The interpretation of noun noun compounds

Davidson, Oliver Geoffrey

Davidson, Oliver Geoffrey (1996) *The interpretation of noun noun compounds*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/5301/

The Interpretation of Noun Noun Compounds

Oliver Geoffrey Davidson

1996

A thesis submitted for the degree of Doctor of Philosophy in the Department of Psychology, University of Durham.

Abstract

This thesis looks at conceptual combination, in particular it investigates how noun noun compounds are interpreted. Several themes run throughout the work. Real compounds (e.g. coat hanger, crab apple) are compared to novel ones (e.g. banjo cactus, zip violin). Also, compounds are examined in each of the possible permutations of artefacts (A) (e.g. coat, banjo) and natural kinds (N) (e.g. crab, cactus), (AA, AN, NA and NN).

Experiments 1 - 4 examine noncompositionality in noun noun compounds. Possible sources of noncompositionality are investigated using both feature listing and feature rating tasks. Although some differences were found, results were similar between different types of compound, evidence of noncompositionality being found in each. The results also confirm that most of the meaning of a noun noun compound is derived from the second constituent (noun2).

Experiments 5 and 6 look at two different types of compound interpretation - slot filling and property mapping. In experiment 5, slot filling is found to be the preferred interpretation type overall, but property mapping is more common in compounds composed of two natural kinds (NN). Experiment 6 examines possible factors influencing the choice between slot filling and property mapping interpretations. It was found that constituent similarity plays an important role, and also that this interacts with whether or not the constituents have important properties which clash.

Experiment 7 looks at compound identification. Results suggest that the first constituent (noun1) may be critical in such tasks. Experiment 8 compares the importance of noun1 and noun2 in determining the type of interpretation given to a compound. Neither position is found to be more influential than the other, although relational information does seem to be associated with specific nouns in each position.

Throughout the thesis findings are related to current theories of conceptual combination, such as prototype models, the concept specialisation model and theories of compound interpretation by analogy.

Acknowledgements

Thanks first to my supervisor, Rosemary Stevenson, for her guidance and support throughout this research. I'm also grateful to others at Durham, particularly Jan, Bruce and Graham, and of course all those tolerant people who participated in these experiments.

This work was carried out as a research project funded by the Human Communication Research Centre, so my thanks go to them.

This work was also made possible by the love and support of Agnieszka, Giles, Matthew, Harry and Mum. My thanks to them, and I hope they enjoy this thesis.

Table of Contents

Declaration

The material contained in this thesis has not previously been submitted for any degree.

Chapter one: Introduction and Literature Review

What Is A Concept ?

A seemingly appropriate issue to address before discussing details of research, is that of what counts as a concept. However, perhaps unsurprisingly, this is not an easy question to answer. There are two identifiable and distinct forms of circularity related to this issue, the first is straight-forward and problematic, the second is less simple, but perhaps more helpful.

(i) We ask the question, "What is a concept ?", as a prelude to the study of concepts, since it makes sense to have some idea of what we are meant to be investigating before we begin. But, inevitably, a principle reason for wanting to study concepts is that we don't yet know the answer to this very question, and we are trying to find it out. So, in some sense there is no possible, logical starting point for such a programme of investigation.

(ii) In asking, "What is a concept ?" we are formulating a conceptual question in a sense beyond the trivial - it could be rephrased, "What is the concept "concept" ?" We presuppose some answer, moreover we are implicitly making a judgement as to what kind of thing we are prepared to accept as an answer. That is to say, the appreciation of what the question itself may entail gives us some indication of what a concept is. Undeniably such inferences are nebulous and inadequate, but they at least provide a starting point.

From such an insight, we can derive an initial understanding of concepts - they are a cognitive tool for information processing. Rosch (1978) describes two "general and basic principles" with respect to the role of categories.

> "The first has to do with the function of category systems and asserts that the task of category systems is to provide maximum information

> with the least cognitive effort; the second has to do with the structure of the information so provided, and asserts that the perceived world comes as structured information rather than as arbitrary or unpredictable attributes. Thus maximum information with least cognitive effort is achieved if categories map the perceived world structure as closely as possible." (Rosch, 1978 p312).

From these two principles, Rosch infers more information about the likely nature of categories. Firstly, that some levels of hierarchy will, in general, be more useful than others, since they better mirror the perceived world, secondly that, since some aspects of our experience are more frequent or more pertinent than others, there should be some representation of these variations in frequency within categories, via some form of prototypicality.

Similarly, a form of introspection led Wittgenstein (1953) to the conclusion that concepts are not bounded, at least not in a simple way,

> "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 ? - Don't say "There MUST be something common or they would not be called 'games" but LOOK AND SEE whether there is anything common to all. -For if you look at them you will see not something that is common to ALL, but similarities, relationships, and a whole series of them at that. To repeat: don't think, but look !" (Wittgenstein, 1953, p31)

Wittgenstein goes on to use the expression "family resemblances" to characterise these similarities. So theoretical analysis of how and why we use concepts has informed us to some extent of the kind of things we are looking at. However, a set of guidelines as to what processes are looked at in experimental work is also required. The following is from Medin and Smith, (1984).

> "The following four functions (after Rey, 1984) seem particularly important:
>
> 1. SIMPLE CATEGORIZATION: the means by which people decide whether or not something belongs to a simple class (e.g. deciding that a particular object is an instance of the concept BOY).
>
> 2. COMPLEX CATEGORIZATION: the means by which people decide whether or not something belongs to a complex class (e.g. deciding that a particular object is an instance of the concept RICH BOY).
>
> 3. LINGUISTIC MEANING: that part of the meaning of a term that explains relations of synonymy, antinomy, and semantic implication (e.g. that part of the meaning of "boy" that explains why it is roughly synonymous to "lad" and implies being male and young).
>
> 4. COMPONENTS OF COGNITIVE STATES: the critical components of beliefs, preferences, and other cognitive states; in this role, concepts are what provide a cognitive explanation of complex thought and behavior (e.g. the roles played by the concepts RICH, BOYS and SPOILED in someone's belief that rich boys are spoiled)." (Medin and Smith, 1984, p114),

In considering what linguistic entities are counted as concepts a comment from the same paper, but attributed to Miller (G.A. Miller, 1982, personal communication), is also worth noting, "... *those interested in categorization think that concept is spelled "N.O.U.N"*" (Medin and Smith, 1984, p.122). Although this comment ignores some work (e.g. Miller and Johnson-Laird, 1976), and is less true now than when first published, it is still a broad truth. It is probably the case that nouns are seen as "safe" examples of concepts, and so are used by researchers most commonly. This thesis also investigates noun concepts. It should be recognised though, that this weight of attention is hard to justify on theoretical grounds, and may lead to an unbalanced representation of concepts, with those concepts related to actions or events, for example, being largely ignored by cognitive psychology.

Theories Of Concepts

In the study of concepts, as with many areas of cognitive psychology, it is very difficult to divorce the findings of research from the theoretical framework of the researcher concerned. This is in part due to the range of techniques used in obtaining data. I shall therefore discuss these three aspects of the literature as an integrated whole, as their interdependence demands. In doing so I adhere to the conventional technique for such a discussion, and subdivide the topic into types of theories.

Unfortunately there appears to be no consensus as to the best way to carry out such a division. There is a so called **classical** view, which is taken as a starting point, and with which other theories are contrasted. This classical view assumes that concepts are defined by lists of features. Other theories reflect the probabilistic nature of categorisation. Amongst classifications, these theories have been lumped together as "prototype" theories (Costello & Keane 1992), divided into "defining- and characteristic-attribute theories" and "characteristic-attribute theories" (Eysenck & Keane, 1990) or into "probabilistic" and "exemplar" theories (Medin & Smith, 1984). I shall first describe the classical view, and then the prototype and exemplar views, both of which are probabilistic in nature.

There is another dimension along which a useful division of theories can be imposed, particularly relevant to recent work - this is the extent to which general world knowledge is held to inform categorical judgements. More traditional views have tended to assume that the intension of a concept is finite, if perhaps probabilistically based. General knowledge, or "theory" based views claim that our extensive theories of the world at large have a major influence. However, the evidence on which such theories are founded comes largely from studies of conceptual combination, so I propose to deal with theory-based views when I discuss the specific issues of complex concepts. In the present discussion, I shall divide the theories into two main sections, classical and probabilistic, with the latter section being further sub-divided into prototype and exemplar views.

The Classical View

The central claim of the classical view is that an item **X**, is a member of concept **Y** if it complies with certain, defined rules. In other words, the intension of a concept is a list of attributes, and that category membership is an "all-or-nothing" affair. As will be immediately apparent, this view entails that concepts should behave in a straight-forward, simple fashion, but this is not in fact found to be the case. As seen from the work of Wittgenstein cited above, concepts, e.g. games, are not characterised by a set of necessary and sufficient features, but are better described as being comprised of items which bear a family resemblance, one to another (Wittgenstein, 1953, or Rosch & Mervis, 1975 for an experimental account). This is a severe failing, perhaps the most severe, of the classical view. Related to this problem are others. There is, for example, no room in such a theory for unclear cases of category membership. Yet the psychological reality is that there is often considerable doubt as to whether or not a particular item belongs in a category: Is a rug furniture? It is even the case that a single subject will give different answers on different occasions (McCloskey & Glucksberg, 1978).

Another major stumbling block is evidence of a typicality range within concepts. According to the classical view an item cannot be a more or less typical member of a class - it is simply a member or not. This is in stark contrast to the actual state of affairs, in which even the concepts most likely to conform to such a principle (e.g. "odd numbers") show typicality effects (Armstrong et al, 1983).

The classical view cannot stand up to the weight of evidence against it. It is still often referred to in the literature though. This is in part because of its intuitive appeal - it is perhaps the view most similar to the folk psychological theories that people have of concepts. The issue of how people believe concepts behave can be important. This possibility also accounts for the fact that it is the view from which later theories are derived, arising from the evidence mounted against it. Finally, it is not yet clear that the classical view has no role to play. It may be significant as a process of categorisation when other techniques are inhibited, or with respect to specific kinds of concept.

Probabilistic Views : prototypes and exemplars

A fundamentally different alternative to the classical view is required if the phenomena recorded from studies of concepts are to be explained. The probabilistic view rejects the notion that any criteria can be laid down as absolutely determining category membership. Rather, membership of a category is determined by the similarity between the item in question, and a representation of the concept. This representation may take the form of a prototype, or of previously encountered category members. These two cases are the **prototype** and **exemplar** views respectively. In both cases the representation conveys typicality, that is, information of a probabilistic nature.

Prototype View

This view contends that a concept is represented by some prototypical entity. The major proponent for this view of concepts has been Rosch, who with other researchers in the 1970's, found much evidence in favour of prototypes.

Rosch (1978) cites research using a range of psychological variables, all of which implies the importance of prototypes :-

Speed of Processing : Reaction Time : people have faster categorisation responses for items rated as more typical members of the category concerned than for less typical members. Rosch et al (1976) demonstrated this even with artificial categories devised in an experiment which controlled factors such as frequency.

Speed of Learning : Rosch et al (1976) found that prototypicality of stimuli predicted speed of learning artificial categories. Anglin (1976) and Rosch (1973) found evidence that children learn good examples of categories before less typical ones.

Order and probability of item output : Battig and Montague (1969) asked subjects to list the members of various superordinate categories. The order in which they are listed correlates with prototypicality. Rosch et al (1976) found that both order and frequency of output correlated well with rated typicality of artificial categories.

Priming : The typicality of items in a category determines how strongly the category name will prime that item. This has been shown across several category types (Rosch, 1975b&c; Rosch et al, 1976b).

Logic of Natural Language use of Category Terms : Various aspects of natural language use suggest that categories are structured by typicality. For example, the acceptability of hedges, such as "virtually", or "technically" correlates with typicality (Lakoff, 1972; Rosch, 1975a).

This body of evidence, however should not be seen as pointing to a particular model. In the words of Rosch (1978, p318), "To speak of a prototype at all is simply a convenient grammatical fiction; what is really referred to are judgements of degree of prototypicality." Prototypicality should be seen as a constraint, which a model of categorisation should satisfy, rather than a theory in its own right. These arguments for prototypicality are not to be confused with the suggestion that a concept is defined by similarity to, for example, an average template image, they might equally well support, for example, propositional or structural representations (Rosch, 1978). Clearly, the notion of prototypicality does away with many of the problems associated with the classical view. There is no need to search for sufficient and necessary characteristics, and there is a built in assumption that some features are more salient or important than others. There is also the important difference of probabilistic rather than definite class inclusion, since the category boundaries are not strict, but fuzzy.

Two principle criticisms are levelled at the prototype approach. The first concerns the validity of inferring a prototypicality structure of representation from evidence of apparent prototypicality effects in the output, or treatment of information. The second relates to the loss of useful information involved in the formation of a prototype. The first of these arguments was made by Armstrong, Gleitman and Gleitman (1983), who demonstrated just the same prototypicality effects in operation when dealing with concepts specifically chosen for their obvious non-probabilistic nature. That is, concepts like "odd number" which have a clear "all or nothing" definition do exhibit typicality effects. Armstrong et al argue that if concepts we know to be non-prototypical exhibit typicality effects, we are not entitled to infer prototypicality of structure for any other concepts from this sort of evidence. This argument might constitute a real problem for proponents of a prototype theory. However, there is still work that needs to be done to clarify the status of the claim that "non-prototypical concepts show typicality effects". On closer examination, all Armstrong et al have shown is that categories which can be classically defined, can also show typicality effects. It is possible that, even though subjects are aware that there are definite criteria for judging "oddness" of numbers, they do not normally represent the concept in such terms. In other words, the assumption that classically defined concepts are necessarily classically represented is itself unfounded.

It seems that subjects employ a probabilistic judgement heuristic to this situation. Whether this is seen as implying that such procedures mask the structure of all concepts in a veil of typicality effects, remains to be seen. An alternative explanation can be sought by reference to a possible two tier representation. This might envisage an explicit, classically defined category "odd numbers", in parallel with an implicit probabilistic representation or identification procedure, which would manifest itself via a graded value of oddness of numbers. There appears to be scope for empirical investigation of this possibility. It is likely that, if there is such a two tier system responsible for these effects, then each component could be individually inhibited. If for example the typicality effects are the result of a probabilistic identification mechanism, this might be bypassed by explicitly supplying information which is absolutely diagnostic of category membership (for example, subjects may be encouraged to attend **only** to the final digit of numbers, and use this to judge whether the number is odd or even).

The second major criticism is that the suggestion of a prototype inevitably condenses the information to a large extent. This is considered unsatisfactory, since there appear to be items of relational information concerning the internal structure of categories to which we have access, but which wouldn't have been stored in a prototype. The best example of this concerns complex concepts derived from simple concepts. The information contained in such complexes could not be derived form simple prototypes (Osherson and Smith, 1981). This problem is addressed in more detail in the discussion of complex concepts below.

Exemplar View

The exemplar view asserts that a concept, rather than being represented by an abstraction of its elements (as in the prototype view), is simply represented by the instances of the concept themselves. It retains the probabilistic nature of the prototype view by determining categorisation on the basis of similarity to existing category members. The two major arguments in favour of this view are firstly that some useful (and apparently used) information, (e.g. about the range of elements within a concept) is discarded by prototype formation, and secondly that such a system is unjustified unless it can be shown to be necessary.

There are however, inevitable problems with the idea that it is simpler to allow the instances of concepts to represent themselves than to construct a representation that is abstracted from the instances. It would require that apparently simple superordinate categories have a huge definition, e.g. consider the representation of "things" under this system. This is surely perverting principles of cognitive economy. It is also founded on the assumption that abstracting features from an element somehow excludes the specific attributes of that element from being accessed. Obviously such an outcome would be intolerable, but it is not clear that it is a necessary implication of prototype systems, in the generous sense of Rosch (1978), who writes, "For natural-language categories, to speak of a single entity that is the prototype is either a gross misunderstanding of the empirical data or a covert theory of mental representation." (Rosch, 1978, p318). Finally, proposing such an exemplar view is rather confusing theoretically. If it is the extensional exemplars themselves which define concepts, then it is not a psychological model. On the other hand, if it is the case that our mental representations of exemplars define our concepts, then the theory is rather empty in that it still requires that some suitable model be found for representing the exemplars themselves.

Assessment Of The Available Theories

In the words of Medin and Smith, in a review of this area, "The present state of affairs is less than satisfying." (Smith and Medin, 1984, p119). They also identified a number of questions which the extant literature failed to address satisfactorily. Two of these questions in particular have received considerable amount of attention in the intervening years. These are the questions of concept combination, and concept coherence. Although some models of conceptual representation have been specifically designed to address both questions, and nearly all make some reference to both, this thesis concentrates on the topic of concept combination.

Are all Types of Concepts the Same?

Complicating the central question of what type of representation concepts might have, is the issue of whether or not different types of concepts might have different forms of representation. It is possible that there are in fact important differences between ontological categories of concepts. The most common distinction drawn is between artefacts (such as tools, clothes, furniture etc.) and natural kinds (such as plants, animals, chemicals etc.). Other categories such as emotions (Wierzbicka, 1994), speech sounds (Barsalou, 1992) and nominal kinds (Keil, 1989) could also be considered distinct, but since the natural kind / artefact distinction is most well researched, I concentrate on those categories here.

Philosophical Evidence

The origins for supposing such a distinction are in philosophy, principally from Kripke (1972a) and Putnam (1975). Putnam noted that for natural kinds such as **lemon** and **water**, it is not possible to specify a list of criteria which absolutely rules any item in or out of that category. This led him to the conclusion that concepts are not formed on the basis of features, but in the belief that there is some underlying essence. The role of surface features (such as colour, taste and texture) in categorisation is as indicators to what the underlying essence of an item may be. Quite what might constitute such an essence did not concern

Putnam, although he suggests such notions as chromosomal and subatomic structure may be the right sort of thing.[1]

Putnam later generalised this essentialist theory beyond natural kind concepts.

> "It follows that "pencil" is not synonymous with a loose description. When we use the word "pencil", we intend to refer to whatever has the same nature as the normal examples of the local pencils in the actual world" (Putnam, 1975, p. 162).

As an illustration, Putnam describes a thought experiment in which microscopic examination of pencils reveals that they are in fact not manmade, but are biological organisms. The critical point is that upon this discovery we are likely to still claim that what we previously believed to be pencils are still pencils, and that we have simply changed our belief about what attributes a pencil has. In doing so we are revealing our belief that pencils have an underlying essence.[2]

However, Schwartz (1978) points out that when confronted with single items which contradict our beliefs, we behave differently with respect to artefacts and natural kinds. If we discover that our pet cat is actually robotic, we alter our beliefs about that individual to exclude it from the category "cats". The case of artefacts is different. It is possible (though highly improbable) that a pencil could be formed by a bolt of lightening happening to fuse various elements in an

---

[1] Stressing that essences are a matter of belief rather than metaphysical fact, Medin and Ortony have used the term "psychological essentialism" for a version of this theory (Medin and Ortony, 1989).

[2] As thought experiments go, this one does rather tax the imagination. However, an informal but large scale `experiment' can be interpreted similarly. An edition of the television programme Panorama (April 1st, 1957) chronicled the harvest of spaghetti from the large bushes on which it was said to grow. That this counterfactual biological origin for an artefact was widely accepted at the time (Dimbleby, 1975) illustrates people's readiness to accept such findings as plausible.

appropriate structure. Although this would certainly not be manmade, we would still believe it to be a pencil. It appears that there is therefore a basis for a theoretical distinction between artefacts and natural kinds.

Cognitive Psychological Evidence

Empirical evidence on these questions is mixed. Malt (1992) found (in agreement with Schwartz's thought experiment) that intended function was neither necessary nor sufficient for membership of artefact categories. In a later study Malt (1994) looked at the case of water being defined by the chemical formula H_2O, with ambiguous results in several different studies. In one study, subjects were asked to judge the acceptability of sentences about substances known as "water" but not composed entirely of H_2O, such as puddle water, or mineral water. Sentences were like (i) "puddle water is a type of H_2O", and (ii) "puddle water is mostly but not entirely H_2O". Sentence (i) contradicts the essentialist view, since puddle water is not itself H_2O, but some kind of mixture, whereas sentence (ii) is consistent with essentialism. Since subjects found both statements about various kinds of water acceptable, Malt concluded that we can use both an essentialist sense and a nonessentialist sense of water. Similarly mixed findings were obtained by Braisby et al (1996). They empirically tested people's judgements about natural kind concepts as influenced by new "discoveries" which might be counter to supposed essential characteristics. Responses were found to vary in their consistency with essentialism, depending upon the type of discovery presented.[3]

In contrast, Barton and Komatsu (1989) found that for most of their subjects both artefacts and natural kinds have some essential characteristics. Subjects were

---

[3] Braisby et al reject essentialism on this evidence, on the grounds that essentialism itself entails an invariance under changes of context.

asked questions of the form "Would a pencil still be a pencil if it were like a pencil in all ways except that....?" where the continuation of the question referred to a change in either chromosomal/molecular structure, function, or a physical feature. For artefacts, only the appropriate function was necessary, and for natural kinds only the chromosomal/molecular structure was necessary.

