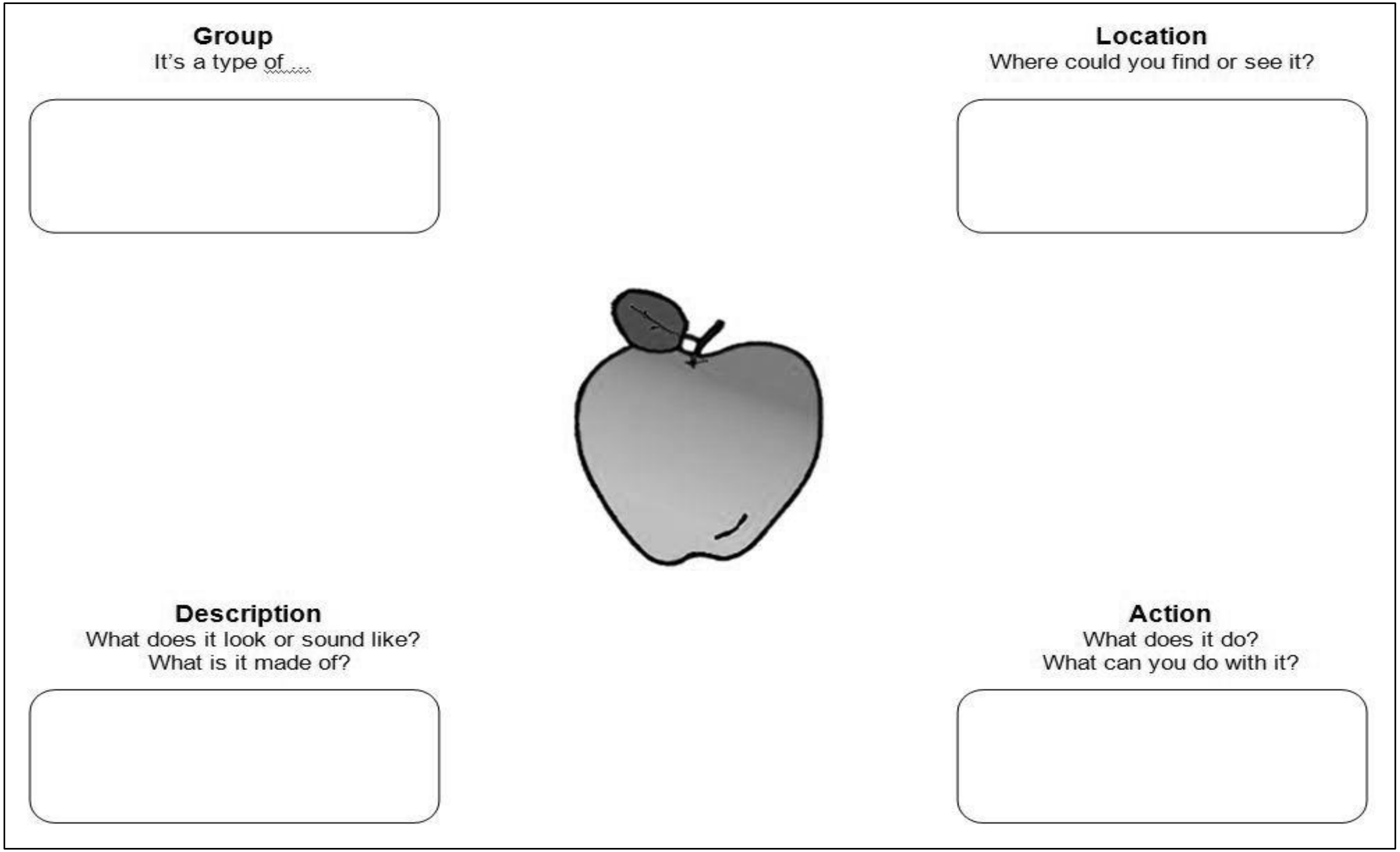

Figure V2. Example of noun-SFA worksheet (apple)

Appendix W Intervention Study – Primary outcome measure data

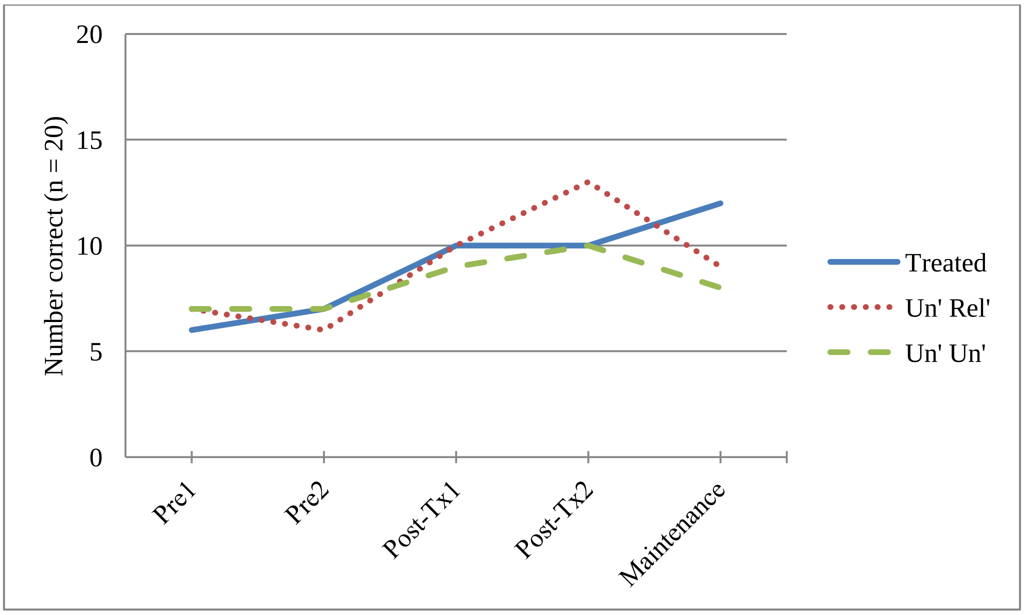

Table W1 Participant AB performance in noun picture naming (by item set; Tx1 – verb-SFA; Tx2 – noun-SFA)

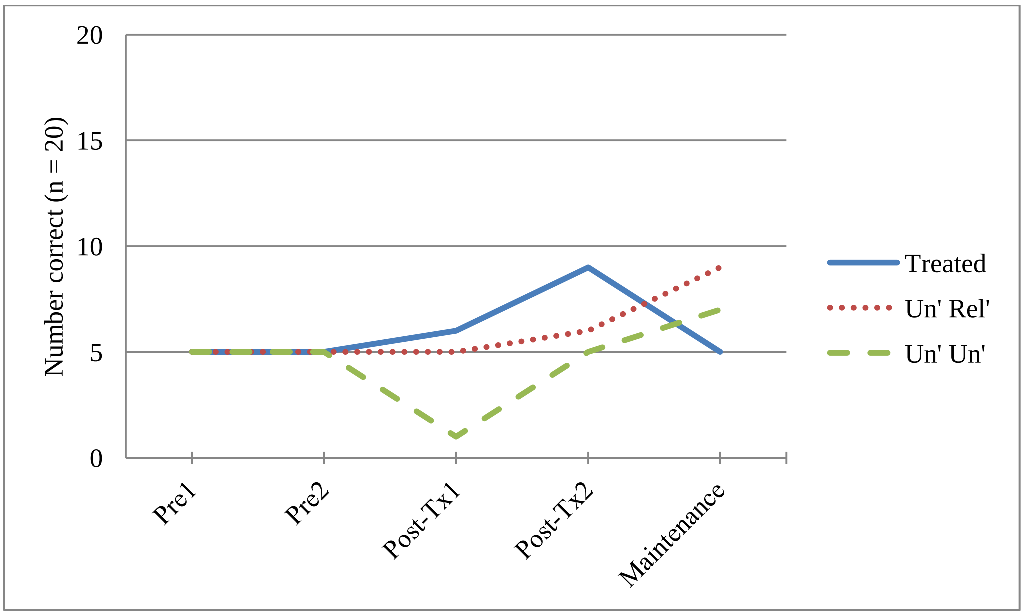

Table W2 Participant AB performance in verb picture naming (by item set; Tx1 – verb-SFA; Tx2 – noun-SFA)