The relevant cognitive psychological evidence is therefore balanced, with various researchers claiming that essentialism holds for both artefacts and natural kinds, or just natural kinds, or neither. The behaviour of subjects seems quite dependent upon the specifics of the task.

Other Evidence

Other studies have also produced evidence that artefact and natural kind concepts are significantly different in their mental representation. Wierzbicka (1994) points out that natural kinds are organised hierarchically (e.g. creature - bird - sparrow) whereas artefacts are not. That superordinate natural kinds are count nouns (e.g. 3 birds, 3 plants), while superordinate artefacts are not (*3 furnitures, *3 clothings) is indicative of such a difference.

There are also several accounts of neurological injury resulting in cognitive deficits relating specifically to one category, while leaving the other intact. Caramazza et al (1994) catalogue a number of such cases.

Overall, there seems good evidence for supposing that there are important and interesting differences between artefacts and natural kinds. Whether or not this consists of a fundamental difference in essentialist terms is a moot point, but there would appear to be ample evidence to suggest that artefact and natural kind concepts will behave differently in some circumstances.

Models Of Conceptual Combination

The Significance Of Complex Concepts

People are very good at combining concepts. Not only do we readily understand complex concepts that we may have heard before, but there is often little difficulty in understanding entirely novel combinations. However, as a topic of research, it is first necessary to decide what qualifies as a complex concept. This entails all the problems referred to earlier relating to simple concepts, and more. For example, concepts can be linguistically combined in more or less explicit ways, e.g. the form "pets which are also fish" may differ from "pet fish". The literature appears to be in general accord with the notion that any non-idiomatic phrase comprised of two or more lexical units describes a complex concept.[4] There are also inevitable problems with mixing grammatically different word types. The more difficult cases, such as verbs, tend not to have been addressed, but work has been carried out on adjective-noun, adverb-adjective-noun, and noun-noun combinations. Apart from the need to understand complex concepts for their own sake, there are related areas of language use which may be understood by the same process. For example, simile and metaphor both seem to require some complexing of two concepts to derive both their literal and non-literal meanings. There is also the promise of understanding something of the nature of simple concepts, from looking at their complexes.

How Theories Relate To Complex Concepts

The various theories of conceptual combination discussed below each make there own predictions about the relationship between a complex concept and its component parts. In discussing these, a number of important findings on the

---

[4] Although such a definition is useful, it is worth bearing in mind that it may ultimately be of little use, since how we define terms such as "concept" and "lexical unit" remains unclear.

behaviour of complex concepts, often in contradiction to the proposed models, will be introduced.

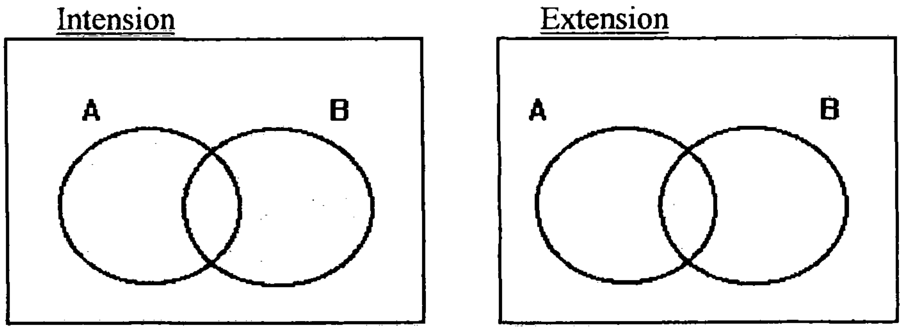
Intensions And Extensions

The relationship between a complex concept and its constituents has often been expressed in terms of set theory, which is a useful way of representing such combinations. Before looking at theories of concept combination, it is therefore important to appreciate the relationship between the intension and the extension in terms of the intersection and union of sets. Set theoretic descriptions of concept combination under different models of concepts are therefore given here.

Classical View

Classically, the union of the intensions ($A_i \cup B_i$) of two categories corresponds with the intersection of the extensions ($A_e \cap B_e$), as shown in fig 1.1. That is, things which have all the characteristics of **A** and all the characteristics of **B** are members of both **A** and **B**.

Fig 1.1 : The Relationship between the Union of the Intension and Extension in the Classical View



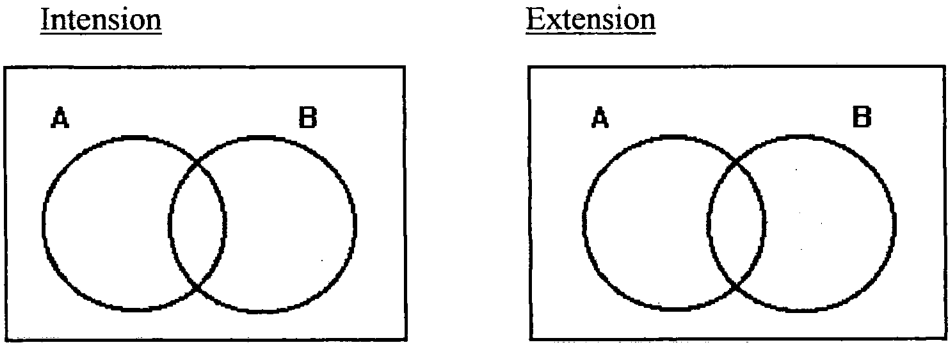
Prototype View

The prototype view provides a less straight-forward relationship, the nature of which is determined by the criteria relating intension to extension used in a particular prototype model. Assuming class-inclusion to be determined by a function of similarity between intension (i.e., the prototype) and the item being categorised exceeding some threshold, we can predict a relationship of the form shown in fig 1.1, that is, identical to the classical view. However, it may be that the threshold is variable, perhaps decreasing when more features are specified, this would introduce a psychologically realistic unpredictability into the situation, allowing inclusion of some items in $A_e \cap B_e$, but not in $A_i \cup B_i$.

Exemplar View

The correspondence in the exemplar view is more straight forward, since there is no logical difference between the extension and intension. The union of the intensions ($A_i \cup B_i$) therefore corresponds with the union of the extensions ($A_e \cup B_e$), as shown in fig 1.2.

Fig 1.2 : The Relationship between the Union of the Intension and Extension in the Exemplar View



Theories of Concept Combination

Having seen how views of concepts differ with respect to concept combination in terms of set theory, I now examine how these differences relate to empirical situations.

Classical View

This tradition predicts that the extensional complex of two concepts will be the conjunction of the sets of members of the two concepts. An example to illustrate this is as follows: the combined concept "orphan girl" contains all those individuals who are both girls and orphans. For this example, the theory works well, but this is the exception rather than the rule. One could think of countless examples where such an prediction fails, consider "mechanical engineer", "chair leg" or "coffee cup". The theory therefore demands very little by way of refutation. It is however useful in what it illustrates. It becomes very clear that different combinations behave as different grammatical forms. This is an observation to which subsequent theories ought perhaps pay more attention than they do, as will be discussed below.

The Probabilistic Views : Prototypes and Exemplars

In order that the probabilistic nature of human conceptualisation can be accurately modelled, researchers have employed fuzzy set theory, in which each item belongs to a set with a certain probability, a formulation of the prototype view of concepts.. Fuzzy sets, just like well defined ones, can be combined. Since fuzzy sets contain information relating to the typicality of a member to the set, it is possible to convey this information to the combined concept too. The precise description of the process is open to debate, but the **min rule** (Zadeh, 1965, 1982) is the most commonly adopted principle. This rule considers the probabilities of a particular item being a member of the individual constituent concepts. It takes the lowest of these to be the probability of that item being a member of the complex concept. There are alternatives to this, e.g. taking the product of the individual probabilities to be the probability of membership of the complex (Costello & Keane, 1992). However it will be seen that by any such process, the probability of being a member of a complex cannot exceed that of being a member of any simple constituent, i.e. The 'min' rule is the most generous formulation possible.

This prediction has caused the prototype view perhaps its biggest problems. Osherson & Smith (1981), demonstrated that in some instances an item is a far more typical member of the complex than of either constituent. E.g. the **guppy** is a relatively atypical example of a pet, and it is not a very typical fish either, but it is highly typical of the concept **pet fish**. Thus the guppy phenomenon proves very awkward for this framework. The exemplar view is not subject to the same problem. Under the exemplar view, typicality is not inherited from constituent concepts, but simply results from similarity to other concept exemplars. In this case, the exemplar view would correctly predict that a guppy is a typical pet fish. However, the exemplar view also has severe problems specifically related to concept combination. For example, its lack of abstraction inhibits creative potential in combining concepts.

Recent Models of Conceptual Combination

There are severe problems with the approaches described above. Not least being the failure to address the issues in specific detailed terms. It has become clear that to predict the nature of a complex concept from its simple constituents one requires more than a manipulation of the extensions. The next generation of models have therefore begun to describe what may be happening in terms of the intensions of concepts.

Concept Specialisation Model

The first such theory to appear has become known as the Concept Specialisation Model (Cohen and Murphy, 1984; Murphy 1988). Importantly, this deals not in extensions, but in intensions. In doing so, some of the problems discussed above are avoided. In making this switch of focus, different ways of combining concepts arise. It is no longer necessary to be confined to the rules of set theory, and the mechanical combinations that they imply. Intensions are meanings, and as such they have complex relations with one another. Combining intensions implies a whole new range of diverse structural relations.

According to this model, we combine concepts by finding some mediating relation between them. Clearly, this could be any of innumerable relations, and it could be more salient to one of the constituent concepts than the other. In adjective noun combinations for example, the mediating relation is often quite straightforward, and is indicated pretty well by the adjective. For a combination like **blue chair**, **colour** is the obvious mediating relation. For noun noun compounds things are often more complicated. For a combination like **house boat**, an appropriate mediating relation is found in **function**, thus it is a boat functioning as a house.

Finding such mediating relations might be a question of comparing concept intensions, but in other cases we may need to refer to things beyond the constituent concepts themselves. The other novel aspect of this model therefore is the inclusion of world knowledge. Prior experience and our own theories about the world therefore influence the way we combine concepts. As with the switch to manipulating intensions, this facility to allow general knowledge to influence proceedings provides a rich source of possibilities, and phenomena not previously explained can now be accounted for. Using world knowledge we can make predictions about what sort of novel characteristics we might expect a combined concept to exhibit. For example, an **apartment dog** (mediating relation = **lives in**) will probably be relatively small and placid compared to other dogs. However, these gains in power attained by using intensional relations and world knowledge come at a price. That is that the model is much under defined. There is no explicitly defined mechanism for how one mediating relation is preferred to any other. Without spelling out how we make such a selection the model fails to explain the mechanism upon which it is founded. As such, although it succeeds in accounting for the range of phenomena reported of concept combination, there is no way that it could be used in a predictive capacity. The description of the mechanism is simply too imprecise to be thought of as a satisfactory model of concept combination.

Setting aside these drawbacks, the theories upon which the model is based point the way for more limited but also more explicit models. It is no coincidence that two of the more successful models, although markedly different in their detail, exhibit a similar general approach. The two models referred to are the Selective Modification Model, of Smith, Osherson, Rips & Keane, (1988), and a Composite Prototype Model, of Hampton, (1987). Their differences are many and great, and will be evident in a more detailed look at each, but for now I shall highlight their significant similarities. Both models take the intension of a concept to be the critical focus of modification, breaking from the more straightforward and traditional approach which attended primarily to the extension. There is also a similarity in the extent to which both models are spelled out in detail. They are both "working" models, based on fairly straightforward, intuitive procedures, and their success in dealing with certain cases is demonstrable. The predictive scope in each case is strictly limited, but this concentration on specific aspects of conceptual combination may serve to isolate and characterise relatively distinct processes. At some future date perhaps such procedures may find their way into a more general framework.

Hampton's Composite Prototype Model, (1987)

Hampton has proposed a model of combining prototype concepts, by taking the union of the sets of features of the constituents as the basis of the new concept. A number of constraints of attribute inheritance are imposed on this process :-

i) The mean level of importance of an attribute across both parent concepts must be above a threshold level for inheritance.
ii) If an attribute is perceived as 'necessary' for either constituent, then it must be inherited.
iii) If an attribute is perceived as 'necessarily impossible' for either constituent, then it cannot be inherited.

The experimental work on which Hampton grounds these assumptions uses complex concepts of the form, "Games that are also sports". This model has proved surprisingly successful at deriving the character of these compounds. There are however several findings which lead to elaboration of the basic model. There is often an imbalance in the contribution of each constituent to the complex, and there are also so called `emergent properties`. Firstly, to consider the imbalance between constituent concepts. This dominance effect was predicted by Hampton (1988), and is thought to be a product simply of a relative imbalance of the 'size' of categories used (i.e., how many attributes each has). Further to this phenomenon is that of noncommutativity, i.e. :- "**X** that are also **Y**" is different from "**Y** that are also **X**". Hampton ascribes this to a predominance of concern with distinguishing the named set from its natural contrast set, i.e. distinguishing items in "**X** that are also **Y**" from those in "**X** that are not **Y**". There are however two other factors which may have influenced this outcome. Firstly, the constituents of all Hampton's pairs were concerned with similar domains, e.g. tools/weapons, sports/games. This may well concentrate attention on the contrast set. Secondly, the form of combination itself is somewhat unusual in such studies, and implies a different grammatical relationship to the more commonly used "**XY**" form. Indeed, this distinction between the two forms of presentation may be significant in other ways too. It is quite possible that the success Hampton has had in taking the union of sets as the critical starting point is related to his chosen form of presentation of complex concepts. Intuitively the form "X that are also Y" appears to indicate that to derive the category concerned one must take the extensional intersection of the two sets, whereas the form "XY" may imply many more involved relationships. This point does not devalue the findings, but does suggest that they may well relate to only a specific type of concept combination, since other combination types are effectively ruled out by the use of a such a precise form of presentation.

Returning to the presence of **emergent features**, a phenomenon that Hampton refers to as **noncompositionality**. There are features given as belonging to the combined concept, which are not accountable by reference to either of the constituents. Hampton actually acknowledges the possibility of an underrepresentation of noncompositionality in his data, again due to the form of presentation of combinations. Given that this effect may be even more extensive than demonstrated in this study, a full explanation of this phenomenon is even more important. Hampton identifies two processes as possibly contributing to this noncompositionality. The first is **extensional feedback**, this operates by the meaning of a concept being derived as described in the model above, and subsequent identification of the concept's extension. This extension is likely to have some attributes that are not specified in its derived intension, these may then be added (fedback) to the intension, to produce a more complete picture. The second process is one of inference in order to maintain coherence of the concept. Hampton cites the example that **pet fish** are probably smaller than average pets or average fish. This could derive from extensional feedback, but it is also the case that we could infer that for a fish to be a pet it must be able to live in a small body of water, and therefore be small itself. It is evident that this process doesn't require any experience of the extension of the concept to operate, and that conversely, extensional feedback does.

Another possible influence which might be added to these two is that of conventional or idiomatic effects. Some combinations are likely to have particular properties associated with them for cultural or linguistic reasons. For example, certain characteristics are commonly associated with **car salesmen** or **traffic wardens** whether or not these agree with our own experiences. Consideration of such possibilities is likely to be required if the noncompositionality of all complex concepts is to be explained. But these two processes are now initially defined, and worthy of further investigation.

The Selective Modification Model of Smith et al (1988)

In contrast to that of Hampton, Smith et al's model concerns complex concepts formed by the combination of adjectives and nouns, the primary example being **red apple**. The process of 'selective modification' used follows quite readily from the prototype structure which forms the basis of their model. The model assumes that a concept is prototypically defined by values attributed to possible options in a number of 'dimensions'. E.g. for **apple**, the dimensions might include **colour, shape**, and **texture**, the possible options for colour being **red, green, brown** etc. The values (referred to as 'votes') attributed to these features act as defaults, so if most apples are thought to be red, then 'red' will have more votes than any other option in the dimension of colour. There is also a 'diagnosticity' value for each dimension in the representation, giving an indication of how pertinent that dimension is in general to the concept. This is used in determining whether or not some object is a member of the category. The typicality of any item to a category is determined by its 'similarity' to the prototype. This notion of similarity is adapted from Tversky's (1977) **contrast rule**. It amounts to subtracting the number of votes weighted on different features from the number weighted on the same features, each element having been modified in accordance with its diagnosticity. The modification of a noun by an adjective is therefore represented as modification of the prototype by the following two processes. First, the adjective causes a shift of votes from the alternatives in its own 'dimension' to itself, and second, the diagnosticity of the dimension concerned rises. This means that a **red fruit** is now definitely red as oppose to any other colour, and that the dimension of colour is more important to **red fruit** than it is to fruit in general.

To test this model Smith et al obtained subject's feature listings of several fruit and vegetable concepts (e.g. apple, potato). From these listings they derived the information such as attributes and diagnosticity needed to characterise these concepts and their superordinates (fruit and vegetable) in their model. The model was then used to produce modified representations of these superordinates by combining them with several adjectives, producing representations for such combinations as **red fruit**. Having produced these representations it is possible to predict how typical any particular instance (e.g. apple) ought to be of a particular combined category (e.g. red fruit). In general, these predictions correlate well with subjects' judgements of the same questions.

Smith et al also proposed that the model could be extended to include adverb-adjective-noun complexes, like **very red apple**. The process is fairly straightforward, consisting of multiplication of the votes for the specified adjective by a scalar in an appropriate range, i.e. (in a simplified form):

'very' = multiply votes by scalar k, where $k > 1$

'slightly' = multiply votes by scalar k, where $0 < k < 1$

'not (or non-)' = multiply votes by scalar k, where $k \leq 0$.

Even this was extended still further, to include complexes of the form, **very slightly red apple**. However, the performance level of the model was reduced by the addition of each of these stages, and some of the adverbs ('very'), were considerably better modelled than others ('slightly').

On the whole, the model has been relatively successful, but only in a very limited field of operation. There are many essential aspects of 'literal' natural language that this model expressly fails to capture. Perhaps the most important to address is the extended influence of adjectives. Those used in the testing of the model are about as focused in their influence as possible, e.g., 'red' and 'smooth' are quite well defined[5]. There are many examples in which this is not the case, Smith et al consider 'shrivelled', but what of an 'old' apple, or a 'cooked' potato? Smith et al also provide a catalogue (from Clark, R, 1970) of 'non-standard' adjectives, which the model seems wholly inadequate to cope with, comprising; "negaters (e.g. 'fake'), enlargers (e.g. 'possible'), fictionalizers (e.g. 'mythical'), defictionalizers (e.g. 'simulated'), and neutralizers (e.g. 'alleged')". In a similar fashion to Hampton's model, the model is best viewed as an attempt to characterise the nature of one process of combining concepts. It is strictly limited in its application, due to the influence of other, as yet unclear, processes, but the model is useful as a working description of one part of conceptual combination.

Correlated Attributes and Emergent Features

---

[5] Apparently straightforward adjectives such as **red** can actually be quite ambiguous. For example, in the following cases, the adjective **red** implies several different modifications, none of which has the effect of asserting that the reference is coloured red: **red pen, red cone cell, red army, red flag** (socialist anthem), **red indian, red sea, red tape**. Some of these are certainly idiomatic, but others are legitimate, productive meanings.

The two models described above are concerned with different parts of language, and treat them in different ways. I have suggested that, in terms of their methodological approach, they have some similarities, but there are also some areas of theory which they share. Both recognise the problematic phenomenon of information being present in the complex representation, which does not appear in the constituents. Both models limit the evidence of this with their choice of materials, or format of presentation, as the authors themselves acknowledge. Smith et al are concerned with the mechanics of how an adjective might modify a schematic representation, and so assume that some adjective could operate on several dimensions of the representation. They suppose their model to be adaptable to cope with this phenomenon, but other research (e.g. Medin & Shoben, 1988) implies that even apparently "one-dimensional" adjectives can affect other attributes. For example, metal spoons tend to be small, wooden spoons tend to be large. Also, the correlation of attributes is something quite specific to the concept concerned, e.g. white and grey are more similar than grey and black with respect to hair, but the reverse is true with respect to clouds (Medin & Shoben, 1988).

Shoben (1991) identifies three processes that may lead to the correlation of attributes:

i) Inherent correlation : e.g. changing the material of construction from lead to aluminium will make the object lighter.

ii) People's beliefs : e.g. something made in Japan may be thought to be of better quality than something from the Philippines.

iii) World knowledge : e.g. size of a tropical fish is associated with a whole range of behavioural, physical and biological properties.

In terms of psychological significance it may be more pertinent to consider a simpler binary division of types of attribute correlation, between necessary and non-necessary correlations. Recall that in Hampton's model, whether an attribute was considered necessary or not for a constituent concept could be critical in determining the attributes of the combination. As such necessity of attribute correlation is likely to be equally important. Our beliefs about the necessity or otherwise of the correlation of attributes will be largely determined by experience. For example, if we witness any counter example to the correlation, then it can be assumed to be non-necessary. Conversely, if we witness what we believe to be a large and representative sample of the population, without coming across a counter example, we assume the correlation to be necessary.

There are also other reasons for assuming a correlation to be necessary or not, pertaining to our theoretical beliefs. If we believe that one attribute necessarily **causes** the other, or if we believe that some other factor is the necessary originator of both attributes we assume the correlation to be necessary. Given this binary division between two forms of correlation (necessary, and non-necessary), it is possible to see a relationship between correlated attributes and emergent features. The process of extensional feedback is identified with non-necessary correlations, while maintenance of coherence inferences are more akin to necessary correlations. So, by this analysis of the situation the problem in Smith et al's model, of accounting for correlated attributes, and the problem of noncompositionality in Hampton's model are closely related.

Noun Noun Compounds

In recent years, the case of noun noun compounds as a special type of conceptual combination has received increasing attention. Smith et al and Hampton focused on particular types of conceptual combination, as described above, in which the relation of one concept to the other was quite restricted. Adjectives tend to operate along quite well defined dimensions of meaning, and Hampton used syntax to specify the combination type. (In fact, as discussed above, in both cases the combinations did have influences beyond what superficially appeared to be their limits.)

Noun noun compounds by contrast clearly encode a high degree ambiguity, which obviously creates extra problems. This inherent ambiguity has been addressed in different ways by different researchers. The basic problem is, given a highly ambiguous compound, how do we settle on one (or even two or three) of the possible interpretations as being the most likely intended meaning. Certainly context has a very important role to play, an assumption which has not really been questioned. Undoubtedly, given a context in which some item is highly salient, a compound will be interpreted accordingly (e.g. Gerrig and Murphy 1992). Often compounds are coined to refer to such specific items. Downing (1975) gives an example in which a colleague refers to a particular student as "the bike girl", because she was seen to leave her bike in the vestibule. These uses are often referred to as **deictic** compounds. Little research has been carried out on such compounds, but clearly context is sometimes strong enough to enable coinages to be made and understood regardless of other influences. Most research however has been conducted on compounds in little or no obvious context.

Several aspects of noun noun compounds have been the subject of research in recent years, and are discussed below. Specifically they concern the relationship between lexicalised and novel compounds and the respective influence of each constituent in determining the compound relation, and the distinction between predicating and nonpredicating compound types.

Lexicalised and Novel Compounds

Some compounds are lexicalised and so are familiar to us and have conventionally accepted meanings, such as "phone box" or "computer disk". Others are novel and newly coined, like "frog ladder" or "robin snake". How these two types of compound contrast with respect to factors like compositionality and cognitive accessibility is important in understanding how we combine concepts generally. However, as well as theoretical comparisons between the two being of interest, it is possible that the relationship between lexicalised and novel compounds is active in the very way we use compounds, as in the theory that novel compounds are interpreted by analogy to their lexicalised counterparts. Theories of analogy are discussed below. Because the empirical research is closely related, included in this is a discussion of the relative influence the two positions of a compound have in determining what relation the constituents bear to each other. This is followed by a comparison of the nature of lexicalised and novel compounds.

Analogical Models

Ryder (1992) looked at the possibility that novel compounds are interpreted by analogy to more familiar compounds. She suggests that when we encounter a novel compound we compare it to linguistic templates to which known compounds adhere. The novel compound is then interpreted accordingly. The linguistic templates, and so the analogy, can be at a range of abstraction, from the most abstract (taking noun2 as the head) to the most concrete (sharing a common head noun and semantically similar modifier). In support of her theory, she demonstrates that when asked to paraphrase novel compounds, subjects usually interpret them in accordance with analogous lexicalised compounds.

The theory certainly has appeal from several perspectives. It seems to be coherent with phenomena from other linguistic domains. Labov (1972) reports that language change in general can occur by analogical processes. Also, there are several contemporary coinages which would seem to rely upon analogy to derive their meaning. **Trolley rage** is a supermarket equivalent of road rage, while a **net potato** is a technologically enhanced couch potato.[6]

Ryder succeeds in experimentally demonstrating that interpretations of novel compounds are in accordance with those of lexicalised compounds. However, as she acknowledges, her results are also explicable without recourse to analogy, but relying solely upon semantic script/schema type information, such as those of Smith et al (1988) or Murphy (1984) described above.

> "For example, since wine generally is packaged in bottles, and since bottles participate in a schema that has a critical slot for what the bottle contains, it is not surprising that *wine bottle* should have the meaning 'a bottle that contains wine.'" (Ryder, 1992, p97).

By the same token, a novel combination of liquid + bottle (e.g. rain bottle, juice bottle) will likely have an analogous meaning, though we may derive it from our knowledge of liquids and bottles rather than direct analogy to familiar

---

[6] Interestingly perhaps, these are not direct analogies, e.g. road rage occurs travelling on roads, and trolley rage occurs in supermarket aisles. It seems that the rhetorical appeal of referring to trolley rage outweighs the need for a direct analogy.

compounds. Essentially the problem is that if there is a likely schematic relation between the two concepts, then there will probably be several lexicalised instantiations of analogous combinations. Separating the two factors is difficult.

Van Jaarsveld et al (1994) devised a new methodology to try to circumvent this problem, and test Ryder's model directly. They reasoned that if novel compounds are interpreted by analogy to lexicalised compounds, then the size of the set of lexicalised analogues will have an effect on the time it takes to arrive at an interpretation. The more possible analogues, the greater the search space, and the more time it should take. They also make the assumption that this difference will be exaggerated if the compound is less interpretable. This is because highly interpretable compounds are easily interpreted because a suitable analogue is easily found, at which point the search for analogue can be terminated. Low interpretable compounds require a full search of all analogues.

They asked subjects to judge whether compounds were lexicalised or novel, and measured the response time to see if any such interaction did occur. The results show that there is no interaction between the two factors, so they concluded that novel compounds are not interpreted by analogy. However, their reasoning is founded on several questionable assumptions. First, that the search is serial. If, as Ryder argues, analogy is characteristic of much linguistic interpretation, then it would seem likely that an efficient mechanism has arisen to cope with it. A serial search of possible analogues is grossly inefficient. For example if one encounters a novel compound like **fire pencil** there is little point in exhaustively searching through the numerous (20 or so, by Ryder's count) lexicalised compounds with fire as the modifier which are to do with combating fires (e.g. fire station, fire engine, fire hydrant etc.). It would be more realistic to propose that we can take this set of compounds as a whole, and accept or reject the linguistic template as appropriate all in one go. Also questionable is Van Jaarsveld et al's account of interpretability - that compounds are more interpretable if a suitable analogue is found at the start of a search. This would mean that a more appropriate analogue might exist, but is not encountered because the search has already been terminated upon finding the first tolerable match. It makes more sense to suppose that all searches are approximately equally extensive, and that interpretability is a product of how good the analogy is, rather than how many items were searched through.

Other researchers have used different techniques, and found evidence that analogy generally can be a significant influence, at least when there are only short time periods between presentation of the analogue and the novel compound. Shoben (1993) found that subjects' interpretations of novel compounds could be influenced by prior exposure to a context supporting one particular interpretation of a similar compound (same noun2, similar noun1). Subjects who read a passage which supported one interpretation (for example about a child who contracted a **cat rash** from contact with a pet) were more likely than control subjects to use that same relation to interpret a similar compound shortly afterwards (for example, to interpret **horse rash** as a rash caused by contact with a horse than as a rash located on a horse). Shoben interprets these results as a kind of semantic priming, and though it doesn't relate directly to Ryder's model, such a phenomenon could be the basis of compound interpretation by analogy.

Further evidence comes from Gerrig and Murphy (1992), who investigated the effect of context on compound interpretation. In the first of their experiments, they presented subjects with a context in four different conditions. They found that subjects' later comprehension of novel compounds was improved if the context supported a particular relation between the two constituents of the compound (explicit condition), or if a compound consisted of one of the constituents (the noun2) and a concept similar to the other constituent (noun1) (implicit condition). When no relation was supported by the context (control condition), subjects' comprehension time was slower. Their second experiment was similar, but this time there was a rather longer time interval (of around 5 - 10 minutes) between the presentation of the context and the test compound. In this study, it was found that the results of their experiment 1 were approximately replicated. It therefore seems that interpretation by analogy could occur over at least slightly longer time intervals. In their final study, Gerrig and Murphy used a similar design containing just explicit and neutral contexts and a novel compound. In this study test sentences contained compounds which were possible analogues for the compound in the context, but which didn't have any lexical overlap with the primed compound. Subjects were required to state how easy the test compound was to comprehend, and to paraphrase it. On both these measures, the explicit condition showed improved comprehension compared to the neutral condition. This is further evidence of some analogy based interpretation, even when neither constituent concept is directly primed. This would correspond to the more abstract levels of Ryder's linguistic templates.

The question of whether or not we do use analogy to familiar compounds is still an open one then. Shoben (1993) and Gerrig and Murphy (1992) have shown that, at least in the short term, interpretation by analogy to primed compounds or relations can occur. Van Jaarsveld et al (1994) have some evidence that a particular model of analogical interpretation derived from Ryder (1992) fails to hold up under experiment. Ryder's own empirical work (1992) was supportive of her more general model, although other schema type approaches would have made similar predictions.

Another relevant issue which arises from the various strands of this research is the degree of influence which each of the constituent nouns has over the interpretation of a compound. Both Van Jaarsveld et al and Shoben used compound pairs with a common noun2 and similar noun1. Gerrig and Murphy also predominantly used such compounds, but also found that they could get similar effects when both constituents were only similar.

In more recent research Gagné and Shoben (1995, unpublished manuscript) measured the time it took subjects to judge the meaningfulness of compounds to compare the influence of noun1 and noun2. They presented subjects with meaningful compounds, along with non-meaningful fillers. The experimental compounds were composed such that the relation between the constituents was either a frequent or an infrequent relation with respect to each of the constituents in other meaningful compounds. So compounds had relations that were High-High, High-Low and Low-High frequencies for noun1-noun2 (it was not possible to construct enough meaningful Low-Low compounds to include this condition, but that comparison is not critical to the experiment). The compounds appeared on a screen and subjects had to indicate whether they were meaningful or not. It was found that High-High and High-Low compounds were responded to equally quickly, but that Low-High compounds took longer. Gagné and Shoben interpret this as indicating that the relational information associated with noun1 (whether it is analogical or schematic) is of primary importance in interpreting compounds, and that such information associated with noun2 is less important.

Ryder (1992) uses the notion of **cue reliability** (from Bates and MacWhinney, 1987) to address the same sort of issue. Cue reliability, in this context, is the proportion of cases in which a constituent noun (the cue) occurs in a compound with a particular relation. From surveying lexicalised compounds in a dictionary, Ryder concludes that various constituent nouns exhibit more or less the full range of cue reliability. Some, such as **box** in the noun2 position are almost absolutely reliable (an **X box** is a **box** used for storing Xs). Others such as **man** in the noun2 position are somewhat reliable, having just a few possible relations, while others still, such as **board** in the noun2 position are very unreliable (e.g. **surf board, chalk board, head board** etc.)

The issue of what contribution each constituent makes to the relation expressed in a compound is clearly an important one. At present, from the research discussed above it is not possible to say if either position in a compound is generally more influential than the other.

Comparing Lexicalised and Novel Compounds

A number of important issues arise from a comparison of lexicalised and novel compounds. The most striking aspect of research into such comparisons is therefore its scarcity. Some authors, as described above, have looked at the possibility of novel compounds being interpreted by analogy to real compounds (e.g. Ryder, 1992; Van Jaarsveld et al, 1994), while others have looked at novel compounds only (e.g. Wisniewski and Gentner, 1991; Shoben, 1991). Studies involving a comparison of lexicalised and novel compounds are conspicuous by their absence. One related area which has received some attention is the study of idioms. Jackendoff (1995) highlights the similarity between understanding idioms, and understanding compounds, and claims that the same mental mechanisms are likely to underlie both processes. Certainly it is the case that several of the key questions relating to idioms are also directly relevant to compounds, and so the two are looked at together below, with particular reference to compositionality, flexibility, storage and access.

Compositionality

That idioms are noncompositional is often taken as a part of their definition. Swinney and Cutler (1979, p522) define idioms as "a string of two or more words for which meaning is not derived from the meaning of the individual words comprising the string". This definition would therefore classify many lexicalised compounds as idioms. Jackendoff (1995) points out that absolute noncompositionality is extremely rare even in idioms, and the overlap between idioms and lexicalised compounds is therefore all the greater. The difference in compositionality that exists between idioms and literal language seems similar to that that exists between lexicalised and novel compounds.

Flexibility

The degree and type of flexibility of idioms is a matter of current research (e.g. Gibbs, 1995). Some types of modification are acceptable (innumerable cats were let out of unnumbered bags), while others are not (*the bucket was kicked by him). This observation may also provide an insight into the way that novel compounds are coined and understood. It has been noted that novel compounds are interpreted in ways analogous to related lexicalised compounds, and it has been suggested that the processes of novel compound generation/interpretation are themselves analogical (e.g. Ryder, 1992). The idea that idioms (and so perhaps lexicalised compounds too) are flexible leads to the possibility that an explanation for this analogical data can be given in terms of modification of lexicalised compounds. That is, the modification of a lexicalised compound would create an analogous novel compound, e.g. the familiar compound **road rage** has a more recent counterpart in **trolley rage,** a supermarket equivalent apparently derived from it. Whether this derivation is best viewed as a process of analogy or modification is not clear.

Storage and Access

Issues of storage and access of lexicalised compounds and idioms are closely related. Both questions are particularly critical to our understanding of idioms. Are idioms single lexical units like simple words, or do they consist of syntactic structures like other phrases? Similar questions are also valid for lexicalised compounds. Several models have been produced to try to describe how idioms are stored and accessed, and quite a lot of data is now available describing the time course of activation of literal and idiomatic word meanings (Titone and Connine, 1994). In summary, it seems that literal meanings are initially accessed, but as the idiom nears completion the idiomatic meaning becomes activated and the literal meaning deactivated (Cacciari and Tabossi, 1988). Although such studies have not been carried out with compounds, it is worth speculating that they may obtain similar results.

Predicating and Non-predicating Compounds

Some authors divide compounds into **predicating** and **non-predicating** (e.g. Gagné and Shoben, 1995). They define the two types of compound thus "Predicating combinations are combinations of the form XY and can be paraphrased as a Y that is X. **Red apple** is an example of a predicating combination. Nonpredicating combinations also have the form XY, but cannot be so paraphrased; thus mountain stream is **not** a stream that is mountain." (Gagné and Shoben, 1995, p4).

Other authors have used an alternative dichotomy of **property mapping** and **slot filling** (e.g. Wisniewski and Gentner, 1992). Slot filling compounds are so called because the meaning on the compound can be attained by using the modifier to fill an appropriate 'slot' in the representation of the head noun. The slot is a relation in a model something like the Selective Modification Model

(Smith et al, 1988) discussed above. Gagné and Shoben's example of mountain stream would be a slot filling compound according to this classification, **mountain** filling the "location" slot of **stream**. Property mapping compounds are those in which a single property from the modifier is mapped across to the head, e.g. elephant shrew, tortoiseshell cat.

On the surface, non-predicating and slot filling compounds seem the same, while predicating and property mapping compounds seem to differ. However, more fundamentally it may be the case that the two classifications actually refer to the same sets of compounds. Before discussing such issues, it is necessary to explore a little further what these various labels refer to. Most investigations into compounds concern non-predicating or slot filling compounds. The two names for this class of compound suggest the two kinds of theory generally used to describe them. The term non-predicating suggests that some predicative element of the compound is omitted, and interpretation involves the discovery of that relation. Slot filling implies a schema model, in which interpretation is a process of fitting the noun1 into the correct slot in the noun2's schema. Examples of each are discussed below.

Levi (1978) describes a model whereby such compounds are interpreted by the listener / reader discovering the **recoverably deletable predicate** (RDP) which the person coining the compound has omitted from the compound's surface form. Among others, Levi (1978), and following her Shoben (1991), have drawn up lists of the kinds of relations they feel it is possible for non-predicating compounds to express, fig 1.3.

Fig 1.3 : Types of Nonpredicating Relation (Shoben 1991)[7]

A.	N causes Adj FLU VIRUS	H.	N uses Adj GAS STOVE
B.	Adj causes N HEAT RASH	I.	N located Adj URBAN RIOTS
C.	N has Adj PICTURE BOOK	J.	N for Adj NOSE DROPS
D.	Adj has N LEMON PEEL	K.	N about Adj TAX LAW
E.	N makes Adj HONEY BEE	F'.	N derived from Adj OIL MONEY
F.	N made of Adj SUGAR CUBE	I'.	Adj located N MURDER TOWN
G.	Adj is N SERVANT GIRL	H'.	N used by Adj FINGER TOY

As Ryder (1992) points out however, to claim that such lists give a full and explicit account of the possibilities would be misleading. For example, Levi herself points out that the relation **Y in X** (corresponding to category I, N located Adj, in fig 1.3 above) is so broad as to cover a multitude of different meanings, fig 1.4.

Shoben (1991) devised an experiment to test the theory that the type of relationship was critical to the ease of interpretation of a non-predicating compound. He theorised that some relations are inherently more complicated than others. For example, it is perhaps reasonable to suppose that causal relations between constituent concepts (e.g. **flu virus, heat rash**) are more complicated than possessives (e.g. **picture book, lemon peel**). However, when he tested this by measuring comprehension time he found no such difference between relation

---

[7] This table preserves Shoben's terminology of Adj for the noun1, and N for noun2.

types. What he did notice was that some compounds within each type took longer

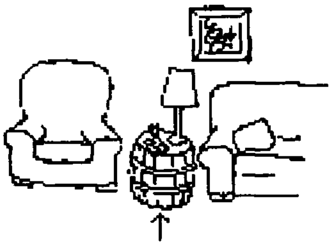
Fig 1.4 : Possible Meanings of IN, from Levi (1978)

Meaning	Examples
inhabit	desert rat, field mouse, city folk
grow in	water lilies, mountain laurel, desert blossoms
according to	logical impossibility, communist tenet, ideological truth
during	spring showers, autumnal rains, night flight
found in	urban parks, city libraries, phonological pattern
occur in	marital adjustments, urban riots, academic transfers

to comprehend than the others. These longer comprehension times seemed to be for those compounds which were difficult to straightforwardly paraphrase. For example **cactus plant** can be quite readily paraphrased as **a plant which is a cactus**, and was quickly comprehended. Conversely, **finger lake** does not paraphrase easily (***a lake which is a finger**), and took longer to comprehend. Similarly **paper money** and **sugar cube** are easily paraphrased and were quickly comprehended, but **mountain range** does not paraphrase readily, and was slow to comprehend. Shoben interprets this in terms of the ease of filling slots in a conceptual schema. Paraphrasing is, in a way, making explicit the slot of the head into which the modifier must be fitted. If the paraphrase is clumsy, perhaps this indicates that the modifier is a relatively ill fit for that slot. This is however a rather speculative idea, and Shoben's judgements of ease of paraphrasing are quite subjective.

Downing (1977) takes an alternative approach, which focuses on the link between the semantics of the constituents and their relationship within the compound. She used several kinds of task with her subjects. First there was a naming task in which subjects named novel objects indicated in illustrations e.g. fig 1.5

Fig 1.5 : An example illustration from Downing (1977)



In other tasks subjects paraphrased lexicalised and novel compounds, or selected appropriate paraphrases from a list. From these data, Downing concluded that several types of relationship within a compound were constrained by the semantic class of the compound constituents, fig 1.6.

Fig 1.6 : Predominant Relationships by Semantic Type, from Downing (1977)

Semantic Type	relationship	example
Humans:	occupational, sexual, and racial identity	police demonstrators, women officers, Negro woman
Animals:	appearance, habitat	giraffe bird, Salt Creek coyotes
Plants:	appearance, habitat	trumpet plant, Texas roadside flowers
Natural objects:	composition, origin, location	granite outcroppings, cow hair, Montana beach
Synthetic objects:	purpose	banana fork

Downing interpreted this in terms of **classificatory relevance** varying with semantic category, that is relations chosen are those with particular significance to the concepts concerned. This is important from a linguistic point of view in that it rejects the idea that there are grammatical constraints on the relation between constituents. Grammatically, the possible interpretations are limitless, but the semantics of compound interpretation bring other pragmatic constraints to bear. In this respect it is surprising that Shoben should have chosen to follow Levi in making finite lists of possible relationships. Downing's insight that the limiting factors are semantic/pragmatic rather than grammatical might be expected to be embraced by the psychological tradition.

In spite of her rejection of a pre-defined list of predicating relationships, Downing apparently accepts that this general kind of slot filling relation is the only possible one. In fact, it has only relatively recently been demonstrated that it is quite common for us to interpret noun noun compounds with an altogether different type of relationship existing between the constituent concepts. Wisniewski and Gentner (1992) asked subjects to paraphrase novel compounds. In these paraphrases they noticed several types of interpretation. As well as the previously recorded and much debated **X relation Y** interpretations, which they called slot filling, there were others which they termed **property mapping** and **structure mapping**. In property mapping, instead of the modifier filling a slot in the representation of the head (e.g. **book box**, in which **book** fills the slot **contains** in the representation of **box**), a single property of the modifier is mapped across and applied to the head. For example, a **robin snake** was described as **a red snake**. Structure mapping occurs when a relation between the properties of the modifier is mapped across to the head. For example, a **pony chair** was said to be a small chair. In this case, it cannot be thought of as simply mapping across size, because ponies are generally bigger than chairs, not smaller. Rather it is seen as mapping across the size relationship between ponies and their superordinate category, horses, onto pony chairs and their superordinate category, chairs. So **pony chairs** are small in relation to other chairs just as ponies are small in relation to other horses.