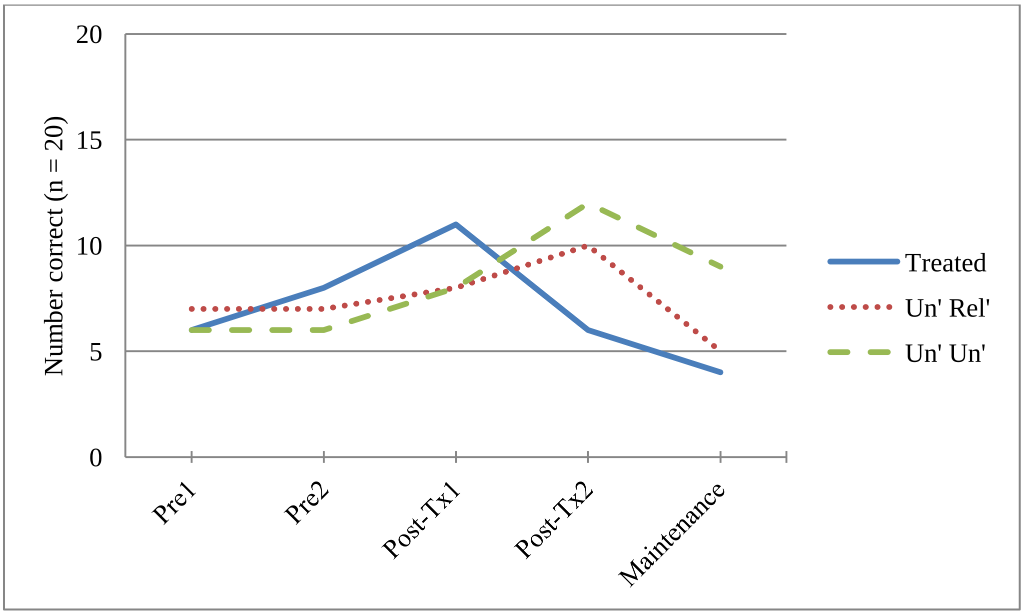

Table W3 Participant GF performance in noun picture naming (by item set; Tx1 – noun-SFA; Tx2 – verb-SFA)

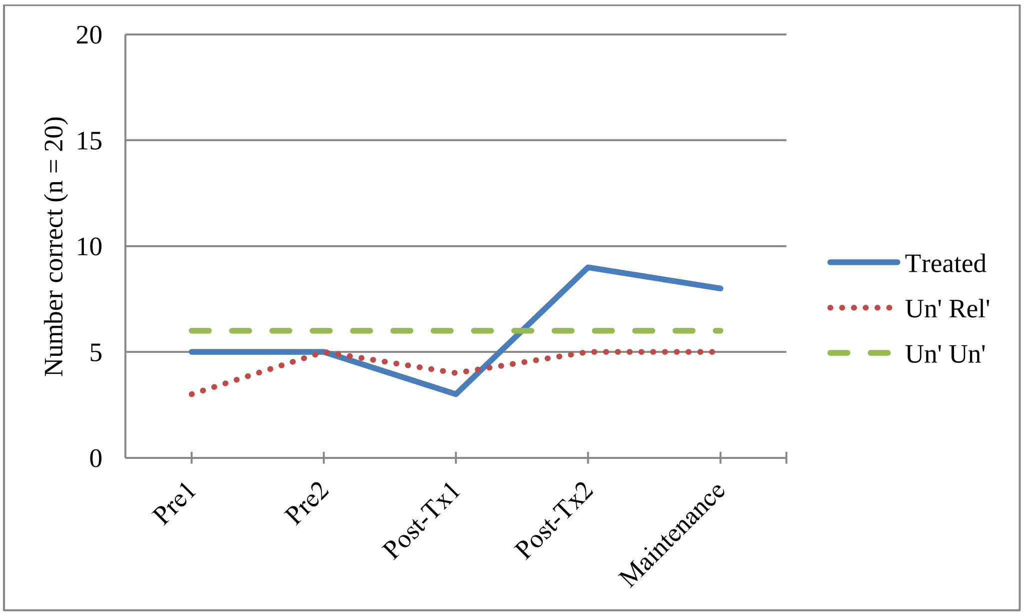

Table W4 Participant GF performance in verb picture naming (by item set; Tx1 – noun-SFA; Tx2 – verb-SFA)

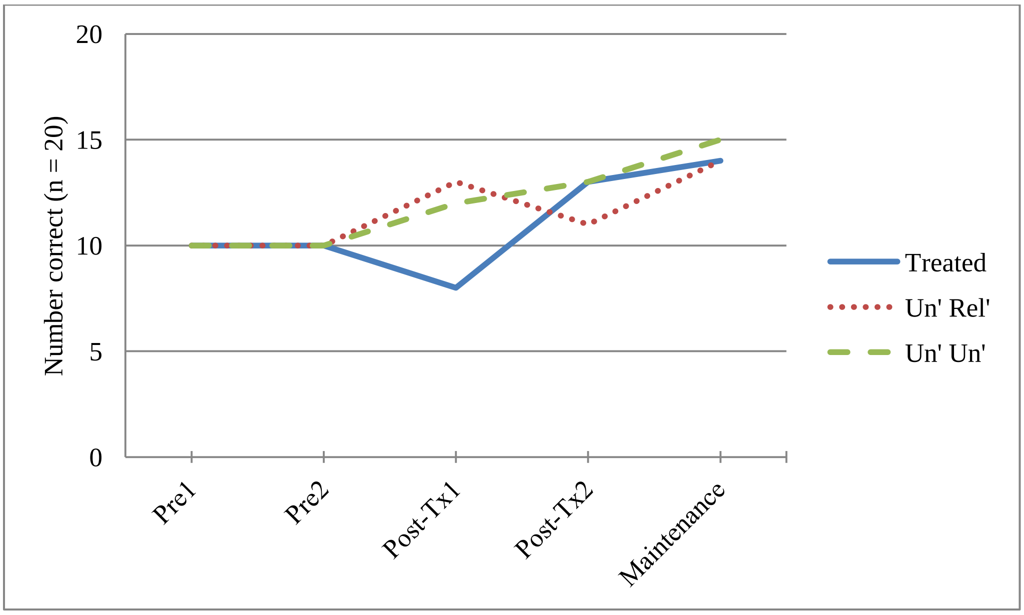

Table W5 Participant JA performance in noun picture naming (by item set; Tx1 – verb-SFA; Tx2 – noun-SFA)

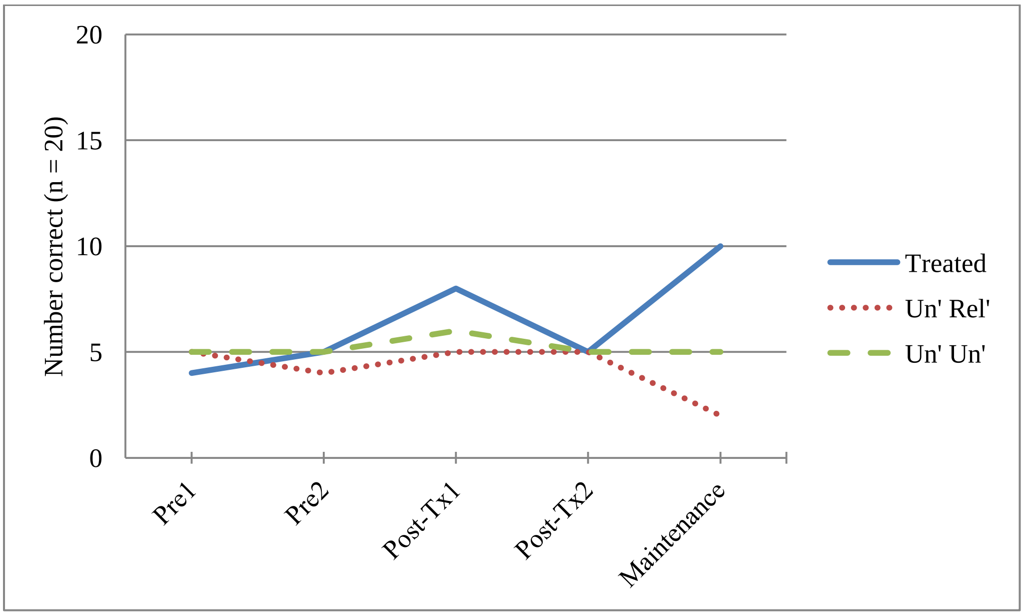

Table W6 Participant JA performance in verb picture naming (by item set; Tx1 – verb-SFA; Tx2 – noun-SFA)