Although Wisniewski and Gentner do manipulate the variable of constituent semantic class (i.e. they generate compounds from artefact and natural kind concepts in each possible configuration), they do not report data according to this dimension. They also offer little by way of quantitative analysis of their data, simply reporting that approximately 30% of paraphrases imply a property mapping interpretation. In a more recent paper Wisniewski (1996) describes a similar paraphrasing experiment, in which he also manipulates the similarity of constituent nouns to each other. In this study he found that property mapping was much more likely when constituent nouns were similar. He interprets this in terms of structural alignment. That is, when two concepts are similar, it is easy to align their representations and so compare them. Property mapping is likely to occur at the point at which the two concepts differ. Again, although he included concepts of various ontological categories in his study, he does not report any analysis based on this variable.

Summary

Throughout this survey of the literature, a range of issues have cropped up. Research into conceptual combination, and into noun noun compounds in particular is quite recent, and so much of the research discussed above has raised more questions than it has answered. Noncompositionality is problematic for models of conceptual combination, although extensional feedback and maintenance of coherence inferences may prove to account for most emergent features. Experiments 1 - 4 of this thesis look at the phenomenon of noncompositionality, and try to identify possible sources of emergent features. Slot filling and property mapping are two important types of compound interpretation. Factors which may effect these are examined in experiments 5 and 6. Experiment 7 examines concept identification. The role played by each constituent position in determining the relationship between constituents is unclear, but important to understand if a full account of compounds is to be reached. This is the subject of experiment 8.

Also important to the psychology of concepts generally are possible differences between artefact and natural kind concepts, a topic which has not been properly addressed with respect to compounds. This is a theme throughout the experiments described in this thesis. The relationship between lexicalised and novel compounds is another potentially important area which has been neglected. Comparisons between these two types of compound also run throughout the experiments in this thesis.

Chapter two: Noncompositionality

Experiment 1 : Measuring Noncompositionality

Introduction

This experiment is designed to investigate "extensional feedback" (EF), and "maintenance of coherence inferences" (MCI) which Hampton identified as probable sources of noncompositionality (Hampton, 1987). It also seems plausible to suggest that some sort of "conventional wisdom", a kind of idiomatic effect, is responsible for some noncompositional meaning. Many of the features we associate with familiar concepts are true by convention rather than by personal experience or personal opinion. So a `gold mine' may be thought of as a source of inevitable wealth and happiness, or `traffic warden' as cruel and vindictive. These are items of received wisdom which we associate with the concepts concerned, whether they agree with our own personal experience or not.

In this experiment, two different categories of concept were looked at.

Real Compounds

e.g. **brick wall, car alarm**.

These are familiar compounds with well known meanings. This would allow all three effects (EF, MCI and `conventional wisdom') to influence the eventual form of the compound, and so would be predicted to have the highest level of noncompositionality.

Novel Compounds

e.g. **parsnip tree, chair mouse**.

These compounds are completely novel, and so have no real reference, but subjects could be asked to imagine what such things would be like if they did actually exist. In this case extensional feedback and conventional wisdom ought to be inhibited, leaving maintenance of coherence inferences as the only source of noncompositionality.

So a comparison of real and novel compounds should give some indication of the amount of noncompositionality due to extensional feedback and conventional wisdom. The remaining source of noncompositionality is maintenance of coherence inferences. In order to isolate this effect, subjects must somehow be inhibited from making inferences about the nature of the compounds. The theory of psychological essentialism (see Medin and Ortony, 1989) is utilised here in an attempt to do this. This theory ignores the question of whether or not extensional categories genuinely have essences, and merely states that people behave as though concepts have essences. The question of whether or not this essentialism is also a real world phenomenon is a topic of current debate. For example, it has been claimed that real categories do have essences, but that only experts in the relevant fields have any knowledge of these. This debate however is not immediately crucial for the purposes of this investigation. What is important is that there appear to be distinct types of concept, some of which have fairly accessible essences, and others for which the essence is rather more elusive.

Two quite distinct types of category stand out; **artefacts** and **natural kinds**. There are theoretical distinctions between these types, but little experimental work has been carried out. Artefacts are "man-made" items, and as such their essence is relatively clear. The essence of an artefact is usually its intended function, or some combination of this with its actual use. In the case of an artefact, to say what it is "essentially", is to say what it is "to all intents and purposes". This much is fairly clear, that people in general have access to what they believe to be the essences of artefacts.

Natural kinds, however, comprise the opposite end of the spectrum. The reasons for this again are moot. It seems likely that, when contrasted with artefacts, the evolutionary/geological/meteorological reasons for the category formation are without intent. Or if we appeal to religion for an explanation, we must at least accept that the divine plan is unclear to us.

This distinction between artefacts and natural kinds is utilised here in an attempt to isolate maintenance coherence inferences. By using an artefact in the noun1 position (as a modifier) and a natural kind in the noun2 position (as a head noun), the subject will be faced with a head noun which has an inaccessible core meaning. The difficulty of knowing what the essentials of the head noun are should inhibit the subject from modifying it. In this case, no elaborative inferences about the nature of the compound can really be made. Three types of compound are therefore compared, real artefact artefact (rAA), novel artefact artefact (nAA) and novel artefact natural kind (nAN) compounds. It is predicted that these will show decreasing levels of noncompositionality in that order.

Noncompositionality can be measured by comparing how descriptive a number of features are of the compound and how descriptive those same features are of the nouns which constitute that compound. This means obtaining a list of features some of which should provide a good range of applicability to the three parts of each compound (noun1, noun2 and the compound itself). These features are then rated as to how descriptive they are of each of these three parts. Noncompositionality is be measured by calculating how much difference there is between the ratings with respect to the compound and the ratings with respect to the constituents. Only the magnitude of these differences is relevant to noncompositionality. Although the direction of difference is of descriptive interest with respect to the meanings of the features, noncompositionality could be manifest in both positive and negative differences - features might be present in the compound but absent from the constituents, or vice versa. For this reason absolute differences are needed to capture the overall noncompositionality without such directional differences being able to cancel out one another. Thus, the bigger the difference, the greater the noncompositionality. Measuring noncompositionality in this way means that it is possible to see how much meaning of the compound is common to each of its constituents independently. In keeping with the idea that the noun2 should act as the head noun, it is predicted that this should have most meaning in common with the compound.

To recapitulate briefly, the motivation behind this experiment is to try to find evidence of different levels of noncompositionality between different types of noun noun compound. It is predicted that real artefact artefact compounds (r.AA) will have most noncompositionality, followed by novel artefact artefact (n.AA) compounds and novel artefact natural kind compounds (n.AN). This is measured by finding the absolute difference between ratings for features with respect to the compound, and ratings for the same features with respect to its constituent parts. It is also predicted that this difference will be greater for noun1 than for noun2.

Method

This experiment contained two stages; **feature generation**, in which subjects are asked to list the features of concepts, and **feature rating,** in which features are rated for their descriptiveness of concepts.

Stage 1 - Feature Generation

The purpose of this stage was to produce feature listings to be used in stage two, the feature rating stage.

Subjects

Twenty-four unpaid volunteers participated in this experiment. They were all undergraduates studying the University of Durham first year psychology course, aged 18 - 30 years. There was an approximately equal number of males and females.

Design

This stage used a repeated measures design. Each subject saw items from each of six different item categories. Details are given in the Questionnaire Design section below. A replicate set of questionnaires was administered to a second group of subjects.

Materials

Questionnaires were designed to elicit from the subjects a range of features associated with twenty-four noun noun compounds in each of the three experimental conditions (r.AA, n.AA and n.AN), and with their constituents (noun1 artefacts, noun2 artefacts and noun2 natural kinds). The selection of materials and design of questionnaire were as follows.

Real Artefact-Artefact Category : Twenty-four noun noun phrases were selected. Each of the forty-eight constituents was an artefact, and each phrase had a real reference, e.g. **seat belt**. The mean frequencies of these words were 29.87 for noun1, and 30.38 for noun2 (Kucera and Francis, 1967).

Novel Artefact-Artefact Category : This category was generated directly from the real artefact-artefact category, by swapping the noun2s with one another, to produce novel combinations. In order that the relative word frequencies of noun1 and noun2 were controlled for, the frequencies of each noun were used to determine which novel compounds were formed. Any frequency imbalance between noun1 and noun2 was maintained by ordering the real artefact-artefact compounds with noun2s in ascending frequency, and then swapping over positions 1 & 2, 3 & 4, etc. In three cases this resulted in the new complex referring to a real object, i.e. shower shelf, refuse bin, and library curtain. In these instances the noun2s were swapped again (with their next closest neighbours) in order that there should be no real world referent of each compound.

e.g. **carpet shampoo** → **carpet hook**
coat hook → **coat shampoo**

Novel Natural kind - Artefact Category : A list of twenty-four single natural kind nouns was also compiled, e.g. **horse**. These were ordered for frequency to match the noun2s of the n.AA category, and then substituted for them to produce n.AN compounds. This maintained any frequency relationship between noun1 and noun2.

e.g. **coat hook** → **coat frog.**

These three categories, r.AA, n.AA and n.AN, along with their constituents, were used as the materials in the investigation. The ordering of the materials was randomised so that they were no longer ordered by frequency, and then twelve different sets of materials were compiled. Each set contained twelve items, two each of each category, that is two each of :-

Artefacts (noun1s), artefacts (noun2s), natural kinds (noun2s), real artefact-artefact combinations, novel artefact-artefact combinations, novel artefact-natural kind combinations. The order of occurrence of these in each set was also randomised. A list of items used in this experiment is given in the appendix, table A1.1.

Questionnaires : Each questionnaire comprised a front instruction page (see below), followed by 6 test pages. Each test page presented two of the materials as shown above, each material item being followed by some blank lines on which its attributes were to be listed. The entire set of 12 questionnaires was replicated with another set of subjects.

Procedure

Subjects were given the questionnaires and the following "instructions to subjects" (also printed on the questionnaire) were read out to them.

"On the following pages are twelve different phrases, each made up of either one or two words. Each phrase names a type of object, and after each is a number of dotted lines. I want you to imagine that you have to describe or define the object to someone unfamiliar with it. Do this by listing on the dotted lines the properties or attributes that you consider important to this object, and which could be used to decide whether or not something is this kind of object. You may use a few short phrases if it helps.

Some of the objects might not actually exist in the real world. If this is the case, then try and imagine what the object might be like in some imaginary world, and list its attributes as though it did actually exist. Try to list as many attributes as possible. Aim for ten, but more if you like. Take two or three minutes over each object. When you've listed all you can for one object, please go back over your lists and rate each attribute according to the following scale (just write the letter);

N = necessarily true for all possible examples
A = a very important part of the definition
B = a fairly important part of the definition
C = typically true, but not defining"

Stage 2 : Feature Rating

The features listed in stage 1 of the study were used as features that subjects were asked to rate in this stage.

Subjects

Forty-eight unpaid volunteers participated in this experiment. They were all undergraduates studying the University of Durham first year psychology course, aged 18 - 30 years. There was an approximately equal number of males and females.

Materials

Questionnaires : The twelve material sets used in stage 1 were now split into twenty-four sets, each containing six items, one from each category. Under each item were listed those features that had been produced for that item in stage 1. In the cases of the heading being a single word, then all the features listed for the compounds in which that word was a constituent were also listed. Similarly, if the heading itself was a compound, then all the features listed under either of its constituents, or other compounds of which they are constituent parts, were also included. These twenty-four material sets were compiled into questionnaires. As in stage one, an instruction page was followed by six test pages. The instructions to subjects were as follows.

"At the top of each of the following pages is a different phrase, which is underlined, each is made up of either one or two words. Each phrase names a type of object, and below each is a number of possible attributes. I want you to rate these attributes as to how characteristic they are of the named object. In doing this, you should decide if the attribute describes this object, and if it could be used to decide whether or not something is this kind of object.

Some of the objects might not actually exist in the real world. If this is the case, then try and imagine what the object might be like in some imaginary world, and list its attributes as though it did actually exist.

Please rate each attribute according to the following scale (just write the letter on the space provided to the left of each attribute);

N = necessarily true for all possible examples
A = a very important part of the definition
B = a fairly important part of the definition
C = typically true, but not defining
D = not usually true
I = impossible for any example

(The scale is printed at the bottom of each page for easy reference.)"

Each test page was headed by one material item, which was followed by a list of possible features, each with a space provided for a rating. The entire set of questionnaires was again replicated for a second subject group.

Design and Procedure

Items data were analysed using a 3 (Compound Type) x 2 (Constituent Position) x 2 (Subject Group) factorial design, with repeated measures on the last two factors. Compound Type was either real artefact-artefact, novel artefact-artefact, or novel artefact-natural kind. Constituent Position was either noun1 or noun2.

Results

Calculation of Results

The ratings given were first converted to numeric form. In an attempt to reflect the non-linear relationship between the ratings, the conversion included an extension of the extremes, i.e.

N = 5 A = 3 B = 2 C = 1 D = 0 I = -2

This is based on assumptions referred to by Hampton (1987), relating to the exaggerated effect of necessity and impossibility compared to other grades of feature attribution.

Tables of means and ANOVA tables of these raw scores are given in the appendix, tables A1.2 and A1.3 respectively.

The ratings given to features listed under each compound were then compared to the ratings given to the same features for the compound's constituents. This provides a measure of noncompositionality (nc). The absolute differences were calculated as shown below.

Calculation of Noncompositionality (nc)

For a compound 'AB' :-

$$nc_A = \frac{\sum_{i=1}^{n} AB_i - A_i}{n} \qquad nc_B = \frac{\sum_{i=1}^{n} AB_i - B_i}{n}$$

where $nc_{AB} = nc_A + nc_B$

The higher the score, the less meaning there is in common between compound and constituents. This provides two results, one for noun1, and one for noun2.

Absolute Differences Analysis

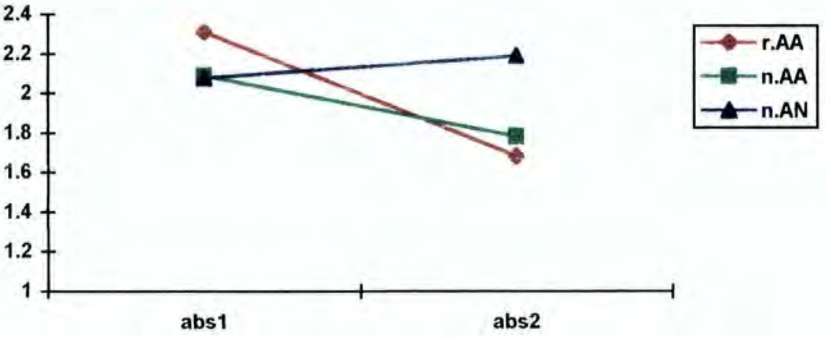
The absolute differences between the ratings for each feature with respect to noun1 and the compound and noun2 and the compound were calculated (absdiff1 and absdiff2 respectively). Because not all of the features rated for any single part were rated for the other parts of that item, only the subset of features common to the compound and noun1 or noun2 were used. The mean absolute difference scores in each condition are shown in Table 2.1.

A 3 (Compound Type) x 2 (Absdiff Part) x 2 (Subject Group) analysis of variance with repeated measures on the final two factors, was conducted for this data. Absdiff Part was the only significant main effect for this analysis ($F(1,61) = 5.28$, $p < 0.05$). Compound Type x Absdiff Part was the only significant interaction ($F(2,61) = 2.72$, $p < 0.05$). As with analyses reported throughout this thesis, means and full anova tables are given in the appendix (see tables A1.2 and A1.3).

Table 2.1 : Compound Type x Absdiff Part x Subject Group

	absdiff part1			absdiff part2			Mean
	subject group 1	subject group2	mean	subject group 1	subject group2	mean	
r.AA	2.461	2.162	2.312	1.716	1.651	1.684	1.999
n.AA	2.137	2.038	2.088	1.765	1.801	1.783	1.936
n.AN	2.144	2.013	2.079	2.224	2.158	2.191	2.135
mean	2.245	2.071	2.160	1.900	1.870	1.886	2.023

Figure 2.1 : Compound Type x Absdiff Part



Absdiff1 (mean = 2.16) is greater than absdiff2 (mean = 1.89). The interaction of this effect with Compound Type is shown in Fig 2.1. Inspection of Figure 2.1 shows that the novel AN Compounds exhibit a reversal of the main effect of Absdiff Part, absdiff2 being greater in this case.

Differences Scores

Signed differences between the two nouns and the compound were also calculated for each common feature. Tables of means and ANOVA tables of these difference scores are given in the appendix, tables A1.4 and A1.5 respectively.

Discussion

The only significant main effect was Absdiff Part (Absdiff1 > Absdiff2). The interaction Compound Type x Absdiff Part was also significant. It had been predicted that noncompositionality would increase from real AA to novel AA to novel AN compounds. Since there was no main effect of Compound Type, this prediction was not upheld.

The main effect of Absdiff Part shows that the compound has more meaning in common with noun2 than with noun1. This confirms the hypothesis that noun2 will tend to act as the head of a compound. However, there is a significant interaction between Compound Type and Absdiff Part. This interaction is as predicted, with absdiff1 being greater than absdiff2 for both r.AA and n.AA compounds, but there is little difference between absdiff1 and absdiff2 for the n.AN compounds. The constituents of the n.AN compounds therefore seem to contribute equally to the meaning of the compound. Superficially, this finding is consistent with the hypothesis that subjects would interpret n.AN compounds by taking the union of the meanings of their constituent nouns. But contrary to prediction, this equal contribution between the two constituents is not due to an increase in the amount of meaning contributed by the noun1. Comparison with r.AA and n.AA compounds shows that it is in fact due to a decrease in the amount of meaning noun2 and the compound have in common. This result would therefore appear to refute the theory that the meaning of a n.AN compound is the union of the meanings of its constituents. However there are a number of other possible influences. Firstly, the n.AN condition is to be contrasted with the other two in that its component nouns are from different semantic groups, and are therefore unlikely to have much in common with each other. This in itself is likely to increase the apparent noncompositionality of the AN compounds.

Also, if the AN compounds are genuinely difficult to interpret, then different people are likely to come up with different interpretations of them. If this is the case then subjects who are rating compounds may well differ in their opinions about the meanings of compounds from those who generated the features. Some evidence for this influence can be found in the signed difference data (see the appendix, table A1.4). Both diff1 and diff2 have positive scores for the real AA compound type, but negative scores for the novel compound types. This means that only in the real case were features rated higher with respect to the compound than with respect to its constituents. The likely reason for this is the high relative ambiguity of the novel compounds compared to real compounds. The increase in negativity of these scores from novel AA to novel AN compounds might therefore be accounted for by a similar increase in ambiguity, as would be predicted from the assumption that natural kind nouns are difficult to modify. However, the raw mean scores (appendix, table A1.2), show that, compared to r.AA and n.AA compound types again, it is an increase in ratings for noun2 rather than a decrease in ratings for the compound which is the source of the highly negative n.AN diff2 score. The n.AN noun2 is natural kind noun condition in the experiment, and has been more highly rated than the other conditions (r.AA noun1 and noun2, n.AA noun1 and noun2 and n.AN noun1) which are all artefacts. It appears therefore that this is a simple effect of semantic category - that the natural kind nouns were rated more highly with respect to the listed features than were the artefact nouns. Thus it seems that to the subjects, the features are somehow generally truer of the natural kinds than they are of the artefacts. A problem in the design of this experiment is that the same representative set of features is not rated with respect to each of the three parts of each compound. This prevents the teasing apart of any simple effect of semantic category which might occur, such as that postulated here, and any more complex effects concerning the interpretations given to different compound types.

The implication of this finding is that the high absdiff2 score for the novel AN compound type is the result of particularly high ratings for the natural kind noun2 rather than low ratings for the AN compound. This effect seems to confound the overall comparison of novel AN compounds with the other two compound types to such an extent that it is not worthwhile attempting to explain the contrasts between them in terms of the meanings of the compounds. An experimental design which more closely controls the features presented to subjects ought to be able to examine any difference in compound types more properly. Such a design is used in the next chapter.

It is still possible to compare the real AA compounds with the novel AA compounds. There was no overall effect of compound type in the analysis, and the interaction between Compound Type and Compound Part was due to the n.AN compounds. It must therefore be concluded that no effect of extensional feedback has been found here. It had been predicted that extensional feedback would lead to an increase in the number of noncompositional features, so it remains to be determined why the real AA combinations did not exhibit significantly higher degrees of noncompositionality than the other categories. One possibility is that the process used to manufacture the novel AA compounds has actually resulted in producing combinations which people feel have real references. Some of those produced were rejected for this reason, e.g. `refuse bin', but these were selected only by the experimenter's intuition, and some of those not rejected may have suffered this drawback. However, it seems only a remote possibility that enough truly referential compounds slipped through the net to make a genuine difference to the results.

An alternative explanation is one offered by Hampton, in his discussion of a pilot study (Hampton 1991). His subjects were asked to draw examples of non-existent items, such as "a teapot that is a computer", "a bicycle that is a kind of stove" or "a fruit that is a kind of furniture". According to Hampton, noncompositional elements abounded (although quantitative analysis of the results was considered impossible). This is as he predicted, his subjects having been asked to "combine the uncombinable" then "struggled to find a consistent composite prototype to represent the new type of object" (Hampton 1991). Hampton suggests that presenting subjects with novel combinations will, if these combinations are improbable enough, cause a greater than normal amount of maintenance of coherence inferences, which would counteract any corresponding decrease in extensional feedback. He goes on to state, in reference to such inferences, "One can see that as each change is implemented, the subject has to think through the consequences, and make further adaptations. Different people will have quite different answers to the problem." This highlights a related concern with the present study. Subjects were only given features to rate, they were not allowed to introduce their own interpretation, which was therefore excluded unless it coincided with that produced by the feature generation stage of the study. This in itself may have distorted the proportions of emergent to inherited features present. This seems most likely to have occurred in the novel AN category, in which the combinations were particularly obscure, and so their resolutions particularly idiosyncratic.

Again, although there is no overall effect of compound type with respect to noncompositionality, absdiffpart1 and absdiffpart2 do seem to behave slightly differently between r.AA and n.AA compounds. The difference between the two absdiff parts is rather greater for the real compounds than for the novel ones. In fact the absdiff1 of both the novel AA and the novel AN compound types are at the same level, both below that of the real AA type. This suggests that in general subjects are interpreting the novel compounds as inheriting more of the noun1s meanings than is the case for the real compounds. In this sense, the novel compounds can actually be considered more compositional than the real compounds.

In conclusion, then, no overall difference in compositionality was found. This is not interpreted as denying that the processes of extensional feedback and maintenance of coherence inference are present, but that they have been confounded with other influences. There is evidence that artefacts and natural kinds may behave differently, making the results of the novel AN compounds difficult to compare with the results of AA compounds. However, it does seem that the contribution that each constituent makes to the compound is more equal for the n.AN compounds than for r.AA or n.AA compounds. There is also some evidence to suggest that real compounds derive less of their meaning from noun1 than do novel compounds.

Chapter three: Factors that Influence Noncompositionality

General Introduction

Experiment 1 suggested that the relationship between the meaning of a noun noun compound and the meanings of its constituent nouns depends to some extent upon the semantic categories of those constituents. The following three experiments investigate this relationship in more detail. Experiments 2, 3 and 4 are best seen as follow-up studies to Experiment 1. Experiment 1 was to some extent inconclusive. It had been predicted that the novel artefact - natural kind case would be distinct from the other two types of compound, in that each of its constituent parts would contribute equally to the meaning of the compound. Although not conclusive, the data does suggest this is the case. However, it had also been predicted that there would be a different level of noncompositionality associated with each compound type - this was not found. Experiments 2, 3 and 4 are therefore intended to re-examine these questions in a more robust design, to look more closely at the ratings given to specific groups of features, and to extend the scope of investigation to more types of compound. Experiment 2 looks at the compositionality by examining the features listed for compounds and their constituents, while experiments 3 and 4 look at feature ratings.

In experiment 1, the factors across which concept combinations were varied, compound type and extensional status, were only represented in 3 of their 8 possible permutations. These experiments include examples of all possible combinations of the factors used, see table 3.1. The factor of Compound Type used in experiment 1 is therefore no longer appropriate, and compounds are varied across two factors: Compound Composition (AA, AN, NA and NN) and Extensional Status (real and novel).

Table 3.1 : Eight possible permutations of the factors Extensional Status and Compound Composition

real artefact - artefact (rAA)	novel artefact - artefact (n.AA)
real artefact - natural kind (r.AN)	novel artefact - natural kind (n.AN)
real natural kind - artefact (r.NA)	novel natural kind - artefact (n.NA)
real natural kind - natural kind (r.NN)	novel natural kind - natural kind (n.NN)

It is difficult to predict the outcome of this more complete manipulation. It was predicted that the three original types of combination would exhibit successively less noncompositionality, however, this was not found to be the case in experiment 1.

Experiments 3 and 4 also include one further factor, that of Feature Set (fs). This factor splits the features into 3 sets according to their derivation, i.e. if they were originally listed as features of noun1, noun2 or the compound (fs1, fs2 and fs3 respectively). Clearly this factor would be expected to interact strongly with Compound Part, each set of features should be highly descriptive of the concept they were originally used to describe. It is also predictable that this interaction should vary with the factor Extensional Status. Real compounds are well defined, and as such feature set 3 should provide a good description of the compound. Novel compounds are probably subject to a good deal more variation in interpretation, and as such fewer of the features in fs3 will be deemed to be highly descriptive of the compound.

A fuller investigation of Extensional Status will also be allowed by experiments 3 and 4, by offering a comparison between real and novel compounds of each combination type. It is possible that the artefact artefact combination used for the real/novel distinction in Experiment 1 is an anomalous case, and that other combinations will exhibit the predicted increase in noncompositionality. Alternatively all combinations may prove consistent with respect to this factor, and so confirm the impression that the real/novel distinction is less powerful than was predicted. The present design will also enable further examination of the artefact natural kind combination. Experiment 1 suggested that n.AN compounds might not conform to the usual pattern of noun1 acting as modifier to noun2 as the head noun. In this study, this finding is retested, and compared to other compound types not previously examined. Each of these compound types is also investigated at a more detailed level, in that the features used to measure the meanings of the compounds are divided into feature sets according to their origin. Thus the source of any variation from the standard modifier-head relationship is traceable. In looking at other compound types, particular attention is directed towards the r.AN compounds, and the n.NA compounds to see if either of these types (which are clearly related to the n.AN type) behave in any similar way.

In general, it is also expected that there will be an overall difference between the behaviour of noun1 and noun2 in the compounds. With the possible exceptions mentioned above, noun2 is expected to act as head noun in most instances, and therefore to determine the majority of the meaning of the compound. This should have implications for several aspects of the results, including a higher mean rating of features with respect to noun2 than noun1, and a greater difference of meaning between the compound and noun1 than there is between the compound and noun2. Similarly, it is also expected that in most cases the compound will exhibit a high degree of compositionality, and so features from all feature sets will be rated relatively high for the compound. This is in contrast to each of the compound parts, which should only have high ratings for a more limited range of features, particularly those from the respective feature set. Thus a distinction between the mean ratings received with respect to the compound and with respect to the constituent parts is predicted.

Experiment 2 : Measuring Noncompositionality by Feature Lists

This experiment looks at the features generated by subjects, comparing those listed for simple concepts with those listed for compounds. The factor of Extensional Status is included in an attempt to isolate the effect of extensional feedback. Also the different combinations of artefact and natural kind nouns are investigated to see if any differences occur with respect to such compound types. Such differences could be due to different conceptual structures.

Method

Design and Questionnaire Composition

This experiment has a 4 (Compound Composition) x 2 (Extensional Status) x 3 (Compound Part) factorial design, with repeated measures on the last factor. The three compound parts were simply **noun1**, **noun2**, and **compound**. The four compound types were determined by the type of constituent noun used (either **artefact** or **natural kind**) in each position (either **noun1** or **noun2**). Extensional Status was either **real** (the compound describes some entity that exists in the world) or **novel** (it describes something which does not exist, and can only be imagined). The concepts investigated therefore fall into eight separate categories, an example of each of which is shown in Table 3.2 below. (The order of mention of the constituents matches their position in the compound).

Table 3.2 : Compound Types and Examples

Compound Composition	Real	Novel
Artefact - Artefact	coat hanger	weapon clamp
Artefact - Natural kind	computer mouse	carpet slug
Natural kind - Artefact	daisy wheel	giraffe motor
Natural kind - Natural kind	crab apple	whale pineapple

Six compounds were selected in each condition, giving a total of forty-eight combinations, which are listed in the appendix, table A2.1. The **real** compounds were chosen first, and the **novel** compounds were generated by choosing words (artefact or natural kind nouns, as required) matching the word frequencies (from Kucera and Francis, 1967) to those in the **real** condition. Details of frequencies are also in the appendix.

Subjects

Forty eight unpaid volunteers participated in this experiment. They were all undergraduates studying the University of Durham first year psychology course, aged 18 - 30 years. There was an approximately equal number of males and females.

Materials

Each of the forty-eight compounds was divided into a triple, the three elements of this triple being **noun1** (e.g., "carpet"), **noun2** (e.g., "slug"), and the **compound** itself (e.g., "carpet slug"), resulting in 144 items.

Each subject was presented with items of the same composition (e.g., Artefact - Natural kind), but from both categories of referential status (i.e., real and novel), so the materials were divided into four groups along these lines. Thus, the six real instances of each composition type were paired with the six novel instances of same composition type. From each of these four groups of twelve triples, twelve materials were selected for three different questionnaires, each questionnaire containing one part (noun1, noun2, or the compound) of each triple in the group. These three questionnaires were from only one compound type, so all four compound types resulted in a total of twelve different questionnaires.

The elements within each questionnaire were randomised, into four different random orders, resulting in forty-eight different sets of materials, with each of the 144 items being presented four times.

Each questionnaire contained twelve concept names (two to a page), and under each concept name were twelve dotted lines.

Procedure

Each subject was given a questionnaire, the first page giving the following instructions.

"On the following pages are twelve different phrases, each made up of either one or two words. Each phrase names a type of object, and after each is a number of blank lines. I want you to imagine that you have to describe or define the object to someone unfamiliar with it. Do this by listing on the blank lines the properties or attributes that you consider important to this object, and which could be used to decide whether or not something is this kind of object. You may use a few short phrases if it helps.

Some of the phrases might describe things that don't actually exist as things in the real world. If this is the case, don't worry, they are meant to be like that. Just try and imagine what they might be like in some other world, and list these attributes as though it did actually exist. Try to list as many attributes as possible. Aim for ten, but more if you like. Take two or three minutes over each phrase."

Results

The lists of features produced by subjects were compared to see how many features were common to the elements of each concept triple. Of interest is the number of features (indicating the amount of meaning) common to the constituent concepts (noun1 and noun2) and the compound concept. There are several different ways of counting the number of features that these concepts have in common. The counts presented here cover two alternative dimensions. First, there is the level of precision. Clearly there is a good deal of ambiguity in the notion of two features being "the same". In order to try to get consistent measures across all of the different categories involved, two levels of precision were used - **exact** and **liberal.** Although these categories are also ambiguous, using them should help to allay concerns that the notion of `sameness' is so broad as to be useless. Features were considered to be **exactly the same** if their meanings were thought synonymous. This would included verbatim matches (e.g. "is black", "is black") but also some features which are worded differently, but cannot really be thought of as being different in meaning (e.g. "obscure", "rare/unusual"; "keeps animals in", "used to hold animals"). Features were considered **liberally the same** if they referred to the same underlying meaning, or if one feature was more specific or informative than the other (e.g. "vegetable", "yellow vegetable"; "circular container", "used to keep things in").

The other dimension included in the analysis of results was that of **types** and **tokens**. Between two lists of features there will be a certain number of matches between features such that a feature on one list matches a corresponding feature on the other list. These are **token** matches. There may, for example be 3 token matches between two lists. But these matches may cover only a single **type** of feature. That is they may all refer to the same meaning (e.g. they might all state that the item concerned "is round"). So this would constitute only a single match of **types**. It is clearly sensible to distinguish this situation from one in which the three token matches were all of different types (e.g. "is round", "is green", "is edible"). So the feature lists generated by the subjects were worked through, counting the number of matches in each of the four different combinations of **exact** and **liberal** counts, and **type** and **token** counts.

An analysis of variance was conducted for each of these counts according to a 4 (Compound Composition) x 2 (Extensional Status) x 2 (Part) design, with repeated measures on the last two factors. Results from each of these analyses were similar, the significant effects being the same in all except for one case, which is detailed below, and all significant differences were in the same direction. For this reason, only the results for the **exact type** count are given here, since these are representative of the other results which are given in the appendix, tables A2.4 - A2.9.

The main effects of Compound Composition ($F = 8.00$, $df = 3, 40$, $p < 0.001$) and Part ($F = 12.36$, $df = 1, 40$, $p < 0.001$) were significant. Also significant was the interaction Compound Composition x Part ($F = 3.92$, $df = 3, 40$, $p < 0.015$) (this interaction was not significant in the analysis of the data of the liberal type count).

As can be seen in Table 3.3, the main effect of Compound Composition seems to arise because there are fewer feature matches for AA and NA compounds than AN, which in turn had fewer than the NN compounds. The main effect of

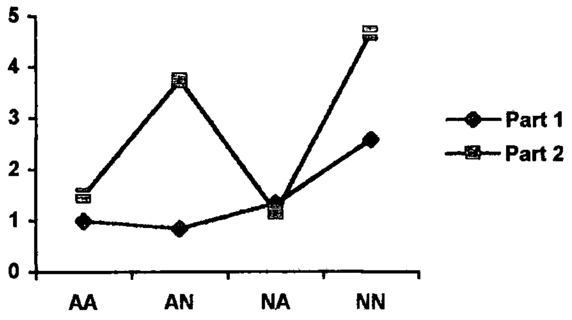
Compound Part is due to there being more features common to noun2 and the compound than there are between noun1 and the compound.

Table 3.3 : Mean number of exact feature type matches for Compound Composition x Part

	AA	AN	NA	NN	mean
Part 1	1.00	0.85	1.34	2.59	1.45
Part 2	1.50	3.75	1.17	4.67	2.77
mean	1.25	2.30	1.26	3.63	

The interaction Compound Composition x Compound Part is shown in Figure 3.1. Inspection of figure 3.1 shows that the main effect of Compound Part is due to the AN and NN compound compositions, the differences between part1 and part2 being negligible for the AA and NA compounds.

Figure 3.1 : Compound Composition x Compound Part



Discussion

This experiment investigates the compositionality of compounds, in a design relying upon subjects producing descriptive lists of features. The number of features common to the descriptions of the constituent concepts and the compound give a measure of this compositionality. Firstly, with respect to the factor Extensional Status. There was no difference between Real and Novel conditions. This result confirms and extends the findings of experiment1 - that novel compounds are no more compositional than real compounds. As such, it provides no direct evidence of extensional feedback. However, this should not be interpreted as conclusively indicating that extensional feedback does not occur. It remains quite possible that extensional feedback is occurring here, but that some other influence is causing a similar increase in noncompositional features amongst the novel compounds, thus eliminating any difference between the two. Furthermore, there is a good candidate for causing such an increase in noncompositional features in novel compounds. This is the inferences needed to make novel compounds seem sensible and coherent, as Hampton (1991) suggested.

To consider next the effects which were significant. The difference between compound parts is straightforward to account for. It is usually assumed that noun2 is the head in noun noun compounds, so the compound will inherit most of its meaning from this concept, and less from noun1. It therefore follows that noun2 should have more features in common with the compound than noun1 does, which is found to be the case in this experiment. This result therefore gives support to the belief that noun2 is usually the head in both real and novel compounds.

However, the interaction between Compound Composition and Compound Part modifies this conclusion. The greater number of common features for noun2 only holds for AN and NN compounds. So the main effect of Compound Part was due only to the AN and NN compounds. The NN compounds also have more features in common for noun1 than other compounds have. The explanation of this finding would seem to lie in the number of common features in different concepts. Or more particularly between and within different types of concept. It seems that for the NN compounds features of both constituents are quite common in the description of the compound. In this case all of the concepts concerned are natural kinds. For the AA compounds fewer features are common, and this is true for the NA compounds too. However, in the case of the AN compounds few features are common between noun1 and the compound, but many more are common between the noun2 and the compound. Following the principle than noun2 acts as head, we can assume that the AN compounds themselves are natural kinds, so this finding is consistent with the conclusion that features are relatively common between natural kind concepts, but rather less so between artefact concepts.

In conclusion, no difference has been found between the levels of compositionality of real and novel compounds. But this cannot be taken as evidence that extensional feedback does not occur in real compounds, since it cannot be ruled out that some other influence causes similar noncompositionality among novel compounds. Differences in the number of features common to the two constituents of compounds were detected. This is simply due to noun2 commonly acting as the head noun and so passing on most of its meaning to the compound. Differences in the compositionality of compounds composed of different types of concept are attributed to differences between how common features are between artefacts and natural kinds.

Experiment 3 : Measuring Noncompositionality by Feature Ratings

Findings from experiment 1 suggest that the approach used to investigate noncompositionality is worthwhile pursuing. Experiment 3 therefore uses a similar design, with a more extensive range of materials. Experiment 1 was composed of a feature generation stage, followed by a feature rating stage. For experiment 3, the features used are those generated and analysed in experiment 2, so experiment 3 uses a feature rating task.

Method

Features for this experiment were those generated in experiment 2. Subjects were required to rate the features derived from experiment 2 for their importance to given concepts.

Subjects

Twenty-four unpaid volunteers participated in this experiment. They were all undergraduates or postgraduates studying psychology at the University of Durham. There was an approximately equal number of males and females, aged 18 - 30 years.

Materials

The features to be rated were selected from those generated in experiment 2. Features were selected as follows. The lists of generated features were grouped into their triples, i.e. the list of generated features for a compound concept was grouped with those of its corresponding noun1 and noun2 constituent concepts. From each of these three lists, 10 features were selected, and so a 30 feature list was compiled. This list was then presented for rating with each of the three concept headings.

The 10 features from each list were selected according to the following rules, in descending order of priority.

- features were selected first from those generated for noun1, then noun2 and finally the compound concept,
- no feature was included more than once. Thus, if a feature was included from noun1 (e.g. the feature "has wings" from `sparrow') it will not be duplicated if it occurs in the compound as well (e.g. if the compound were `sparrow hawk' which also had "has wings" listed as a feature.)
- if a feature was generated more than once for a concept, then it was included,
- features which were judged to be progressively less similar in meaning to others generated for the same concept were included until 10 features had been selected from that concept. E.g., if "small" had already been included, then some similar term such as "microscopic" would only be included when other features of a more distinct meaning had been selected.

In some cases it was deemed necessary to adapt this procedure to particular circumstances. A full description of these adaptations is given here.

Some of the concepts had obviously been ambiguous, and in these cases only those features concerning the predominant sense of the word (i.e. that with which most features were concerned) were considered. Thus, for "gum" the sense of "sticky substance" was used, and "mug" was taken to mean "a type of cup". In one case, the compound "weapon clamp", only 6 features were generated in total. To make up for this 2 extra features were added from each of the constituent lists. In the case of "daisy wheel" none of the features listed by any of the subjects implied that the subjects knew the real reference of the phrase. It was decided that this lack of recognition was likely to be unfortunate coincidence rather than an accurate representation of the common understanding of the phrase. It was therefore assumed that subjects in the latter part of the investigation would be familiar with the term, and would be very confused at not having any appropriate features to rate for this item. Accordingly this concept was resubmitted to different subjects, who affirmed that they felt they knew the meaning of the term "daisy wheel". The features generated in this way were selected according to the procedure described above.

When the lists were completed, the concept headings, along with the appropriate feature lists were compiled into questionnaires in the same random orders as the concept headings had been for stage 1. The features within each list were reordered randomly. Each questionnaire therefore contained a front page of instructions and rating scale, followed by 12 test pages. Each of these test pages was headed with a concept name, and had 30 features listed below in random order. Next to each feature was a short dotted line where the rating was to be recorded, and at the foot of the page the rating scale was reprinted for ease of reference.

Design

Again, the design with respect to subjects differs from that with respect to materials. A 4 (Compound Composition) x 2 (Extensional Status) x 3 (Compound Part) x 3 (Feature Set) design, was used, with repeated measures on the last 3 factors for subjects, and on the last two factors for materials.

The Compound Composition and Extensional Status factors are as in experiment 2. The factor feature set referred to the origin of the features used. The 10 features derived from the noun1 constituent listing in stage one of this study comprised feature set 1, the features from noun2 were feature set 2, and those from the compound comprised feature set 3.

Procedure

Subjects were either posted the questionnaire and asked to complete it according to the given instructions, or approached in person. All subjects completed the questionnaire in their own time. The following instructions to subjects were printed on the front page of each questionnaire.

"Instructions

At the top of each of the following pages is a different phrase, which is printed in **bold**, each is made up of either one or two words. Each phrase names a type of object, and below each is a number of possible attributes. I want you to rate these attributes as to how characteristic they are of the named object. In doing this, you should decide if the attribute describes this object, and if it could be used to decide whether or not something is this kind of object.

Some of the objects might not actually exist in the real world. If this is the case, then try and imagine what the object might be like in some imaginary world, and list its attributes as though it did actually exist.

Please rate each attribute according to the following scale (just write the letter on the space provided to the left of each attribute).

N = necessarily true for all possible examples

A = a very important part of the definition

B = a fairly important part of the definition

C = typically true but not defining

D = not usually true

E = very unlikely, but possible

I = impossible for any example

You need not take long over this, but please try to accurately give your own opinion.

The scale is printed at the bottom of each page for easy reference."