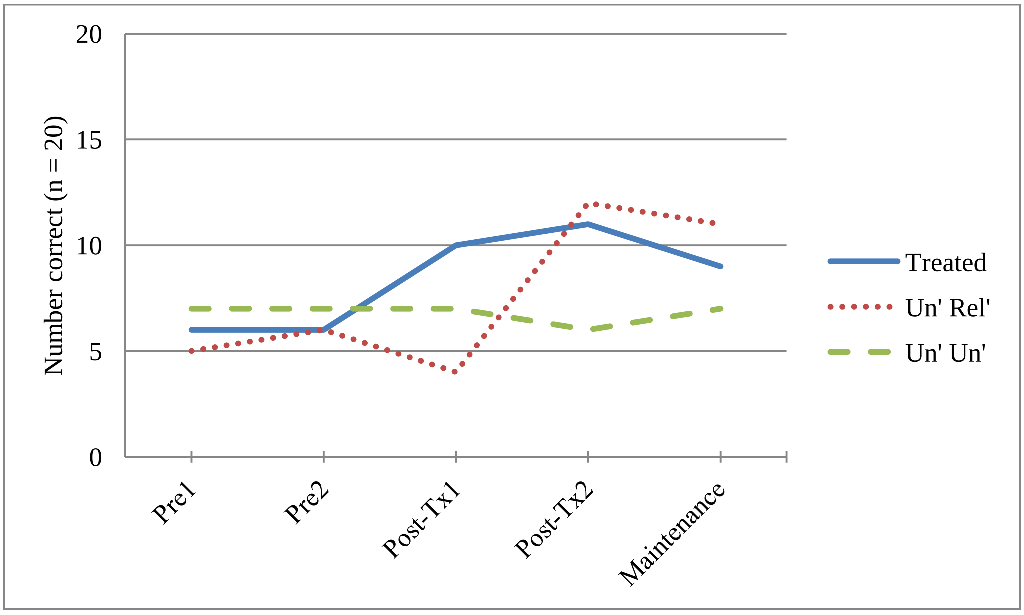

Table W7 Participant RH performance in noun picture naming (by item set; Tx1 – noun-SFA; Tx2 – verb-SFA)

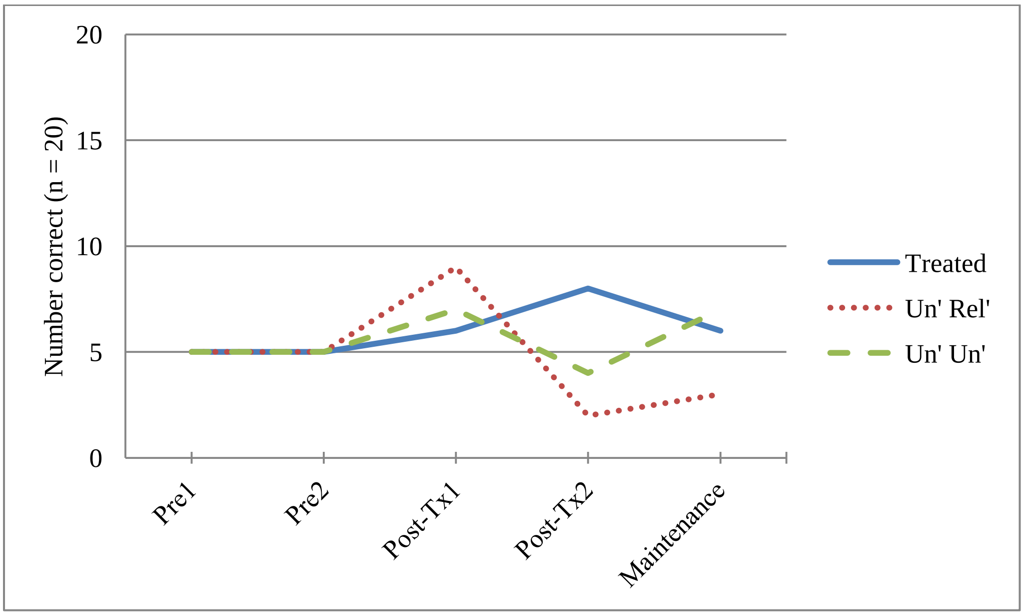

Table W8 Participant RH performance in verb picture naming (by item set; Tx1 – noun-SFA; Tx2 – verb-SFA)

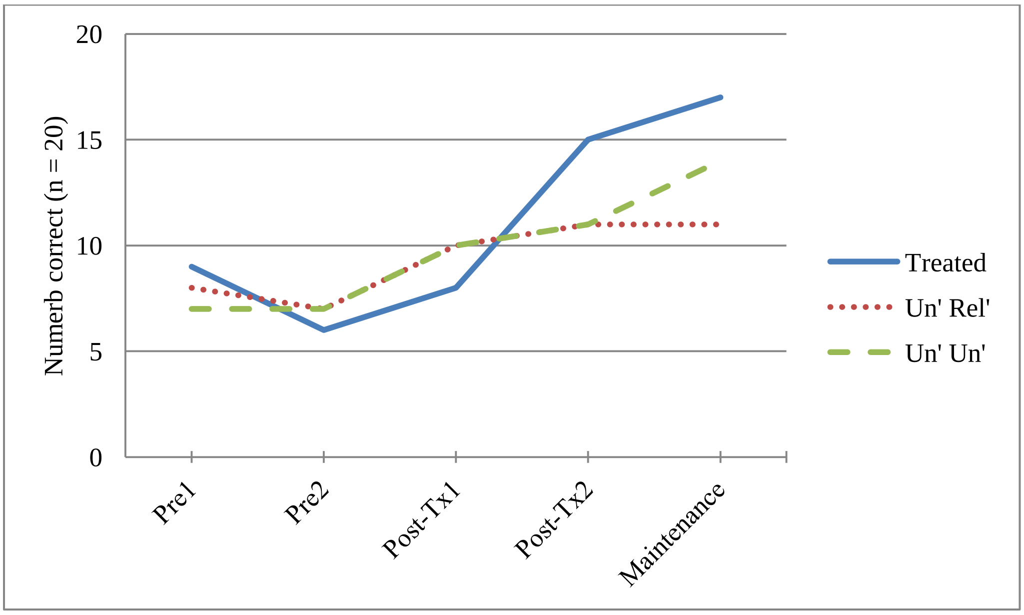

Table W9 Participant SH performance in noun picture naming (by item set; Tx1 – verb-SFA; Tx2 – noun-SFA)

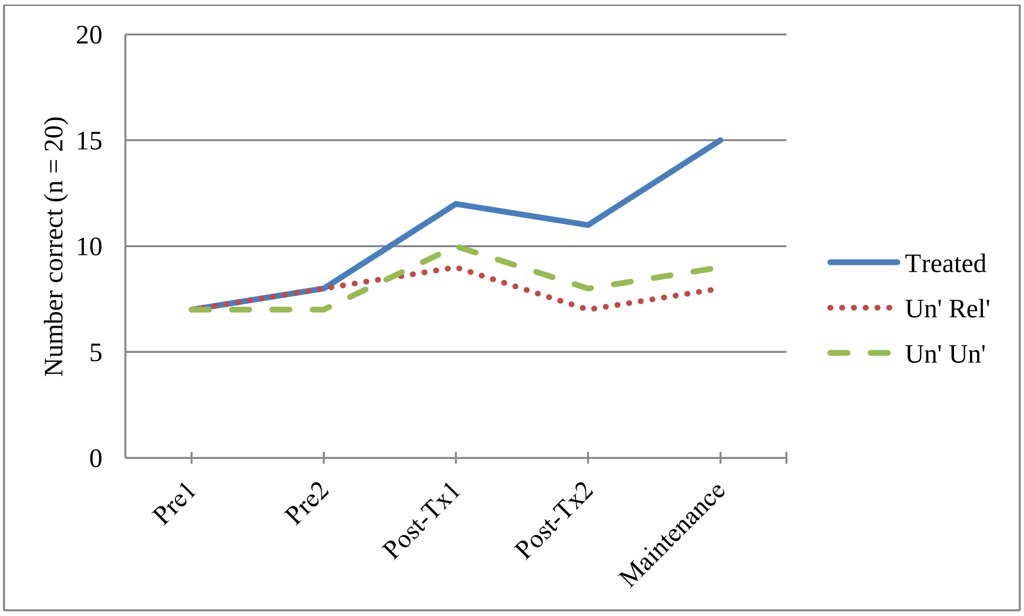

Table W10 Participant SH performance in verb picture naming (by item set; Tx1 – verb-SFA; Tx2 – noun-SFA)