This is an adaptation of the scale used in Experiment 1, following Hampton (1987). The extremes of this scale (i.e. "necessary" and "impossible") are mapped to extended numeric values. This is in accordance with the principle that necessity, and its inverse, impossibility, should not be considered as simple continuations of probability, but rather that they imply a distinct order of difference from other sorts of possibility. The use of extended extremes here is intended to be a simple way of approximating this situation, rather than implying a specific quantitative description of the relationships between necessity, impossibility and intermediate probabilities. It was also felt that the scale in experiment 1 didn't allow sufficient expression of the status of features which were deemed not to be valid descriptions of a concept. Given that many features would, in the context of this experiment, inevitably fall into this category, the rating option "E = very unlikely, but possible" was added.

Results

Analysis of data

The ratings given to each item were converted to numeric data. This was carried out in a similar fashion to that used in Investigation 1, observing the pattern of extended extremes suggested by Hampton (1988). The following mappings (table 3.4) were used :

Table 3.4 Numeric mappings used

N → 8
A → 6
B → 5
C → 4
D → 3
E → 2
I → 0

Treatment of Results

The absolute differences between ratings for features with respect to the constituents and compounds were calculated using the same function as in experiment1. This is taken as a measure of noncompositionality.

Following Clark (1973), two analyses are conducted for each set of data. One analysis treats subjects as a random factor, and collapses over items, the other treats items as a random factor, and so collapses over subjects. The F values of these analyses are known as F1 and F2 respectively. Significant values of F1 imply that results are generalisable across the subject population, and significant values of F2 imply that results are generalisable to the population of items. However, to ensure that an effect is generalisable to both subjects and items a further statistic is needed. The quasi F-ratio (F') fulfils this role, but because of the computational complexity of F', its simpler and more conservative relation Min F' is used in this and subsequent experiments[8]. Whenever Min F' is significant, it is reported. F1 and F2 are reported if either of them is significant and Min F' is not significant.

Analyses of variance were conducted for the raw data, the absolute difference scores, and for the signed difference scores. Means and analyses for the raw data and signed difference scores are given in the appendix, tables A3.1 - A3.6. The absolute differences analysis is reported here.

Absolute Differences Analysis

The size of the differences between the ratings with respect to noun1 and the compound and respect to noun2 and the compound were calculated. In the materials analysis this could be done for each feature, since features were common to each compound part. In the subjects analysis the difference between

---

[8] Only the items analysis (collapsing across subjects) was conducted for earlier experiments (1 and 2) because the design of the experiments did not allow the computation of subjects analyses.

the mean of the ten features of each feature set was taken, and the differences calculated from these means. This was necessary since the compound parts rated by subjects were from different items, and therefore the features of each compound part were different.

Absdiff1 = Absolute difference (c - n1), Absdiff2 = Absolute difference (c - n2).

A 4 (Compound Composition) x 2 (Extensional Status) x 2 (Absdiff Part) x 3 (Feature Set) analysis of variance was conducted for this data, with repeated measures on the final three factors by subjects, and the final two factors by materials.

Compound Composition (min F' (3,80) = 5.12, $p < 0.01$), Absdiff Part (min F' (1,63) = 34.89, $p < 0.01$) and Feature Set (min F' (2,107) = 4.44, $p < 0.05$) were significant by min F'. Compound Composition x Feature Set (F1 (6,86) = 7.06, $p < 0.01$; F2 (min F' (6,80) = 2.36, $p < 0.05$) and Compound Composition x Extensional Status x Absdiff Part (F1 (3,43) = 14.55, $p < 0.01$; F2 (3,40) = 2.74, $p < 0.05$) were significant by F1 and F2 independently, but not by min F'.

Inspection of Table 3.4 shows that the main effect of Compound Composition is due to the overall differences of the AN and NA compounds being greater than those of the AA and NN compounds. Also, NN differences appear slightly greater than AA differences. The main effect of Absdiff Part can be seen in the same table. Clearly the difference in ratings of features with respect to noun1 and the compound is much greater than the difference between the ratings with respect to noun2 and the compound.

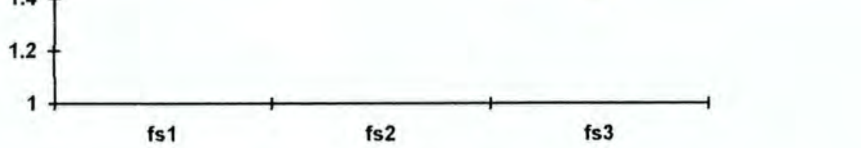
Table 3.4 : Compound Composition x Extensional Status x Absdiff Part

	Absdiff1		Absdiff2		Mean
	Real	Novel	Real	Novel	
AA	7.044	5.410	3.010	3.554	4.883
AN	8.372	10.480	4.074	3.227	6.538
NA	9.083	6.505	5.576	4.700	6.466
NN	8.450	6.400	3.132	4.846	5.707
Mean	8.237	7.199	4.077	4.082	
Mean	7.718		4.079		5.899

Table 3.5 : Compound Composition x Feature Set

	fs1	fs2	fs3
AA	1.623	1.707	1.554
AN	2.203	2.035	2.303
NA	2.301	2.340	2.061
NN	1.945	2.283	1.481
Mean	2.018	2.091	1.850

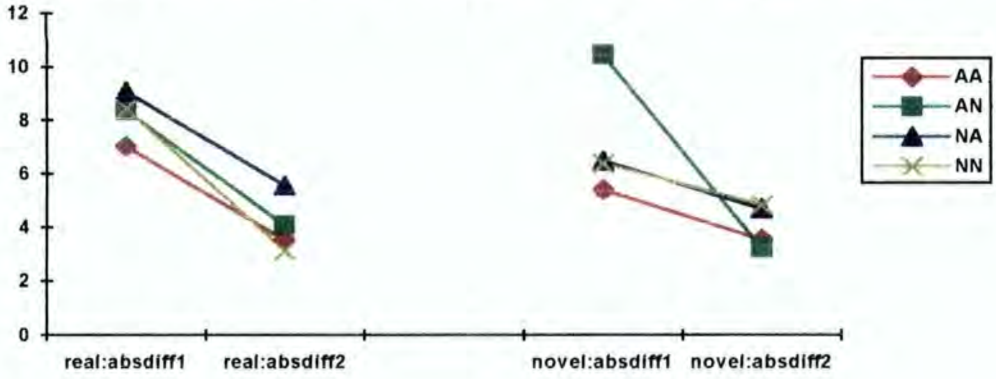
Figure 3.2 : Compound Composition x Feature Set



The main effect of Feature Set can be seen from Table 3.5. Feature Sets 1 and 2 have similar absolute difference scores, whereas the scores for Feature Set 3 are smaller.

The interaction of this Feature Set x Compound Composition is shown in Figure 3.2 (means in Table 3.5). This figure shows that the effect of Feature Set holds for all compounds apart from AN. For AN compounds Feature Set 2 has the smallest absolute difference scores.

Figure 3.3 : Compound Composition x Extensional Status x Absdiff Part



The interaction Compound Composition x Extensional Status x Absdiff Part is shown in Figure 3.3 (means in Table 3.4). Here it is clear that although all Compound Compositions have a similar difference between the two Absdiff Parts in the real condition, the novel condition does show an interaction. In the novel cases the AN compounds show a much greater difference between Absdiff Parts than do the other compounds. Also Absdiff1 is generally larger in the real condition than in the novel condition, except for the AN compounds.

Discussion

The principle motive for this experiment was the investigation of noncompositionality, and as such results with direct bearing on this will be discussed primarily. Results of the analysis carried out on the raw data (i.e. data before computation of absolute differences) will be discussed where relevant to the noncompositionality analysis. However, the discussion will mainly focus on the absolute difference scores.

It had been predicted that real compounds would show more noncompositionality than would novel compounds. In fact the results showed that there was no difference between the two groups. Interpretation of this result is not straightforward. The simplest explanation would imply that neither extensional feedback nor conventionality effects cause noncompositionality. However such a reading would be counter to intuition. Recalling such examples as "phone card" and "wooden spoon", it can confidently be stated that a process of extensional feedback really does operate in at least these cases. Moreover, the very simple nature of this effect suggests that it is unlikely to be limited to such a small set of examples, it should occur in many other cases too. It is also worth noting that any such effect operating on any compound is likely to be quite small. In this study, 30 features were used to describe each item. If extensional feedback did occur, but only with respect to one or two features in each case, then any resulting effect would be small. Some effect operating in the opposite direction would not need to be on a large scale to eliminate evidence of extensional feedback.

Hampton (1991) refers to a process which would seem to be a likely candidate for causing just such an effect. In a pilot study that he ran, he asked subjects to draw pictures of unlikely sounding things, the names of which he had made up by putting together pairs of semantically unrelated nouns, e.g. "A teapot that is a computer" or "A fish that is also a vehicle". The pictures produced included many emergent features. Hampton attributes this high degree of noncompositionality to `conflict resolution'. That is, the bizarre combinations required of the subjects necessitated that incompatibilities within the pairs of concepts be accommodated. In many cases this was achieved by considerable elaboration of the original concepts. This effect would seem in fact to be the same as the `maintenance of coherence' effect referred to previously. It seems reasonable to suppose that such a process also occurred in the present experiment. It is therefore quite conceivable that two effects, extensional feedback and maintenance of coherence both occurred in this study, but neither led to a significant effect as they cancelled out one another by operating in opposing directions.

In spite of this potential counter balancing effect, the data can in places be seen to suggest some of the differences between real and novel compounds which might be expected. Overall absdiff1 was larger for real compounds than for novel ones, while absdiff2 was more or less constant across the two conditions (see Table 3.4). In fact, were it not for the anomalous behaviour of the AN compounds this difference would be very considerable. This would suggest that while both real and novel compound inherit a great deal of their meaning from the noun2, novel compounds (apart from AN compounds) have more meaning in common with noun1 than do their real counterparts.

One clear trend in the data is that the heterogeneous compounds (AN and NA) show more noncompositionality than do the homogeneous types (AA and NN). This is attributed to a general commonality of features between both constituents and the compound in the homogeneous cases, whereas in the heterogeneous cases this commonality of features is restricted to the compound and one of its constituents. The fact that there is a difference between these two types is not seen as resulting from a different type of combination between homogeneous and heterogeneous compounds. Rather it is simply that features derived from one kind of concept are more likely to be apt descriptions of another concept of the same kind, than they are of a concept of a quite different type, e.g. the features of the concept 'lion' - 'is furry', 'runs', and 'eats meat' - will be more likely to be applicable to another natural kind, than to an artefact. Since homogeneous compounds will almost certainly be of the same general category as both of their constituents, they will exhibit this commonality to a greater extent than the heterogeneous compounds, which will be of a different kind to one of there constituents. This difference between homogeneous and heterogeneous compounds is also evident in several other results. It is particularly clear in the main effect of Compound Composition in the raw scores (see appendix, table A3.1).

Returning to the issue of apparent noncompositionality, it had originally been predicted that the AN compounds would be less noncompositional than the NA compounds, due to the supposed inhibition of maintenance of coherence inferences in AN compounds. Clearly this prediction was not borne out, there being virtually no difference between the two with respect to mean absolute difference scores (mean absdiff: AN = 6.446, NA = 6.466). However, underlying this there is also a significant interaction between Compound Composition, Absdiff Part and Extensional Status (see fig 3.3). In the real condition no clear difference emerges between the Compound Compositions, but in the novel condition the difference between Absdiff Parts 1 and 2 is much more extreme in the AN case than is the case for the other compounds. In fact, Absdiff Part 1 of the AN compounds is much higher than that of other compounds, while Absdiff

Part 2 of the AN compounds is the lowest. Evidently the apparently high level of noncompositionality of AN compounds is based on the very low amount of meaning common to the noun1 and the compound. If only the relationship between the noun2 and the compound is considered, then it can be seen that these two are closer in meaning than is the case in other compounds. Therefore, rather than outright rejection of the original prediction with respect to these results it seems more appropriate to modify it. It appears that when confronted with AN compounds subjects do, as predicted, find them difficult to interpret. However, rather than becoming more liberal in their interpretation as a consequence of this difficulty of interpretation, subjects appear to react by adhering more rigidly to the standard grammar of such compounds. As a result, AN compounds are very close in meaning to the noun2, but they have very little in common with the noun1.

There were two other significant main effects within the analysis. That of Absdiff Part is straightforward to account for. Absdiff Part 1 (7.672) is greater than Absdiff Part 2 (3.991). This result is in accord with the expectation that the compounds should be right headed, there being less difference between the meaning of noun2 and the compound than there is between noun1 and the compound. The final main effect is that of Feature Set. This effect shows that noncompositionality of the compounds is lower with respect to Feature Set 3 than Feature Sets 1 and 2. This demonstrates the principle that the apparent compositionality of compounds is dependent upon the features by which they are investigated. It seems that compounds will appear more compositional when compositionality is measured using features of the compound than of the constituents.

A final issue is the different results in this experiment from those in experiment 1. Both this experiment and experiment 1 found that novel AN compounds behaved differently to other compounds. In experiment 1 the AN compounds showed a big difference in meaning from noun2, yet in the present experiment they should a big difference in meaning from noun1. There are several differences between the two experiments which might account for this. Firstly, the context in which subjects carried out the experimental task differed in that there were many more types of compound present in experiment 3 questionnaires. Secondly, the rating scale changed, with the addition of an extra choice. Both of these could have altered subjects responses. However, perhaps the most significant difference was in the features presented to subjects. As shown by the significant effects involving Feature Set, the features used in experiments such as these can make critical differences. Although there is no ideal set of features for such a comparison, perhaps the more tightly controlled selection of features in the current experiment provides the more suitable range of features.

In summary, therefore, several conclusions are drawn. Firstly, that there seems to be little difference between real and novel compounds with respect to how much meaning they have in common with their constituents. Such differences as do exist between these two types of compound are confined to artefact natural kind compounds (of the compound types studied here) and seem to result from the real compounds having a well known meaning while the novel compounds can be difficult to interpret. It is concluded that novel artefact natural kind compounds are probably more difficult to interpret than other such combinations. This seems to cause stricter adherence to the standard grammar of noun noun compounds. Secondly, any demonstration that extensional feedback and maintenance of coherence inferences contribute substantially to noncompositionality has proved hard to come by, since it was concluded that both of these may have been operating, and so cancelled each other out.

This experiment also provides further evidence that compounds have more in common with noun2 than noun1. This is interpreted as supporting the assumption that noun2 will tend to act as the head of a compound. Finally, the results with respect to the factor Feature Set lend support to the concern that care must be taken to ensure that a properly representative range of features is used, if comparisons between the meanings of compounds and their constituents are being made. Results using features which were listed for the compounds are different to those which include features listed for the constituents.

Experiment 4 : Feature Ratings with Reversed Compounds

It is difficult to discount the possibility that the data from experiment 3 might represent materials which contain inherent imbalances. This experiment therefore uses the same constituent nouns to construct novel compounds in the reverse order to that used in experiment 3. Therefore AA and NN compounds are reversed but remain the same compound composition types, AN and NA compounds are reversed and exchange compound compositions with each other. The design and procedure of this experiment are along the same lines as experiment 3. This study compares the data from the novel condition of experiment 3 with data from a new set of novel compounds.

Method

Again, this was a two stage study; feature generation and then feature rating. For the feature generation stage however most of the necessary data was already available. Feature sets 1 and 2 for each item were the same as those in experiment 3 (although they swapped places because of the constituents swapping position), features therefore had to be generated for feature set 3 only.

Part 1 : Feature Generation

This stage is similar to experiment 2. Two lists were compiled such that each contained twelve compounds, three of each compound composition. Four subjects for each list were each asked to provide features of the compounds, which were presented in random orders in paper questionnaires. The same instructions were used as in experiment 2.

Part 2 : Feature Rating

Subjects

Twenty-four unpaid volunteers participated in this experiment. They were all undergraduates studying psychology at the University of Durham. There was an approximately equal number of males and females, aged 18 - 30 years.

Materials

Only novel compounds were used in this experiment, the nouns of each novel compound used in experiment 3 were reversed in order, so a compound "AB" became "BA".

Design

The experiment was of a 4 (Compound Composition) x 2 (Word Order) x 3 (Compound Part) x 3 (Feature Set) design, with repeated measures on the last two factors by both subjects and materials. The factor Word Order is composed of Word Order 1, from experiment 3, and Word Order 2, new data from this experiment.

Procedure

The procedure used was the same as that in experiment 3 above.

Results

As with experiment3, tables of means and analyses for raw and signed difference scores are to be found in the appendix (tables A4.1 - A4.6).

Absolute Differences Analysis.

Absolute difference scores were calculated as in experiment 3. A 4 (Compound Composition) x 2 (Word Order) x 2 (Absdiff Part) x 3 (Feature Set) analysis of variance was conducted for this data, with repeated measures on the final two factors.

Compound Composition (min F' (3,93) = 8.75, $p < 0.01$), Absdiff part (min F' (1,77) = 20.19, $p < 0.01$) and Feature Set (min F' (2,99) = 10.80, $p < 0.01$) were all significant main effects.

Compound Composition x Absdiff Part (min F' (3,75) = 3.21, $p < 0.05$) and Absdiff part x Feature Set (min F' (2,198) = 9.43, $p < 0.01$) were significant interactions. Word Order x Absdiff part x Feature Set (F1 (2,174) = 4.42, $p < 0.05$; F2 (2,80) = 5.24, $p < 0.05$) was significant by F1 and F2 independently, but not by min F'.

Table 3.6 shows that as with experiment 3 the main effect of Compound Composition is caused by the heterogeneous compounds having higher differences than the homogeneous ones. In this case however, NN compounds also have considerably larger differences than AA compounds. The main effect of Absdiff Part is also clear from this table. Absdiff1 is again much greater than Absdiff2.

Table 3.6 : Compound Composition x Absdiff Part

	abs1	abs2	mean
AA	4.985	4.695	4.840
AN	8.682	4.516	6.599
NA	7.517	5.658	6.588
NN	7.311	4.999	6.155
	7.124	4.967	6.046

Fig 3.4: Compound Composition x Absdiff Part

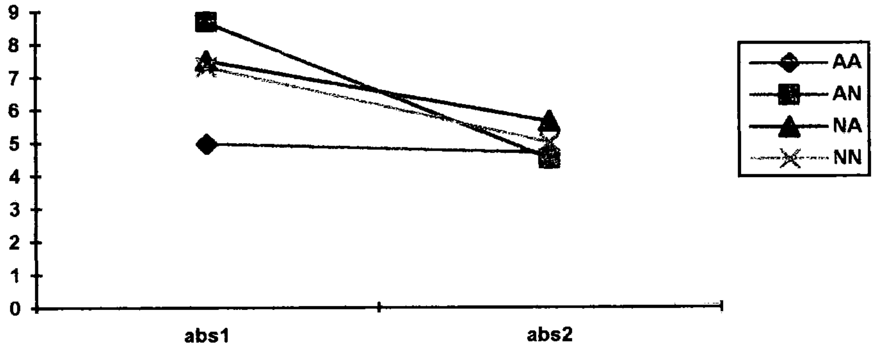
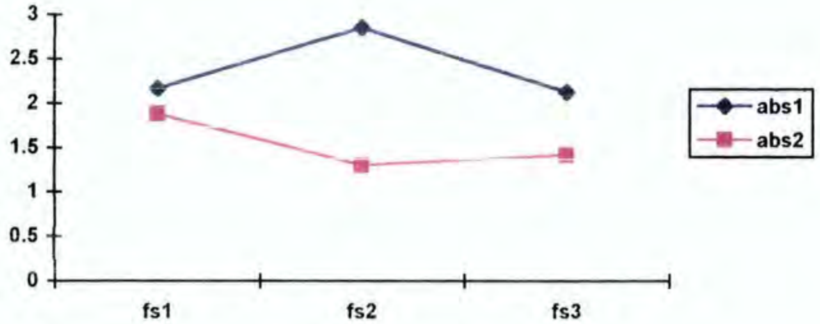


Fig 3.4 shows that the AN compounds again have the highest absdiff1 and lowest absdiff2, causing the interaction between Compound Composition and Absdiff Part.

Fig 3.5 : Feature Set x Absdiff Part



The effect of Feature Set is similar to that of experiment 3, feature sets 1 and 2 are approximately equal, but feature set 3 is lower, as shown in Table 3.7. The interaction between Feature Set and Absdiff Part is illustrated in fig 3.5 It is apparently due to Feature Set 2 which has the highest of the three Absdiff1 scores, and the lowest of the Absdiff2 scores.

Table 3.7 : Word Order x Feature set x Absdiff Part

	FS1		FS2		FS3		Mean
	Abs1	Abs2	Abs1	Abs2	Abs1	Abs2	
WO1	2.418	1.576	2.684	1.261	2.097	1.428	5.732
WO2	1.919	2.195	3.033	1.346	2.172	1.425	6.045
	2.169	1.886	2.859	1.304	2.135	1.427	5.890
	2.028		2.082		1.781		

Fig 3.6 : Word Order x Feature Set x Absdiff Part

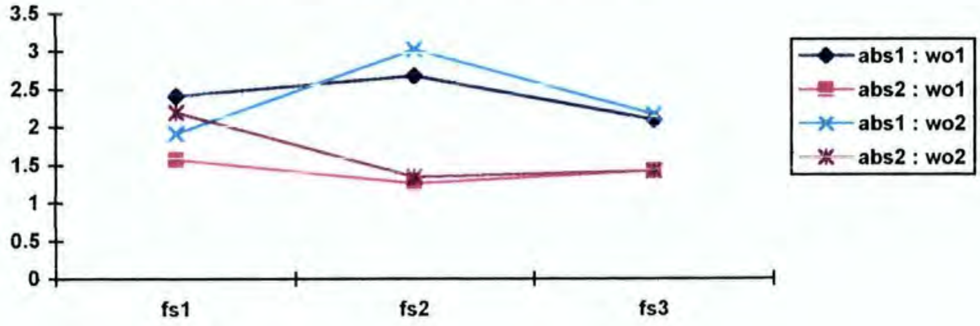


Fig 3.6 shows the interaction Word Order x Feature Set x Absdiff Part. The interaction is caused by Feature Set 1 in Word Order 2, for which Absdiff Part 1 is greater than Absdiff Part 2, counter to the main effect of Absdiff Part.