Appendix X Intervention Study – OANB Quantitative Error Data

Broad error class	Mätzig et al (2009) error class	AB (pre)	AB (post)	GF (pre)	GF (post)	JA (pre)	JA (post)	RH (pre)	RH (post)	SH (pre)	SH (post)
Semantic	Coordinate	4	8	9	5	9	14	8	5	4	-
	Superordinate	-	-	-	-	1	-	-	1	-	-
	Subordinate	-	1	-	1	-	-	-	-	-	-
	Associative	3	10	7	8	4	1	6	9	1	1
	Circumlocution	1	-	1	-	-	1	2	2	3	-
Visual	Frank visual	-	2	2	2	3	1	4	1	-	1
Lexical	Misinterpretation	34	26	36	37	17	21	27	29	20	35
Phonological	Phonological	1	4	4	2	3	5	-	-	3	1
Other	Other / Mixed	2	3	3	4	2	1	1	2	6	1
No response	No response	18	6	2	1	8	12	10	5	-	1
	Total	63	60	64	60	47	56	58	54	37	40

Table X1. Error raw frequencies in Action subtest of OANB pre- and post-therapy

Broad error class	Mätzig et al (2009) error class	AB (pre)	AB (post)	GF (pre)	GF (post)	JA (pre)	JA (post)	RH (pre)	RH (post)	SH (pre)	SH (post)
Semantic	Coordinate	6.35	13.33	14.06	8.33	19.15	25	13.79	9.26	10.81	-
	Superordinate	-	-	-	-	2.13	-	-	1.85	-	-
	Subordinate	-	1.67	-	1.67	-	-	-	-	-	-
	Associative	4.76	16.67	10.94	13.33	8.51	1.79	10.35	16.67	2.7	2.5
	Circumlocution	1.59	-	1.56	-	-	1.79	3.45	3.7	8.1	-
Visual	Frank visual	-	3.33	3.13	3.33	6.39	1.79	6.9	1.85	-	2.5
Lexical	Misinterpretation	53.97	43.33	56.25	61.67	36.17	37.5	46.55	53.7	54.05	87.5
Phonological	Phonological	1.59	6.67	6.25	3.33	6.39	8.93	-	-	8.1	2.5
Other	Other / Mixed	3.18	5	4.69	6.67	4.26	1.79	1.72	3.7	16.22	2.5
No response	No response	28.57	10	3.13	1.67	17.02	21.43	17.24	9.26	-	2.5
	Total	63	60	64	60	47	56	58	54	37	40

Table X2. Error percentage proportions in Action naming subtest of OANB pre- and post-therapy

Broad error class	Mätzig et al (2009) error class	AB (pre)	AB (post)	GF (pre)	GF (post)	JA (pre)	JA (post)	RH (pre)	RH (post)	SH (pre)	SH (post)
Semantic	Coordinate	7	10	3	8	10	12	8	12	1	3
	Superordinate	-	-	4	2	-	3	-	-	-	-
	Subordinate	1	1	2	2	2	1	4	3	-	-
	Associative	7	7	16	7	6	4	12	16	3	3
	Circumlocution	-	-	2	-	-	0	3	4	1	1
Visual	Frank visual	4	1	5	1	2	7	3	2	1	-
Lexical	Misinterpretation	7	13	15	11	11	9	13	15	7	3
Phonological	Phonological	2	2	17	17	3	8	-	-	26	12
Other	Other / Mixed	6	8	13	6	2	1	5	-	5	2
No response	No response	22	6	5	7	7	6	42	25	3	4
	Total	56	48	82	61	43	51	90	77	47	28

Table X3. Error raw frequencies in Object naming subtest of OANB pre- and post-therapy

Broad error class	Mätzig et al (2009) error class	AB (pre)	AB (post)	GF (pre)	GF (post)	JA (pre)	JA (post)	RH (pre)	RH (post)	SH (pre)	SH (post)
Semantic	Coordinate	12.5	20.83	3.66	13.12	23.26	23.53	8.89	15.58	2.13	10.71
	Superordinate	-	-	4.88	3.28	-	5.88	-	-	-	-
	Subordinate	1.79	2.08	2.44	3.28	4.65	1.96	4.44	3.9	-	-
	Associative	12.5	14.58	19.51	11.48	13.95	7.84	13.33	20.78	6.39	10.71
	Circumlocution	-	-	2.44	-	-	-	3.33	5.2	2.13	3.57
Visual	Frank visual	7.14	2.08	6.10	1.64	4.65	13.73	3.33	2.6	2.13	-
Lexical	Misinterpretation	12.5	27.08	18.29	18.03	25.58	17.65	14.44	19.48	14.89	10.71
Phonological	Phonological	3.57	4.17	20.73	27.87	6.98	15.69	-	-	55.32	42.86
Other	Other / Mixed	10.71	16.67	15.85	9.84	4.64	1.96	5.56	-	10.64	7.14
No response	No response	39.29	12.5	6.10	11.48	16.28	11.77	46.67	32.47	6.39	14.29
	Total	56	48	82	61	43	51	90	77	47	28

Table X4. Error percentage proportions in Object naming subtest of OANB pre- and post-therapy

Appendix Y Intervention Study – Sentence Processing (SCAPA) Outcome Data

	Sentence comprehension			Verb comprehension		
	Pre-therapy	Post-therapy	p-value (McNemer test)	Pre-therapy	Post-therapy	p-value (McNemer test)
AB	31	33	.824	52	46	.629
GF	36	31	.359	48	44	.388
JA	29	34	.275	45	49	.285
RH	38	33	.332	48	46	.791
SH	39	37	.794	50	48	.774

Table Y1, Pre- and post-therapy performance on Sentence Comprehension subtest of SCAPA

	Pre-therapy (n = 60)	Post-therapy (n = 60)	p-value (McNemer)
AB	3	1	.625
GF	2	1	1.000
JA	6	9	.508
RH	4	6	.687
SH	30	32	.832

Table Y2. Pre- and post-therapy performance on Sentence Production subtest of SCAPA

	Verb			Subject noun			Object nouns			Thematic completeness			Syntactic completeness		
	Target	Sub	Omit	Target	Sub	Omit	Target	Sub	Omit	Target	Rev	Other	Target	Rev	Other
AB (pre)	29	23	8	40	13	7	21	17	22	12	7	41	8	12	40
AB (post)	30	24	6	42	12	6	17	18	25	12	8	40	9	11	40
GF (pre)	22	21	17	43	5	12	30	15	15	11	9	40	11	9	40
GF(post)	30	14	16	38	6	16	34	9	17	8	8	44	7	9	44
JA (pre)	12	40	8	51	9	0	33	17	10	27	18	15	23	24	13
JA (post)	20	34	6	57	3	0	39	13	8	30	16	14	26	20	14
RH (pre)	16	37	7	49	5	6	21	23	16	34	7	19	20	21	19
RH (post)	23	31	6	53	4	3	29	18	13	32	15	13	24	23	13
SH (pre)	40	17	3	59	1	0	55	3	2	41	10	9	40	11	9
SH (post)	42	16	2	59	1	0	57	2	1	41	13	6	41	13	6

Table Y3. Pre- and post-therapy noun and verb production on Sentence Production subtest of SCAPA

		AB	GF	JA	RH	SH
Verb retrieval	Target	1.000	.134	.115	.230	.791
	Sub	1.000	.167	.345	.327	1.000
	Omit	.774	1.000	.774	1.000	1.000
Subject noun retrieval	Target	.845	.458	.109	.454	1.000
	Sub	1.000	1.000	.109	1.000	1.000
	Omit	1.000	.523	-	.508	-
Object noun retrieval	Target	.523	.596	.286	.170	.687
	Sub	1.000	.307	.454	.383	1.000
	Omit	.700	.850	.791	.690	1.000
Thematic completeness	Target	1.000	.523	1.000	.345	.453
Syntactic completeness	Target	1.000	.523	1.000	.345	.453

Note.: Target – target production according to SCAPA scoring criteria; Sub – innapropriate substitution; Omit – omitted

Table Y3. Pre- versus post-therapy verb production, noun production, thematic and syntactic completeness statistical comparisons (p-value, McNemer test) on Sentence Production subtest of SCAPA

Appendix Z Intervention Study – Semantic Feature Production by Word Class

	Nouns	Verbs	Adjective/Adverb
AB	255	99	76
GF	259	125	59
JA	236	86	93
RH	239	106	82
SH	254	204	38
M	248.6	124.0	69.6
SD	10.4	46.9	21.5
Proportion	0.56	0.28	0.16

Table Z1. Feature production in verb-SFA by word class

	Nouns	Verbs	Adjective/Adverb
AB	271	75	86
GF	315	109	37
JA	290	86	57
RH	305	74	50
SH	298	185	48
M	295.8	105.8	55.6
SD	16.6	46.5	18.4
Proportion	0.65	0.23	0.12

Table Z2. Feature production in noun-SFA by word class

References

Alario, F-X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology, 53A*(3), 741-764.

Allport, D. A. (1985). Distributed memory, modular subsystems and dysphasia. In S. Newman, & R. Epstein (Eds). *Current perspectives in dysphasia* (pp32-60). Edinburgh: Churchill Livingstone.

Almor, A., Aronoff, J. M., MacDonald, M. C., Gonnerman, L. M., Kempler, D., Hintiryan, H., et al. (2009). A common mechanism in verb and noun naming deficits in Alzheimer's patients. *Brain and Language, 111* (1), 8-19.

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition, 13*, 263-308.

Ashcraft, M. H. (1978). Property norms for typical and atypical items from 17 categories: A description and discussion. *Memory & Cognition*, *6*(3), 227-232.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995).*The CELEX Lexical Database [CD ROM]*. Philadelphia, PA: The Linguistic Data Consortium, University of Pennsylvania.

Baddeley, A. D. (2004). The psychology of memory. In A. D. Baddeley, M. D. Kopelman, & B. A. Wilson (Eds). *The Essential Handbook of Memory Disorders for Clinicians* (pp 1-13). Chichester: John Wiley & Sons, Ltd.

Bak, T. H., & Hodges, J. R. (2003). Kissing and dancing - a test to distinguish the lexical and conceptual contributions to noun/verb and action/object dissociation. Preliminary results in patients with frontotemporal dementia. *Journal of Neurolinguistics, 16*, 169-181.

Barde, L. H., Schwartz, M. F., & Boronat, C. B. (2006). Semantic weight and verb retrieval in aphasia. *Brain and Language, 97*, 266-278.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11*(3), 211-227.

Basso, A. (2005). How intensive/prolonged should an intensive/prolonged treatment be? *Aphasiology*, *19*(10/11), 975-984.

Basso, A., & Marangolo, P. (2000). Cognitive neuropsychological rehabilitation: The emperor’s new clothes? *Neuropsychological Rehabilitation*, *10*(0), 219-229.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, *80*(3), 1-46.

Berndt, R. S., Haendiges, A. N., Burton, M. W., & Mitchum, C. C. (2002). Grammatical class and imageability in aphasic word production: Their effects are independent. *Journal of Neurolinguistics, 15*, 353-371.

Berndt, R. S., Haendiges, A. N., Mitchum, C. C., & Sandson, J. (1997). Verb retrieval in aphasia: 2. Relationship to sentence processing. *Brain and Language, 56,* 107-137.

Berndt, R. S., Mitchum, C. C., Haendiges, A. N., & Sandson, J. (1997). Verb retrieval in aphasia: 1. Characterizing single word impairments. *Brain and Language, 56*, 68-106.

Best, W., & Nickels, L. (2000). From theory to therapy in aphasia: Where are we now and where to next? *Neuropsychological Rehabilitation, 10* (3), 231-247.

Bi, Y., Han, Z., Shu, H., & Caramazza, A. (2007). Nouns, verbs, objects, actions, and the animate/inanimate effect. *Cognitive Neuropsychology, 24* (5), 485-504.

Bird, H., Howard, D., & Franklin, S. (2003). Verbs and nouns: The importance of being imageable. *Journal of Neurolinguistics, 16*, 113-149.

Black, M., & Chiat, S. (2003). Noun-verb dissociations: A multi-faceted phenomenon. *Journal of Neurolinguistics, 16*, 231-250.

Bock, K. J. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355-387.

Bock, J. K. (1995). Sentence production: From mind to mouth. In J. Miller, & P. Eimas (Eds.), *Handbook of Perception and Cognition: Speech, Language and Communication* (pp. 181-216). New York, US: Academic Press.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341-345.

Boo, M., & Rose, M. L. (2011). The efficacy of repetition, semantic, and gesture treatments for verb retrieval and use in Broca's aphasia. *Aphasiology, 25* (2), 154-175.

Boster, J. S. (1988). Natural sources of internal category structure: Typicality, familiarity, and similarity of birds. *Memory & Cognition, 16*(3), 258-270.

Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech-Language Pathology, 13*, 236-249.

Boyle, M. (2010). Semantic feature analysis treatment for aphasia word retrieval impairments: What's in a name? *Topics in Stroke Rehabilitation, 17* (6), 411-422.

Boyle, M., & Coelho, C. A. (1995). Application f semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology, 4*, 94-98.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*, B13-25.

Breedin, S. D., & Martin, R. C. (1996). Patterns of verb impairment in aphasia: An analysis of four cases. *Cognitive Neuropsychology, 13* (1), 51-92.

Breedin, S. D., Saffran, E. M., & Schwartz, M. F. (1998). Semantic factors in verb retrieval: An effect o complexity. *Brain and Language, 63*, 1-31.

British National Corpus, version 3 (BNC XML Edition). (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: <http://www.natcorp.ox.ac.uk/> [Accessed 2nd July 2011]

Capitani, E., Laiacona, M., & Barbarotto, R. (1999). Gender affects word retrieval of certain categories in semantic fluency tasks. *Cortex*, *35*, 273-278.

Capitani, E., Laiacona, M., Mahon, B., & Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cognitive Neuropsychology*, *20*(3/4/5/6), 213-261.

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature, 349*, 788-790.

Casey, P. J. (1992) A re-examination of the roles of typicality and category dominance in verifying category membership. *Journal of Experimental Psychology: Learning, Memory & Cognition, 18*(4), 263-273.

Chang, T. M. (1986). Semantic memory: Facts and models. *Psychological Bulletin, 99*(2), 199-220.

Coelho, C. A., McHugh, R. E., & Boyle, M. (2000). Semantic feature analysis as a treatment for aphasic dysnomia: A replication. *Aphasiology, 14* (2), 133-142.

Collina, S., Marangolo, P., & Tabossi, P. (2001). The role of argument structure in the production of nouns and verbs. *Neuropsychologia, 39*, 1125-1137.

Collina, S., & Tabossi, P. (2007). Semantic interference effects in the production of verbs: The role of response set. *The Mental Lexicon, 2*(1), 65-78.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407-428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 240-247.

Coltheart, M., Inglis, L., Cupples, L., Michie, P., Bates, A., & Budd, B. (1998). A semantic subsystem of visual attributes. *Neurocase, 4,* 353-370.

Conroy, P., Sage, K., & Lambon Ralph, M. A. (2009a). Errorless and errorful therapy for verb and noun naming in aphasia. *Aphasiology, 23* (11), 1311-1337.

Conroy, P., Sage, K., & Lambon Ralph, M. A. (2009b). The effects of decreasing and increasing cue therapy on improving naming speed and accuracy for verbs and nouns in aphasia. *Aphasiology, 23* (6), 707-730.

Conroy, P., Sage, K., & Lambon Ralph, M. A. (2006). Towards theory-driven therapies for aphasic verb impairments: A review of current theory and practice. *Aphasiology, 20* (12), 1159-1185.

Conti-Ramsden, G. & Jones, M. (1997). Verb use in Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 40,* 1298-1313.

Cree, G. S., McNorgan, C., & McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 643-658.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General, 132(2),* 163-201.

Crutch, S. J., & Warrington, E. K. (2008). Contrasting patterns of comprehension for superordinate, basic-level, and subordinate names in semantic dementia and aphasic stroke patients. *Cognitive Neuropsychology, 25*(4), 582-600.

Damasio, A., & Tranel, D. (1993). Nouns and verbs are retrieved with differentially distributed neural systems. *Proceedings of the National Academy of Sciences, 90,* 4957-4960.

Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition, 81,* B77-B86.

De Bleser, R., & Kauschke, C. (2003). Acquisition and loss of nouns and verbs: Parallel or divergent patterns? *Journal of Neurolinguistics, 16,* 213-229.

den Ouden, D-B., Fix, S., Parrish, T. B., & Thompson, C. K. (2009). Argument structure effects in action verb naming in static and dynamic conditions. *Journal of Neurolinguistics, 22,* 196-215.

Devlin, J. T., Gonnerman, L. M., Andersen, E. S., & Seidenberg, M. S. (1998). category-specific semantic deficits in focal and widespread brain damage: A computational account. *Journal of Cognitive Neuroscience, 10*(1), 77-94.

Druks, J. (2002). Verbs and nouns—a review of the literature. *Journal of Neurolinguistics, Vol. 15,* 289-315.