Discussion

This study was conceived principally in order to asses the generalisability of results in experiment 3. The factor critical to this question is that of Word Order. Word Order was a significant main effect. However, the only significant interaction in which it participated was Word Order x Feature Set x Absdiff Part (fig 3.6). In this interaction Absdiff1 is less than Absdiff2 for Feature Set 1 of Word Order 2. In all other conditions Absdiff1 is greater than Absdiff2. There is no account in terms of compound structure for this finding, since there is no principled difference between the two sets of compounds and subject groups. All that can be said is that this finding is quite possibly unrepresentative. The difference between the two scores is small (WO2 FS1 Absdiff1 = 1.919, WO2 FS1 Absdiff2 = 2.195), and could have arisen by chance. As was concluded in the discussion of experiment 3, findings concerning just a limited set of features can be quite misleading.

There is no interaction of Word Order in other effects. The main effect of Word Order itself seems to be due simply to the second group of subjects rating everything slightly higher than did the first subject group. This could be due to the difference in the range of materials presented to the two groups. In Word Order 1 (experiment 3) subjects were asked to rate features with respect to both real and novel compounds. This factor was not present in the Word Order 2 condition. Word Order does not interact with any other effects, and experiment 4 can therefore be seen as dispelling the notion that significant results between Compound Compositions in experiment 3 might have resulted from initial chance biases in the materials.

Chapter four: Slot Filling and Property Mapping

Introduction

Previous experiments have investigated the amount of meaning, in terms of features, common to a compound and its constituent concepts. This chapter departs from a purely quantitative approach, and instead looks at the types of grammatical relationship which exist between the two nouns, and how this relationship may be influenced by a number of factors. As discussed in chapter 1, several researchers have proposed descriptions of the types of combination which are found in noun noun compounds. Levi (1978) and Shoben (1991) provide a number of possible relations which may be asserted to exist between the two constituent concepts. These descriptions can be likened to the more psychologically descriptive models of Smith et al (1988), and Murphy (1988). Underlying all of these seems to be the assumption that modifiers fill slots in the meanings of head nouns. In other words, the relationship between the modifier and the head noun is given by the slot that it fills. For the compound **car door** the slot 'part of' in the representation of **door** is filled with the concept **car**, and so a whole-part relationship is given. A number of problems are associated with this kind of account, several of which were discussed in chapter 1. This chapter addresses one of these problems - that slot filling may characterise only a proportion of noun noun compounds, and is not at all an appropriate description of the remainder.

Wisniewski and Gentner (1991) made up a large number of novel noun noun compounds by combining artefact, natural kind and mass nouns. Subjects were asked to write down descriptions of the most likely meanings of these compounds. Postgraduate judges classified these definitions as to whether or not they described a slot filling relation. Only 40% of the definitions were found to be slot filling relations.

Although it would be unwise to assume that this figure is a reliable measure of the proportion of slot filling among novel compounds as we use them, it does suggest that a theory of noun noun compounds ought to have slot filling as only one of its components. There are several questions which this finding provokes. Firstly, if slot filling fails to account for the full range of noun noun compounds, what other relationship or relationships make up the rest ?

Wisniewski and Gentner report that two other types of relationship in particular were present in their data, what they called property mapping and structure mapping. Property mapping is said to occur when a single feature or property is mapped across from the modifier to the head noun. For example, in the compound **robin snake**, the property of having a red front is mapped to the head noun, resulting in the compounds meaning - `a red-fronted snake'. Structure mapping is along the same lines, although it is not a single property which is mapped from the modifier to the head, but a relationship between multiple properties. Wisniewski and Gentner give the example pony chair, which was interpreted as meaning `a small chair'. The authors suggest that this cannot simply be a property mapping interpretation, since if the size of the chair became equal to the size of the average pony, then it would actually be a big chair, as far as chairs go. Wisniewski and Gentner propose that what is mapped is the relation between the size of ponies and the size of their superordinate category, horses. This relation is mapped onto the corresponding relation of the compound, i.e. the relative size of pony chairs compared to their superordinate category, chairs. Thus, ponies are somewhat smaller than horses and pony chairs are somewhat smaller than chairs. Experiment 5 investigates the proportion of slot filling, property mapping and structure mapping interpretations in a range of noun noun compounds.

Secondly, what are the conditions which favour each of these types of interpretation? Any of several factors could have an important bearing upon which type of compounding is favoured. Clearly context would have an important role to play. Interpreting a compound in a particular context is likely to bias towards a particular meaning. For example:

> A small island is the only home to a number of rare species. In particular, conservationists are concerned about the decline of the local robin population. Its numbers have declined rapidly since the accidental introduction of robin snakes from the mainland.

Given this context it seems likely that most readers would arrive at a slot filling interpretation, that is something like, "a snake that eats robins", counter to the property mapping interpretation made out of context that Wisniewski and Gentner report. However, although a powerful context may be able to override other influences, this does not imply that interpretations will just randomly vary in the absence of such contexts. It may be that each noun has a preference for the type of compounds it will enter as a head noun or modifier, as suggested by Gagné and Shoben (1995). There could be more complicated interactions, such as in Murphy's concept specialisation model (1988), or analogical influences as suggested by Ryder (1994). Wisniewski (1996) reports that he has identified a factor which does have such an influence. He found that noun noun compounds which had similar constituent concepts (e.g. **tie scarf**) were likely to have property mapping interpretations, and compounds with dissimilar constituents (e.g. **fork scarf**) tended to have slot filling interpretations. Another plausible influence is the number of salient features a constituent has. Property mapping would seem to occur only when a highly salient feature of the noun1 can be mapped over to the noun2. The potential for this to happen will vary depending upon the salient properties of a particular pair of constituents. Experiment 6 looks at the influence that constituent similarity and number of salient properties have upon slot filling and property mapping.

Experiment 5 : The Extent of Property Mapping and Slot Filling

Introduction

This experiment compares the four compound compositions used in earlier studies, composed of artefacts and natural kinds, to see if any differences arise between them. These are the same sorts of compounds as used by Wisniewski and Gentner. However, the design of their study seems not to have allowed a proper analysis of what influence the factor of compound composition may have had on the type of interpretation given to compounds. If such an analysis is possible for their data, they appear to have chosen not to carry it out, because none is reported. It has been suggested that there are important differences between the representations of artefact and natural kind nouns. If these suggestions are well founded then it would not be surprising to find differences between the different compound compositions. However, since the proposed differences are somewhat ill defined, it's difficult to make a strong prediction about the type of differences that might arise in compounds.

One prediction that is quite straightforward is that homogeneous compounds (AA and NN) should have more property mapping than heterogeneous compounds (AN and NA). This is because it can reasonably be expected that the properties of any concept will be more likely to be applicable to another concept of the same general kind than to a radically different concept. For example, a

wolf is regarded as a voracious predator. While such an attribute may be true of a **wolf fish**, it is unlikely to be true of a **wolf door**. Data from this experiment will also provide an opportunity for comparison with Wisniewski's findings about the effect of constituent similarity.

Finally, why have non-slot-filling interpretations been overlooked until recently? Given Wisniewski and Gentner's findings it is rather puzzling that so many people interested in noun noun compounds have overlooked non-slot filling interpretations. Perhaps the reason lies in a difference between real and novel compounds. Wisniewski and Gentner only looked at novel compounds, so as yet we are unable to say whether or not slot filling can account for all real compounds. If it can then this would provide some explanation for the attention paid to slot filling. Of course, any such difference between real and novel compounds would require an explanation itself. Real and novel compounds are included here to compare the types of constituent relations in each.

Overall experiment 5 is similar to Wisniewski and Gentner (1991), but several changes have been made, such as the inclusion here of the real/novel comparison. The other principle difference is that Wisniewski and Gentner obtained only a single definition for each compound. This feature of their study would seem to make it vulnerable to idiosyncratic responses, and the results difficult to interpret, so by contrast eight subjects define each item in experiment 5. The Wisniewski and Gentner study did include very many materials (400 compounds), so obtaining multiple definitions is at the cost of reducing this number. The number of materials is further reduced by omitting mass noun compounds.

The design of Experiment 5 includes categorisation of paraphrases with respect to which of the compound's constituent nouns is the referent of the paraphrase. It is commonly assumed that noun2 will be the referent. Wisniewski and Gentner tested this and found that 38% of their paraphrases did not accord with this assumption. However, it's not clear what proportion of this 38% might be exocentric, and what proportion had noun1 as the referent, nor if there was any influence of compound composition on this effect. This is therefore another issue raised by the Wisniewski and Gentner (1991) study for which Experiment 5 attempts to reproduce a reported result, and also answer more specific questions about the nature of that result.

To summarise, experiment 5 looks at AA, AN, NA and NN compounds. The primary focus will be the extent to which slot filling, property mapping and structure mapping occur. Another result which will be looked at is which of the constituent nouns, if either, is the reference of the compounds, and whether this interacts with the experimental variables. Real and novel examples are compared for each compound composition, to see if this affects the type of relation used. A post-test of similarity between the constituents of each compound will be used to compare this data with that of Wisniewski (1996).

Method

Subjects

Thirty six unpaid volunteers participated in this experiment. They were all undergraduates studying psychology at the University of Durham. There was an approximately equal number of males and females, aged 18 - 30 years.

Materials

Each questionnaire consisted of a front page of instructions and two subsequent A4 pages. The two test pages listed twelve noun-noun compounds along with two dotted lines below each. Each questionnaire contained four *real* compounds (e.g. **oven glove**), four *novel-order1* compounds (e.g. **apron filter**) and four *novel-order2* compounds (e.g. **filter apron** - order2 compounds were simply order1 compounds with word order transposed within each compound). The four compounds presented to each subject in each of these conditions comprised each of the four possible compound compositions : artefact - artefact (e.g. **bottle candle**), artefact - natural kind (e.g. **computer mouse**), natural kind - artefact (e.g. **crow pencil**) and natural kind - natural kind (e.g. **tiger moth**). These twelve items were presented in random orders. A list of items is given in the appendix, table A5.1.

Design

This investigation was of a 3 (Compound Category) x 4 (Compound Composition) design. Both factors were within subjects.

Procedure

Stage 1 : Paraphrase generation

Subjects were presented with the questionnaire described above, and asked to write beside each of the noun-noun compounds a definition or paraphrase of what they understood the meaning of that compound to be, verbatim instructions are given below.

"On each of the following pages are six different phrases, which are underlined and enclosed in separate boxes, along with some dotted lines. Each phrase is made up of two words, and names a type of object. I want you to write a *definition* of each of these, that is, write briefly and in your own words, another way of saying what you understand by this phrase. Please write this on the dotted lines directly below the relevant phrase.

Some of the objects referred to might not actually exist in the real world. If this is the case, then try and imagine what the object might be like in some imaginary world, and write a definition of it as though it did actually exist. You need not take long over this, but please try to give your own opinion accurately and clearly. Please do not leave any blank, if you feel that some of the phrases are not possible to provide a definition for, then just write down your best attempt."

No time limit was imposed, and subjects were left to complete the questionnaire unsupervised.

Stage 2 : Reference Categorisation

The paraphrases were categorised according to two criteria. Firstly, which of the two nouns was the referent of the compound, and secondly whether or not the first noun modified the second. These judgements were made independently by two cognitive science postgraduates, and in cases where these judgements were in disagreement, the two judges were asked to come to a consensus opinion. The judges were given the instructions page shown in fig 4.1.

Fig 4.1 : Instructions to Reference Judges.

This booklet contains the responses given by subjects who were asked to define or paraphrase some noun-noun phrases. Some of the phrases are commonly known to English speakers (such as *traffic light*), others are quite novel (such as *beetroot clock*). I would like you to make judgements about these responses. There are two particular questions that are to be answered.

i) What is the referent of the phrase ?

By this question I mean that I want to know what kind of thing the subject had in mind in giving his/her response. For example, if the stimulus was "*beetroot clock*", and the subject wrote

"A clock which has a deep purple colour"

Then I would expect you to record that the referent was "*clock*" But you should do this by ticking the appropriate box on the grid provided, like this,

Referent:	beetroot		clock	✓	both		neither	

because "clock" is the second noun of the phrase "beetroot clock". Had the subject responded something like,

"A root vegetable which has a round shape, like a clock's face"

then I would expect you to indicate that the referent was "*beetroot*" by ticking the first box. You should beware that you are sure which noun is which in each case, because many of the noun-noun phrases used are reversals of one another.

If you think that the subject genuinely thought the noun-noun phrase referred to something which was simultaneously a beetroot and a clock, e.g.

"A genetically engineered root vegetable which is also a clock"

then you should tick the box for "both". If the subject's response seems to refer to neither the first noun, nor the second, e.g.

"A kind of hat"

then you should tick the box for "neither".

The second question that should be addressed is,

ii) Does the first noun modify the second noun ?

This simply requires a "yes" or "no" answer (please tick the appropriate box). If the subject's response implies that some aspect of "*beetroot*" is being asserted of "*clocks*", then answer "yes". This could cover a whole range of possibilities, so don't dismiss it too quickly. If no aspect of the first noun is being asserted of the second one in the subject's response, then you should tick the "no" box. Please be careful to select the answers you feel are most appropriate. Another person is also carrying out this task and I will be asking you to come to agreement with one another about any differences of opinion.

Stage 3 : Modification Classification

Those paraphrases which indicated noun2 to be the head noun (i.e. noun2 is the reference) *and* in which noun1 modifies noun2 were further categorised. Again, the judges were two postgraduate cognitive scientists (different individuals from those who had conducted the first categorisation task). They were asked to indicate what form of modification noun1 made to noun2 in each case. The categories into which judges were able to assign these paraphrases are based upon the possible modification types discussed by Wisniewski and Gentner (1992), i.e. slot filling, property mapping and structure mapping. In addition to these, judges could also categorise a paraphrase as 'appearance mapping' or 'other'. The 'appearance mapping' category is included to cover those cases in which the paraphrase reads something like "an X which looks like a Y", which seemed a probable paraphrase, but which doesn't naturally fit into any other category. The 'other' category is simply present so that judges are not forced to put any paraphrase into a category which it doesn't really fit. Again, differences of opinion were resolved by consensus. Full instructions are given in fig 4.2.

Fig 4.2 : Instructions to judges for the Modification Classification Stage

The following pages contain definitions which people have given for various concepts. Each concept is a noun-noun compound (e.g. book shelf). Some of them really do exist, others have been made up, but you shouldn't worry about this difference.

Your task is to categorise the type of relationship which exists between the two nouns of the compound, as described in the given definition. You should do this by ticking the appropriate box. Here is an explanation (with some examples) of the categories you should choose between.

- Noun1 fills a slot

If the first noun (noun1) *itself* describes a simple, single attribute of the second noun. The terminology here comes from the idea of each concept having a template of attributes, and noun1 is itself the value of one of these.

e.g. phrase: beetroot clock

"A clock made of beetroot"

Seems to mean that the value of the slot "is made of" is beetroot, so you should tick the first box, like this.

Noun1 fills slot	Noun1 provides one property	Noun1 provides structure	Noun1 provides appearance	Other
√				

- Noun1 provides one property

This is the case if a *single property* of noun1 is apparently being asserted of noun2. e.g. if beetroot clock were described as "A deep purple clock", then you would probably want to tick this box.

- Noun1 provides structure

This is the case if, rather than a single property being 'mapped' from noun1 to noun2, some *structural relation* of noun1 is asserted of noun2. e.g. if the concept pony chair were described as "a small chair" this would seem to be applying a structural relation of ponies to chairs, since ponies are not small things themselves, but they are small in relation to other horses, as presumably the chair is small in relation to other chairs.

- Noun1 provides appearance

You should pick this category if the entire appearance of noun1 is being asserted of noun2, and not if just a single attribute of appearance is intended. e.g. for the concept pony chair, the definition "a chair that looks like a pony" should make you tick this box.

- Other

If none of the above categories seem appropriate you should tick this box. If you choose this option, please give an explanation (i.e. say what kind of category it does fall into) on the line below.

Post-test of Constituent Similarity

A measure of similarity between the two constituents of each compound was also collected, using a similar method as that used by Wisniewski (1996).

Subjects

Eight unpaid volunteers participated in this experiment. They were all undergraduates studying psychology at the University of Durham. There was an approximately equal number of males and females, aged 18 - 30 years.

Materials

Each questionnaire was comprised of a single instruction page followed by the full set of test items. Test items were the pairs of nouns used in the noun noun compounds of the main study. Each pair of nouns was presented in both orders (so subjects were asked how similar A is to B, and then later in the questionnaire, how similar B is to A). The items were presented to half of the subjects in one random order, and to the other half in a different random order. The format of each question, and the instructions given to subjects were as follows:

"This booklet contains five pages - this instruction page followed by four test pages.

Your task is to judge the similarity between pairs of concepts, and to record your judgement on a scale of 1 (extremely dissimilar) to 7 (extremely similar), by circling the appropriate number.

E.g.

The following question asks you to compare the concept `tortoise' with the concept `book'.

{r3na8}

How similar is tortoise to book ? : 1 2 3 4 5 6 7

In this example, it was decided that books and tortoises are very dissimilar, and so that has been recorded by putting a circle around the "2". (By the way, you can ignore the code in curly brackets ({r3na8}), it just identifies the questions for me).

You should work through the questions at your own pace. None of the judgements should take long, but make sure that you do record what you feel is the most appropriate response, so don't rush through them. I want your own opinion, so please don't discuss these with other people.

Feel free to make use of the full range of the scale, from 1 right through to 7.

Please make sure that you record your judgement for all of the questions. It is important that none are missed out. If you find some of them difficult then just record what you feel is your best assessment.

Do make sure you get the scale the right way round !

extremely dissimilar	very dissimilar	dissimilar	neutral	similar	very similar	extremely similar
1	2	3	4	5	6	7"

Results

Reference and Modifier Categorisation

The judges originally gave the same opinion on 84% of the cases. After consultation the remaining 16% of cases were agreed. The resulting categorisations are shown in Tables 4.1 and 4.2 below. From Table 4.1 it can be seen that noun2 is highly preferred as referent in all cases. In all but one condition, noun2 is the referent in over 72% of the paraphrases. The exception is the real AN case, for which noun2 is only the referent on 47% of occasions. The same condition has the equal highest proportion of cases in which noun1 is the referent (14%), and the highest proportion of cases in which neither noun is the compound referent (39%). In the other three Compound Compositions (AA, NA and NN) the real condition has a vast majority of paraphrases with a noun2 referent, and no cases at all of noun1 being the referent. Thus the real condition can be seen to follow the same pattern as the novel condition, but more extremely. Overall, noun2 (including `both noun1 and noun2') was the referent in 81% of the cases.

Table 4.1 : Percentage of paraphrases in which each noun provides the reference

	noun1	noun2	both	neither	other
AA					
novel1	11	75	6	3	6
novel2	3	75	11	8	3
real	0	97	0	3	0
AN					
novel1	0	78	0	22	0
novel2	3	86	0	8	3
real	14	47	0	39	0
NA					
novel1	8	78	0	11	3
novel2	3	92	0	0	6
real	0	94	0	6	0
NN					
novel1	14	72	6	6	3
novel2	6	78	6	11	0
real	0	97	3	0	0
overall %	5	81	3	10	2

Table 4.2 : Percentage of paraphrases in which noun1 modifies noun2

AA	noun1 modifies noun2	noun1 does not modify noun2	other	missing
novel1	69	25	3	3
novel2	81	17	3	0
real	77	22	0	0
AN				
novel1	86	14	0	0
novel2	77	19	3	0
real	81	19	0	0
NA				
novel1	77	19	0	3
novel2	86	8	3	3
real	97	3	0	0
NN				
novel1	72	25	0	3
novel2	92	8	0	0
real	56	44	0	0
overall %	79	19	1	1

Table 4.2 shows that in all cases noun1 is likely to modify noun2. Overall there seems to be little difference either between Compound Compositions or real and novel conditions. The one result which may be slightly anomalous is the real NN case, in which noun1 modifies noun2 in only 56% of the cases.

Modification Type Classification

299 paraphrases (69% of the cases) were found to satisfy both the condition that they have a noun2 reference, and that noun1 modified noun2. These paraphrases were further categorised, as described in the method section.

Original agreement between the two judges occurred for 76% of the cases. The remaining cases were agreed after consultation. These classifications are shown in Table 4.3. The percentages of slot filling and property mapping are illustrated by Figure 4.3. Because novel1 and novel2 results are similar, they have been collapsed to a single novel condition in this figure.