Druks, J., & Masterson, J. (2000). *An Object and Action Naming Battery*. Hove: Psychology Press.

Dufour, S., & Peereman, R. (2003). Lexical competition in phonological priming: Assessing the role of phonological match and mismatch lengths between primes and targets. *Memory & Cognition*, *31*(8), 1271-1283.

Edmonds, L. A., Nadeau, S. E., & Kiran, S. (2009). Effect of Verb Network Strengthening Treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. *Aphasiology*, *23* (3), 402-424.

Edwards, S., & Tucker, K. (2006). Verb retrieval in fluent aphasia: A clinical study. *Aphasiology, 20* (7), 644-675.

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General, Vol. 120*(4), 339-357.

Faroqi-Shah, Y., & Graham, L. E. (2011). Treatment of semantic verb classes in aphasia: Acquisition and generalization effects. *Clinical Linguistics & Phonetics*, *25* (5), 399-418.

Faul, F., Erdfelder, E., Lang, A-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Fillmore, C J, & Atkins, B. T. S (2000). Describing polysemy: The case of "crawl". In C. Leacock (Ed). *Polysemy: Theoretical and computational approaches* (pp 91-110). Oxford: Oxford University Press.

Fink, R. B., Martin, N., Schwartz, M. F., Saffran, E. M., & Myers, J. L. (1992). Facilitation of verb retrieval skills in aphasia: A comparison of two approaches. *Clinical Aphasiology, 21*, 263-275.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116-124.

Funnell, E. (2000). Models of semantic memory. In. W. Best, K. Bryan, & J. Maxim (Eds). *Semantic Processing: Theory and Practice* (pp1-27). London: Whurr Publishers.

Garrard, P., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelations: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology, 18*(2), 125-174.

Garrard, P., Perry, R., & Hodges, J. R. (1997). Disorders of semantic memory. *Journal of Neurology, Neurosurgery, and Psychiatry, 62*, 431-435.

Garrett, M. F. (1982). Production of speech: Observations from normal and pathological language use. In A. W. Ellis (Ed.), *Normality and Pathology in Cognitive Functions*. London: Academic Press.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczak (Ed). *Language development, Volume 2: Language, though and culture* (pp 301-334). Hillsdale, NJ.; Lawrence Erlbaum.

Gordon, J. K., Dell, G. S. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science, 27*, 1-40.

Graesser, A. C., Hopkinson, P. L., and Schmid, C. (1987). Differences in interconcept organization between nouns and verbs. *Journal of Memory and Language, 26*(2), 242-253.

Hamburger, M., & Slowiaczek, L. A. (1996). Phonological priming reflects lexical competition. *Psychonomic Bulletin & Review, 3*(4), 520-525.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18*, 441-461.

Hampton, J. A. (1997). Associative and similarity-based processes in categorization decisions. *Memory & Cognition, 25*(5), 625-640.

Hampton, J., & Bubois, D. (1992). Psychological models of concepts: Introduction. In. I. V. Mechelen (Ed). *Categories and Concepts: Theoretical Views and Inductive Data Analysis* (pp11-33). London: UK Academic Press.

Hampton, J., & Gardiner, M. M. (1983). Measures of internal category structure. A correlational analysis of normative data. *British Journal of Psychology, 74*(4), 491-516.

Hantsch, A., Jescheniak, J. D., & Schriefers, H. (2005). Semantic competition between hierarchically related words during speech planning. *Memory & Cognition, 33*(6), 984-1000.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley & Sons

Hemeren, P. E. (1996). Frequency, ordinal position and semantic distance as measures of cross-cultural stability and hierarchies for action verbs. *Acta Psychologica, 91*, 39-66.

Hillis, A. (1998). Treatment of naming disorder: New issues regarding old therapies. *Journal of the International Neuropsychological Society, 4*, 648-660.

Hines, D., Czerwinski, M., Sawyer, P. K., & Dwyer, M. (1986). Automatic semantic priming: Effect f category exemplar level and word association level. *Journal of Experimental Psychology: Human Perception and Performance, 12*(3), 370-379.

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences, 9*, 1-66.

Holmes, S. J., & Ellis, A. W. (2006). Age of acquisition and typicality effects in three object processing tasks. *Visual Cognition, 13*(7-8), 884-910.

Horton, S., & Byng, S. (2002). "Semantic Therapy" in day-to-day clinical practice: Perspectives on diagnosis and therapy related to semantic impairment in aphasia. In A. E. Hillis (Ed.), *The Handbook of Adult Language Disorders* (pp. 229-249). New York: Psychology Press.

Howard, D. (2000). Cognitive neuropsychology and aphasia therapy: The case of word retrieval. In T. Papathanassiou (Ed.), *Acquired Neurogenic Communication Disorder*s (pp. 76-99). London: Whurr Publishers.

Howard, D., & Hatfield, F. M. (1987). *Aphasia Therapy: Historical and Contemporary Issues*. London: Lawrence Erlbaum.

Howard, D., & Patterson, K. E. (1992). *The Pyramids and Palm Trees Test*. Bury St. Edmunds: Thames Valley Test Corporation.

Howard, D., Patterson, K., Franklin, S., Orchard-Lisle, V., & Morton, J. (1985). Treatment of word retrieval deficits in aphasia: A comparison of two therapy methods. *Brain, 108*, 817-829.

Huttenlocher, J., & Lui, F. (1979). The semantic organization of some simple nouns and verbs. *Journal of Verbal Learning and Verbal Behavior, 18*, 141-162.

Jackendoff, R. S. (1983). *Semantics and Cognition*. Cambridge, Mass, US: MIT Press.

Keefe, D. E., & Neely, J. H. (1990). Semantic priming in the pronunciation task: The role of prospective prime-generated expectancies. *Memory & Cognition, 18*(3), 289-298.

Kemmerer, D., & Tranel, D. (2000). Verb retrieval in brain-damaged subjects: 1. Analysis of stimulus, lexical, and conceptual factors. *Brain and Language, 73*, 347-392.

Kertesz, A. (1982). *Western Aphasia Battery*. New York, US: Grune Stratton.

Kim, M., & Thompson, C. K. (2000). Patterns of comprehension and production of nouns and verbs in agrammatism: Implications for lexical organization. *Brain and Language, 74*, 1-25.

Kim, M., Adingono, M. F., & Revoir, J. S. (2007). Argument structure enhanced verb naming treatment: Two case studies. *Contemporary Issues in Communication Science and Disorders, 34*, 24-36.

Kiran, S. (2008). Typicality of inanimate category exemplars in aphasia treatment: Further evidence for semantic complexity. *Journal of Speech, Language, and Hearing Research, 51*, 1550-1568.

Kohn, S. E., Lorch, M. P., & Pearson, D. M. (1989). Verb finding in aphasia. *Cortex, 25*, 57-69.

Kiran, S., Ntourou, K., & Eubank, M. (2007). The effect of typicality on online category verification of inanimate category exemplars in aphasia. *Aphasiology, 21*(9), 844-866.

Kiran, S., Shamapant, S., & DeLyria, S. K. (2006). Typicality within well defined categories in aphasia. *Brain & Language, 99*, 159-161.

Kiran, S., & Thompson, C. K. (2003). Effect of typicality on online category verification of animate category exemplars in aphasia. *Brain and Language, 85*, 441-450.

Kuipers, J-R., La Heij, W., & Costa, A. (2006). A further look at semantic context effects in language production: The role of response congruency. *Language and Cognitive Processes, 21*(7-8), 892-919.

Laiacona, M., Barbarotto, R., & Capitani, E. (2006). Human evolution and the brain representation of semantic knowledge: Is there a role for sex differences? *Evolution & Human Behavior, 27*, 158-168.

Lakoff, G. (1989). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: The University of Chicago Press.

Lambon Ralph, M. A., & Howard, D. (2000). Gogi aphasia or semantic dementia? Simulating and assessing poor verbal comprehension in a case of progressive fluent aphasia. *Cognitive Neuropsychology, 17*(5), 437-465.

Larochelle, S., & Pineau, H. (1994) Determinants of response times in the semantic verification task. *Journal of Memory and Language, 33*(6), 796-823.

Larochelle, S., Richard, S., & Soulières, I. (2000). What some effects might not be: The time to verify membership in “well-defined” categories. *The Quarterly Journal of Experimental Psychology, 53A*(4), 929-961.

Laws, K. R. (2004). Sex differences in lexical size across semantic categories. *Personality and Individual Differences, 36*, 23-32.

Le Dorze, G., Boulay, N., Gaudreau, J., & Brassard, C. (1994). The contrasting effects of a semantic versus a formal-semantic technique for the facilitation of naming in a case of anomia. *Aphasiology, 8* (2), 127-141.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-45.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: The University of Chicago Press.

Lowell, S., Beeson, P. M., & Holland, A. L. (1995). The efficacy of a semantic cueing procedure on naming performance of adults with aphasia. *American Journal of Speech-Language Pathology, 4*, 109-114.

Lupker, S. J. (1988). Picture naming: An investigation of the nature of categorical priming. *Journal of Experimental Psychology: Leaning, Memory, and Cognition, 14*(3), 444-455.

Luzzatti, C., Aggujaro, S., & Crepaldi, D. (2006). Verb-noun double dissociation in aphasia: Theoretical and neuroanatomical foundations. *Cortex, 42*, 875-883

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology, 60*, 27-51.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 503-535.

Maki, W. S., Krimsky, M., & Muñoz, S. (2006). An efficient method for estimating semantic similarity based on feature overlap: Reliability and validity of semantic feature ratings. *Behavior Research Methods, 38*(1), 153-157.

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior, 23*, 250-269.

Marques, J. M. (2007). The general/specific breakdown of semantic memory and the nature of superordinate knowledge: Insights from superordinate and basic-level feature norms. *Cognitive Neuropsychology, 24*(8), 879-903.

Marshall, J. (2003). Noun-verb dissociations—evidence from acquisition and developmental and acquired impairments. *Journal of Neurolinguistics, 16*, 67-84.

Marshall, J., Chiat, S., Robson, J., & Pring, T. (1996). Calling a salad a federation: An investigation of semantic jargon. Part 2 – Verbs. *Journal of Neurolinguistics, 9*(4), 251-260.

Marshall, J., Pound, C., White-Thompson, M., & Pring, T. (1990). The use of picture/word matching tasks to assist word retrieval in aphasic patients. *Aphasiology, 4* (2), 167-184.

Marshall, J., Pring, T., & Chiat, S. (1998). Verb retrieval and sentence production in aphasia. *Brain and Language, 63*, 159-183.

Marshall, J., Pring, T., Chiat, S., & Robson, J. (1996). Calling a salad a federation: An investigation of semantic jargon. Part 1 – Nouns. *Journal of Neurolinguistics, 9*(4), 237-250.

Mätzig, S., Druks, J., Masterson, J., & Vigliocco, G. (2009). Noun and verb differences in picture naming: Past studies and new evidence. *Cortex, 45*, 738-758.

McCarthy, R., & Warrington, E. K. (1985). Category specificity in an agrammatic patient: The relative impairment of verb retrieval and comprehension. *Neuropsychologia, 23* (6), 709-727.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition, 6*(4), 462-472.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Hove: Psychology Press.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Leaning, Memory, and Cognition, 24*(3), 558-572.

McRae, K., & Cree, G. S. (2002). Factors underlying category specific semantic deficits. In. E. M. E. Forde, & G. W. Humphreys (Eds). *Category Specificity in Brain and Mind* (pp 211-249). Hove: Psychology Press.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, & Computers, 37*(4), 547-559.

McRae, K., Cree, G. S., Westmacott, R., & de Sa, V. R. (1999). Further evidence for feature correlations in semantic memory. *Canadian Journal of Experimental Psychology: Special Issue on Models of Word Recognition, 53,* 360-373.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General, 126*(2), 99-130.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44*(12), 1469-1481.

Mervis, A. B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society, 7*(3), 283-284.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227-234.

Miceli, G., Silveri, M. C., Villa, G., & Caramazza, A. (1984). On the basis for the agrammatic's difficulty in producing main verbs. *Cortex, 20*, 207-220.

Miller, G. A., & Fellbaum, C. (1991). Semantic networks of English. *Cognition, 41*, 197-229.

Miller, G. A., & Fellbaum, C. (2007). WordNet then and now. *Language Resources & Evaluation, 41*, 209-214.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and Perception*. Cambridge: Cambridge University Press.

Mirman, D., & Magnuson, J. S. (2009). The effect of frequency of shared features on judegments of semantic similarity. *Psychonomic Bulletin & Review, 16*(4), 671-677.

Mitchum, C. C., Greenwald, M. L., & Berndt, R. S. (2000). Cognitive treatment of sentence processing disorders: What have we leaned? *Neuropsychological Rehabilitation, 10*(3), 311-336.

Mitchum, C. C., Haendiges, A. N., & Berndt, R. S. (1995). Treatment of thematic mapping in sentence comprehension: Implications for normal processing. *Cognitive Neuropsychology, 12*(5), 503-547.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review, 76*, 165-178.

Morton, J., & Patterson, K. E. (1980). A new attempt at an interpretation, or an attempt at a new interpretation. In M. Coltheart, K. E. Patterson, & J. C. Marshall (Eds). *Deep dyslexia*. London: Routledge & Kegan Paul.

Moss, H. E., Tyler, L. K., Durrant-Peatfield, M., & Bunn, E. M. (1998). 'Two eyes of a see-through': Impaired and intact semantic knowledge in a case of selective deficit for living things. *Neurocase, 4*, 291-310.

Moss, H. E., Tyler, L. K., & Jennings, F. (1997). When leopards lose their spots: Knowledge of visual properties in category-specific deficits for living things. *Cognitive Neuropsychology, 14*(6), 901-950.

Moss, H. E., Tyler, L. K., & Taylor, K. I. (2007). In M. G. Gaskell (Ed). *The Oxford Handbook of Psycholinguistics* (pp217-234). Oxford: Oxford University Press.

Murphy, G. L. (2002). *The Big Book of Concepts*. Cambridge, MA: The MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*(3), 289-316.

Nadeau, S. E., & Kendall, D. L. (2006). Significance and possible mechanisms underlying generalization in aphasia therapy: Semantic treatment of anomia. *Brain and Language, 99*, 8-219.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds). *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Erlbaum.

Nickels, L. (2000). Semantics and therapy in aphasia. In W. Best, K. Bryan, & J. Maxim (Eds.), *Semantic Processing: Theory and Practice* (pp. 108-124). London: Whurr Publishers.

Nickels, L., & Best, W. (1996). Therapy for naming disorders (Part I): Principles, puzzles and progress. *Aphasiology, 10* (1), 21-47.

O'Connor, C. M., Cree, G. S., & McRae, K. (2009). Conceptual hierarchies in a flat attractor network: Dynamics of learning and computations. *Cognitive Science, 33*, 665-708.

Pashek, G. V., & Tompkins, C. A. (2002). Context and word class influences on lexical retrieval in aphasia. *Aphasiology, 16*(3), 261-286.

Patterson, K. (2007). The reign of typicality in semantic memory. *Philosophical Transactions of the Royal Society, 362*, 813-821.

Patterson, K. E., & Shewell, C. (1987). Speak and spell: Dissociations and word-class effects. In M. Coltheart, R. Job, & G. Sartori (Eds.).*The Cognitive Neuropsychology of Language* (pp. 273-294). Hillsdale, NJ: Lawrence Erlbaum.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, Mass.: MIT Press.

Plant, C., Webster, J., & Whitworth, A. (2011). Category norm data and relationships with lexical frequency and typicality within verb semantic categories. *Behavior Research Methods, 43*, 424-440.

Plaut, D. C. (1996). Relearning after damage in connectionist networks: Toward a theory of rehabilitation. *Brain and Language, 52*, 25-82.

Ramsay, C. B., Nicholas, M., Au, R., Obler, L. K., & Albert, M. L. (1999). Verb naming in normal aging. *Applied Neuropsychology, 6* (2), 57-67.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 2: The Coloured Progressive Matrices*. San Antonio, TX, US: Harcourt Assessment.

Raymer, A. M., & Ellsworth, T. A. (2002). Response to contrasting verb retrieval treatments: A case study. *Aphasiology, 16* (10/11), 1031-1045.

Raymer, A. M., Ciampitti, M., Holliway, B., Singletary, F., Blonder, L. X., Ketterson, T., et al. (2007). Semantic-phonologic treatment for noun and verb retrieval impairments in aphasia. *Neuropsychological Rehabilitation, 17* (2), 244-270.

Raymer, A., & Kohen, F. (2006). Word-retrieval treatment in aphasia: Effects of sentence context. *Journal of Rehabilitation Research & Development, 43* (3), 367-378.

Rifkin, A. (1985). Evidence for a basic level in event taxonomies. *Memory & Cognition, 13*(6), 538-556.

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition, 47*, 59-87.

Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1996). The relation between typicality and semantic similarity structure. *Journal of Quantitative Anthropology, 6*(1-4), 1-14.

Rosch, E. (1978). Principles of categorisation. In E. Rosch, and B. B. Lloyd (Eds). *Cognition and Categorisation* (pp27-48). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology, 8*, 382-439.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance, 2*(4), 491-502.

Rose, M., & Sussmilch, G. (2008). The effects of semantic and gesture treatments on verb retrieval and verb use in aphasia. *Aphasiology*, *22*(7-8), 691-706.

Rösler, F., Streb, J., & Haan, H. (2001). Even-related brain potentials evoked by verbs and nouns in a primed lexical decision task. *Psychophysiology, 38*, 694-703.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*, 217-236.

Sage, K., Snell, C., & Lambon Ralph, M. A. (2011). How intensive does anomia therapy for people with aphasia need to be? *Neuropsychological Rehabilitation, 21*(1), 26-41.

Schnur, T. T., Costa, A., & Caramazza, A. (2002). Verb production and the semantic interference effect. *Journal of Cognitive Science, 3*, 1-26.

Schwartz, M. F., Saffran, E. M., Fink, R. B., Myers, J. L., & Martin, N. (1994). Mapping therapy: A treatment program for agrammatism. *Aphasiology, 8*, 19-54.

Sebastian, R., & Kiran, S. (2007). Effect of typicality of ad hoc categories in lexical access. *Brain & Language, 103*, 138-139.

Shapiro, K., & Caramazza, A. (2003). The representation of grammatical categories in the brain. *Trends in Cognitive Sciences*, *7*(5), 201-206.

Shapiro, L. P., & Levine, B. A. (1990). Verb processing during sentence comprehension in aphasia. *Brain and Language, 38*, 21-47.

Shapiro, L. P., Zurif, E., & Grimshaw, J. (1987). Sentence processing and the mental representation of verbs. *Cognition, 27*, 219-246.

Shapiro, L. P., Zurif, E., Grimshaw, J. (1989). Verb processing during sentence comprehension: Contextual impenetrability. *Journal of Psycholinguistic Research, 18*(2), 223-243.

Smith, E. E., & Medin, D. L. (1981). *Categories and Concepts*. Cambridge, Massachusetts: Harvard University Press.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*(3), 214-241.

Snowden, J. (2002). Disorders of semantic memory. In A. D. Baddeley, M. D. Kopelman, & B. A. Wilson (Eds). *The Handbook of Memory Disorders (2nd Edition)* (pp293-314). Chichester: John Wiley & Sons, Ltd.

Stanczak, L., Waters, G., & Caplan, D. (2006). Typicality-based learning and generalisation in aphasia: Two case studies of anomia treatment. *Aphasiology, 20* (2/3/4), 374-383.

SurveyMonkey.com, LLC. (n.d). Palo Alto, California, USA. <http://www.surveymonkey.com/> [accessed 2nd July 2011].

Swinburn, K., Porter, G., & Howard, D. (2004). *The Comprehensive Aphasia Test*. Hove: Psychology Press.

Tabossi, P., & Collina, S. (2002). The picture-word interference paradigm: Conceptual effects in the production of verbs. *Rivista di Linguistica, 14*(1), 27-41.

Taylor, J. R. (2003). *Linguistic categorization (3rd Edition)*. Oxford, UK: Oxford University Press.

Thompson, C. K., Lange, K. LO., Schneider, S. L., & Shapiro, L. P. (1997). Agrammatic and non-brain-damaged subjects' verb and verb argument structure production. *Aphasiology, 11*(4/5), 473-490.

Thompson, C. K., Shapiro, L. P., Kiran, S., & Sobecks, J. (2003). The role of syntactic complexity in treatment of sentence deficits in agrammatic aphasia: The Complexity Account of Treatment Efficacy (CATE). *Journal of Speech, Language, and Hearing Research, 46*, 591-607.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving (Ed). *Organization of memory* (pp381-403). New York: Academic Press.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science, 247*, 301-306.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113*(2), 169-193.

Tyler, L. K., & Moss, H. E. (1997). Functional properties of concepts: Studies of normal and brain-damaged patients. *Cognitive Neuropsychology, 14*(4), 511-545.

Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of masked priming: A meta-analysis. *Psychological Bulletin, 135*(3), 452-477.

Vanoverberghe, V., & Storms, G. (2003). Feature importance in feature generation and typicality rating. *European Journal of Cognitive Psychology, 15*(1), 1-18.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory & Language, 50*, 289-335.

Vigliocco, G., & Vinson, D. (2007). Semantic representation. In M. G. Gaskell (Ed). *The Oxford Handbook of Psycholinguistics* (pp195-215). Oxford: Oxford University Press.

Vigliocco, G., Vinson, D. P., Damian, M. F., & Levelt, W. (2002). Semantic distance effects on object and action naming. *Cognition*, *85*(3), B61-B69

Vigliocco, G., Vinson, D. P., Lewis, L., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology, 48*, 422-488.

Vinson, D. P., & Vigliocco, G. (2002). A semantic analysis of grammatical class impairments: Semantic representations of object nouns, action nouns and action verbs. *Journal of Neurolinguistics, 15,* 17-351.

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods, 40*(1), 183-190.

Vinson, D., Vigliocco, G., Cappa, S., & Siri, S. (2003). The breakdown of semantic knowledge: Insights from a statistical model of meaning representation. *Brain and Language, 86*, 347-365.

Vitkovitch, M., & Tyrrell, L. (1999). The effects of distractor words on naming pictures at the subordinate level. *The Quarterly Journal of Experimental Psychology, 52A*(4), 905-926.

Waechter, S., Stolz, J. A., & Besner, D. (2010). Visual word recognition: On the reliability of repetition priming. *Visual Cognition, 18*(4), 537-558.

Wambaugh, J. L., & Ferguson, M. (2007). Application of semantic feature analysis to retrieval of action names in aphasia. *Journal of Rehabilitation Research & Development, 44* (3), 381-394.

Wambaugh, J., Cameron, R., Kalinyak-Fliszar, M., Nessler, C., & Wright, S. (2004). Retrieval of action names in aphasia: Effects of two cueing treatments. *Aphasiology, 18* (11), 979-1004.

Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology, 27*, 635-357.

Warrington, E. K., & McCarthy, R. (1987). Categories of Knowledge: Further fractionations and an attempted integration. *Brain, 110*, 1273-1296.

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain, 107*, 829-854.

Webster, J., Franklin, D., & Howard, D. (2007). An analysis of thematic and phrasal structure in people with aphasia: What more can we learn from the story of Cinderella? *Journal of Neurolinguistics, 20,* 363-394.

Webster, J., & Gordon, B. (2009). Contrasting therapy effects for verb and sentence processing difficulties: A discussion of what worked and why. *Aphasiology, 23* (10), 1231-1251.

Webster, J., Morris, J., & Franklin, S. (2005). Effects of therapy targeted at verb retrieval and the realisation of the predicate argument structure: A case study. *Aphasiology, 19* (8), 748-764.

Webster, J., Morris, J., Whitworth, A., & Howard, D. (2009). *Newcastle University Aphasia Therapy Resources: Sentence Processing*. Newcastle upon Tyne: Newcastle University.

Webster, J., & Whitworth, A. (in prep). *Sentence Comprehension and Production in Aphasia*.

Wheeldon, L. R., & Monsell, S. (1994). Inhibition of spoken word production by priming a semantic competitor. *Journal of Memory and Language, 33,* 332-356.

Whitworth, A., Webster, J., & Howard, D. (2005). *A Cognitive Neuropsychological Approach to Assessment an Intervention in Aphasia: A Clinician's Guide*. London: Psychology Press.

Williams, S. E., & Canter, G. J. (1987). Action-naming performance in four syndromes of aphasia. *Brain and Language, 32,* 124-136.

Wilshere, C. E., Keall, L. M., Stuart, E. J., & O'Donnell, D. J. (2007). Exploring the dynamics of aphasia word production using the picture-word interference task: A case study. *Neuropsychologia, 45*, 939-953.

Wilson, M. (1988). MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers, 20*(1), 6-10.

Wiseburn, B., & Mahoney, K. (2009). A meta-analysis of word-finding treatments for aphasia. *Aphasiology, 23* (11), 1338-1352.

Zingeser, L. B., & Berndt, R. S. (1990). Retrieval of nouns and verbs in agrammatism and anomia. *Brain and Language, 39*, 14-32.