Figure 4.3 : Percentages of slot filling and property mapping definitions

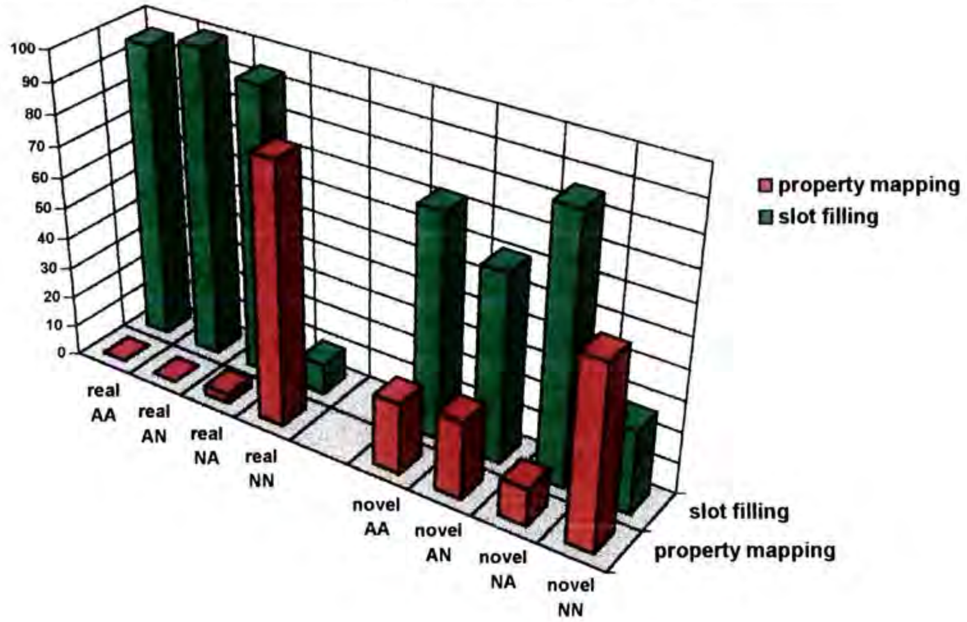


Table 4.3: Percentages of definitions in each modification type category

	slot filling	property mapping	structural relation mapping	appearance mapping	other	n
AA						
n1	83	9	0	4	4	23
n2	62	38	0	0	0	21
real	96	0	0	0	4	28
AN						
n1	80	12	0	0	8	25
n2	42	38	0	8	12	26
real	100	0	0	0	0	15
NA						
n1	85	12	0	0	4	26
n2	87	13	0	0	0	31
real	94	3	0	0	3	33
NN						
n1	35	61	4	0	0	23
n2	21	57	0	21	0	28
real	10	85	0	5	0	20
Overall %	67	26	0	3	3	299

Overall, slot filling occurred in 201 cases (67% of the sample), property mapping accounted for 78 of the definitions (26%). In the novel conditions slot filling occurred in 126 cases (62%) and property mapping in 60 cases (30%). In the real condition there were 75 cases (78%) which were slot filling, and 18 cases (19%) which were property mapping. Figure 4.3 shows that in the AA, AN and NA cases the slot filling relationship is most common, whereas in the NN composition the property mapping is strongly preferred. The difference between the proportion of slot filling to property mapping is more extreme in each of the real cases than it is in the novel cases. Binomial tests were carried out between slot filling and property mapping for each condition. The results of these are given in Table 4.4. The results in Table 4.4 confirm that the difference between slot filling and property mapping was significant in each condition.

Table 4.4 : Binomial tests between slot filling and property mapping for each condition

condition	result
real AA	$p < 0.00003$
real AN	$p < 0.00003$
real NA	$p < 0.001$
real NN	$p < 0.001$
novel AA	$p < 0.0007$
novel AN	$p < 0.0006$
novel NA	$p < 0.00003$
novel NN	$p < 0.0119$

Results of Post-test of Similarity

The mean similarity rating for each compound composition is given in Table 4.5. Scores were on a scale of 1 (extremely dissimilar) to 7 (extremely similar), and is illustrated by figure 4.4.

Figure 4.4 : Mean Similarity Rating for each Compound Composition

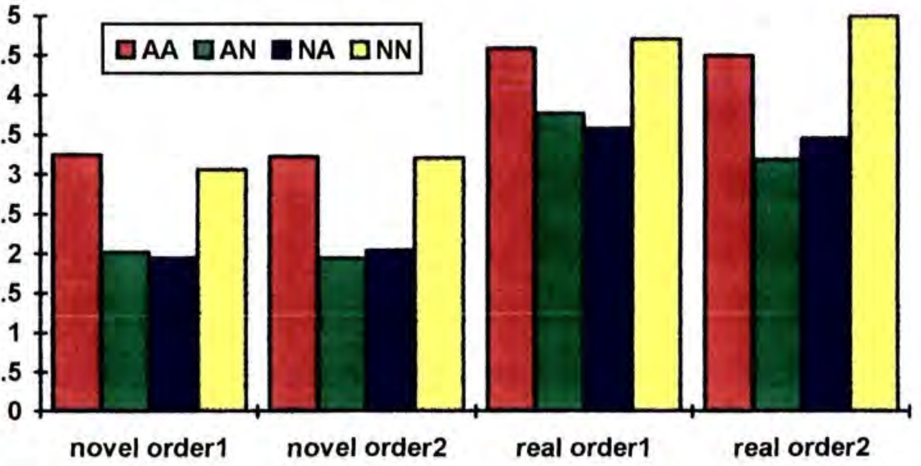


Table 4.5 : Mean Similarity Rating for each Compound Composition

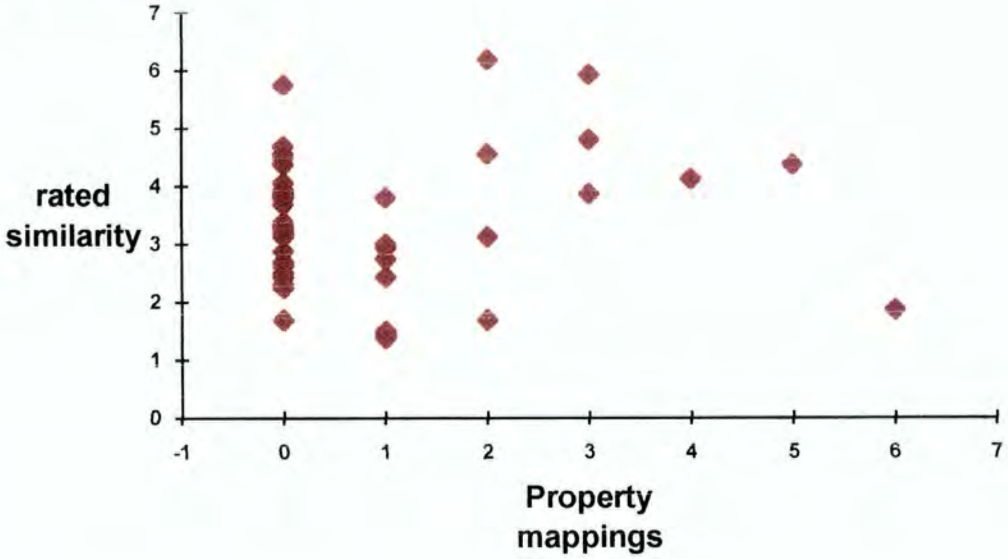
	AA	AN	NA	NN
novel order1	3.25	2.02	1.94	3.06
novel order2	3.23	1.94	2.04	3.21
real order1	4.60	3.77	3.58	4.71
real order2	4.50	3.19	3.46	5.00

As can be seen from figure 4.4, the constituents of the homogenous compound compositions (AA and NN) are more similar to each other than are those of the heterogeneous compounds (AN and AN). The real compound constituents also seem to be rated as more similar than the novel compound constituents. Order of presentation has no effect on rated similarity. To see if there was any significant difference between the homogeneous compounds, t-tests were conducted, both by subjects and by materials. These proved nonsignificant in both cases: by subjects, (paired samples, 2-tailed), t = -0.73, df = 7; by materials (independent samples, 2-tailed), t = -0.10, df = 34.

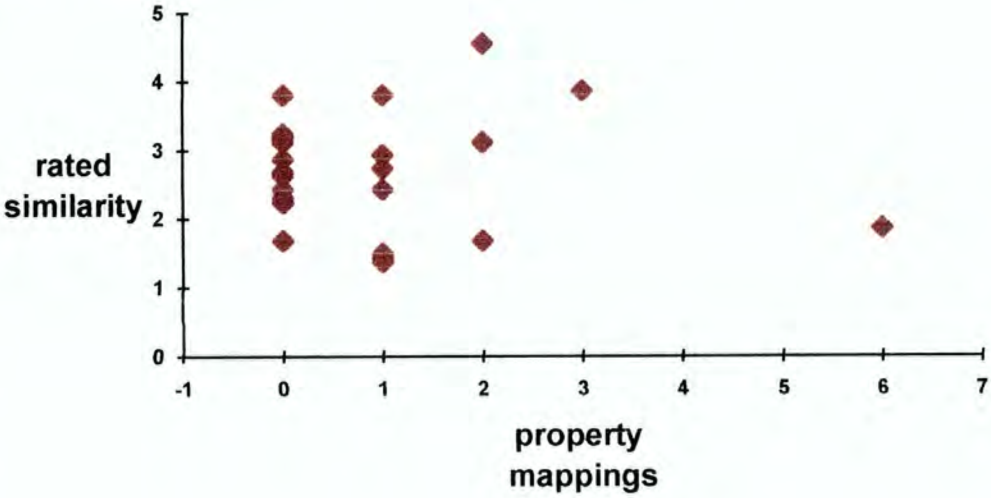
In order to see if there was any correlation between rated similarity and likelihood of property mapping interpretations, scatter plots of mean similarity rating by number of property mapping interpretations were produced. These are shown in fig 4.5. Fig 4.5(i) displays data for both real and novel order1 compounds, novel order2 compounds are excluded because the similarity ratings for these are the same as for novel order1. Fig 4.5(ii) displays only novel order1 data, as this provides the appropriate comparison with Wisniewski's (1996) Experiment 2, which did not consider real compounds.

Figure 4.5 Scatter Plots of Mean Rated Similarity with Number of Property Mapping interpretations

(i) Real and Novel1 data



(ii) Novel1 data



Inspection of the scatter plots in fig 4.5 suggests that there is no correlation, and this is confirmed by the Spearman correlation coefficients (Spearman's rho): real and novel order1, $\tau = 0.042$, $N = 48$, n.s.; novel order1, $\tau = 0.021$, $N= 24$, n.s.

Discussion

Slot Filling and Property Mapping

As in the study by Wisniewski and Gentner (1992), it was found that a slot filling mechanism did not account for all of the noun noun compound paraphrases. However, this study produced a rather greater proportion of slot filling responses (67%) than did Wisniewski and Gentner (40%). The clear implication is that models which rely upon a slot filling mechanism, such as that proposed by Smith et al will not be able to account for all compounds. The present study also found that of the two proposed alternative strategies, property mapping (26%) was much preferred to structure mapping, which was recorded in only one case (<1%).

Furthermore, the relative frequencies of slot filling and property mapping were subject to change with respect to Compound Composition. For the AA, AN and NA compounds slot filling was most frequent, but for the NN compounds property mapping was preferred. This is in partial support of the prediction that homogeneous compounds would favour property mapping because of the greater transferability of features within superordinate categories than between them. Clearly however, the fact that AA compounds do not favour property mapping requires some qualification of this hypothesis.

A possible explanation would seem to lie in the different way that information in artefacts and natural kinds is represented. It can reasonably be expected that if a modifier is to map one of its properties to a head noun, then it will be a property which is particularly salient to that modifier - in other words, some property for which the modifier noun itself can act as a label (e.g. the red chest of a robin in **robin snake**, or the elongated curving shape of a snake in **snake vase**). In the case of most artefacts, the only such property will be its function (e.g. a train is used for transporting things along rails, and a telephone communicates sounds over long distances). However, if this function is now mapped onto another artefact, then a conflict arises because the new function is competing with the head's original function (i.e. the head's most characteristic feature) and the compound cannot tolerate two conflicting values of the same attribute. In the case of salient features being mapped from one natural kind to another, such a conflict is much less likely to arise since the most salient features of any two natural kinds are much less likely to refer to the same attribute. The red chest of a robin can be mapped to a snake quite easily to make **robin snake** a property mapping compound, whereas the transportation function of a train would conflict with the communicative function of a telephone for the compound **train telephone**, making a property mapping relation unlikely.

An alternative explanation can be devised in that there is some reason to suppose that natural kinds and artefacts might differ both in how well they act as slot fillers and offer slots to be filled, and in how well they offer and accept properties. Artefacts have by their very nature got slots ready to fill, in that they have a function, and therefore may have a predisposition to fill slots such as `used for' (e.g. **telephone exchange, animal house**). They also have many slots such as `made of', `part of' and `made by' which could be filled. Fewer available slots are obvious for natural kinds, 'preys on' being perhaps the most likely. There are also differences in the type of slots artefacts and natural kinds might fill. An artefact for example is likely to be a good filler of a `part of' slot, but much less good at filling a `preys on' slot.

Differences between artefacts and natural kinds of a similar type would seem to exist for property mapping. As has been stated above, natural kinds tend to have several salient features compared to the artefacts' one. In general it might therefore be supposed that natural kinds are better at engaging in property mapping relationships from the point of view of both modifier and head noun - they seem like good property donors and acceptors. This difference alone could account for the effect of compound composition in this experiment. It is possible that property mapping will occur if and only if the modifier is a good property donor and the head noun is a good property acceptor, otherwise slot filling occurs.

Two different explanations have therefore been proposed to account for the finding that NN compounds tend to be property mapping, and other compound compositions tend to be slot filling. Both explanations are founded on the assumption that artefacts tend to be more or less defined by their functions, and that natural kinds are represented more by several salient features.

These analyses of the situation seems intuitively appropriate, but do pose a challenge to an assumption made by Wisniewski and Gentner, that is that slot filling is the first choice solution to the interpretation of compounds. They write,

> "....in terms of computation, slot filling may be an easier strategy than others. In general, one only has to check the meaning of the predicate noun rather than alter its structure. .....In contrast, we will suggest that other strategies [such as property mapping] require one to dismantle and significantly alter the meaning of one or both nouns." (Wisniewski and Gentner, p267, 1992).

The present explanations of the AA - NN difference both entail that property mapping is a relatively simple process for some noun types, and that AA compounds tend be interpreted as slot filling because property mapping is difficult. The high proportion of property mapping paraphrases in the NN condition (68%) implies that for these compounds, either slot filling is very troublesome, or a property mapping solution is found prior to slot filling taking place. Support for the idea that property mapping is a primary interpretation mechanism can be found in theories relating to the interpretation of `nonliteral' language, particularly metaphor. The notion of one thing `standing for' another (as in the **robin snake** example) is very reminiscent of metaphorical meaning, and property mapping relations may be considered metaphorical. Also, the idea of a concept lending its name to a metaphorical meaning relating to its most salient feature is in keeping with some current theories of metaphor (e.g. Glucksberg and Keysar, 1990), as is the notion of this kind of meaning being simultaneous to, or even pre-literal (e.g. Gibbs, 1984). Taking this argument a step further, considering property mapping and slot filling as competing or alternative mechanisms may itself be unnecessary. This is because property mapping in a sense implies slot filling. If, in the process of combining two concepts, a subject identifies a property to be mapped from one concept to the other, he/she must then apply this mapping. This application of the property to the head noun is really the same process as slot filling. In the **robin snake** example, once the feature "has a red chest" is identified as the appropriate modification implied by **robin,** all that remains to be done is to fill the slot "chest colour" of the concept **snake**. By this view then, the difference between slot filling and property mapping compounds is not that different mechanisms are employed, but that a different aspect of noun1 acts as the modifier, either its whole or a part of it.

The findings of Wisniewski (experiment 2, 1996) had suggested that property mapping is more likely when constituent similarity is increased. As a post-test of the materials in this experiment, similarity ratings were obtained for the pairs of concepts used as the constituents of each compound. It was found that the constituents of the heterogeneous compounds (AN and NA) were less similar to each other than were the homogenous (AA and NN) constituents, but that there was no difference between the homogeneous constituents. This shows that the difference in proportions of property mapping and slot filling obtained for NN compounds in this experiment is not due to differences in constituent similarity. A test of correlation (Spearman's rho) also found that there was no relationship between rated constituent similarity and the number of property mapping interpretations generated for the compounds used in this experiment. These results do not readily agree with those of Wisniewski.

Since the experimental procedures in this experiment and in Wisniewski's are really quite similar, the reason for the different results probably lies in the materials used. Wisniewski was deliberately manipulating similarity as an experimental variable, and as such compounds in his 'similar' condition are composed of constituents which are very similar indeed, e.g. **coat shirt, tie scarf, organ piano, saxophone trumpet**. This very high level of similarity (not present in any of the materials in the present experiment) seems to have facilitated property mapping. The results of this experiment, and those of Wisniewski can therefore be seen as compatible. Wisniewski has demonstrated that a very high level of similarity of constituents does have an effect on the relative frequency of property mapping and slot filling interpretations. The present experiment has demonstrated that Compound Composition also has an effect at lower levels of constituent similarity.

Table 4.6 Percentages of property mapping interpretations for some compound compositions (adapted from Wisniewski, 1996, experiment 1)

Compound Composition	Percentage Property Mapping
artefact-artefact	38
artefact-animal	45
animal-artefact	41
animal-animal	74

Further evidence in support of this finding is present in Wisniewski's experiment 1. Wisniewski's experiment 1 did not manipulate constituent similarity, but did manipulate Compound Composition in a way similar to the present experiment. Among other Compound Compositions Wisniewski included four which are very similar to the four used in this study, artefact - artefact, animal - artefact, artefact - animal and animal - animal (the only difference being that Wisniewski used `animal' instead of natural kind). He found that his animal - animal compounds seemed to show more property mapping than the other compound compositions, as shown in Table 4.6, which is a very similar result to that for natural kind - natural kind compounds in this experiment.

Real / Novel Effect

The difference in slot filling and property mapping for the four compound compositions seems to interact with whether the compound is real or novel. Differences are in the same direction, but are more marked for the real compounds than for novel ones. There are two possible explanations for this effect. First, such a situation would arise if novel compound interpretation is carried out by analogy to real compounds. This supposes that when confronted with a novel compound to interpret, we attempt to find a real compound which has similar constituents, and whose interpretation we are confident about. The novel compound will then be interpreted in an analogous way. Ryder (1994) has proposed such a theory, with the analogy occurring at any of several levels (e.g. same constituent, similar constituent meanings etc.). Shoben (1993), showed that we can be primed towards a certain type of slot filling relation by presenting subjects with novel compounds in a biased context, and then obtaining paraphrases of similar novel compounds. Such interpretation by analogy would mean that any bias occurring in the population of real compounds would result in a similar but less marked bias in novel compound interpretations, as is found in this data. The problem with such a theory is that it only explains how a bias might arise in novel compounds if such a bias is already present in real compounds. How this original bias might come about cannot be accounted for.

The alternative is to suppose that the biases are present in the population of novel compounds, due perhaps to some of the factors discussed above, such as natural kinds having many salient properties. Most interpretations of novel compounds will be in accordance with these biases, but some will not. Those interpretations that run counter to the trend are likely to have a low communicative value, in that they are less predictable. A compound is only useful if the utterer and interpreter are very likely to agree on its meaning. Such agreement is much more likely if compounds are predictable, i.e. if they agree with the common trend. Therefore compounds with predicable meanings are useful, and stand a good chance of being adopted as a part of the language (i.e. becoming `real'), compounds which are less predictable are less likely to be accepted into the language. Biases evident in the population of novel compounds are thus likely to be more extreme in real compounds.

Both these theories would make similar predictions. Since Wisniewski found that novel compounds composed of highly similar constituents tend to have property mapping interpretations, we can predict a similar, though stronger bias would be found in real compounds with highly similar constituents. (Unfortunately, the practical difficulties of finding enough real compounds with highly similar constituents might render such an experiment impossible.) The same would go for any other bias detected for novel compounds, and vice versa for a bias detected in real compounds.

Referential Status of noun2

The finding that noun2 is the referent of the large majority of compounds will surprise no-one. However, the finding that real AN compounds are much less likely than other compounds to have a noun2 referent is more interesting. Inspection of the data shows that four of the compounds (**toilet duck, clay pigeon, computer mouse** and **mug tree**) are responsible for all of the responses which indicate that noun1, or neither noun1 nor noun2 is the referent. In these cases the noun2 appears to contribute a metaphorical meaning to the compound. At the same time, there is some indication from the data that novel AN compounds are even more likely than other novel compound compositions to have a noun2 referent, and less likely to have a noun1 referent. This effect is therefore in the opposite direction to that of real compounds.

The finding that novel AN compounds often don't have a noun2 head implies that subjects found difficulty in modifying the natural kind head. This is discussed at some length in chapters 2 and 3. The difficulty seems to result from a lack of confidence about how a natural kind concept can be modified by an artefact without violating its essential characteristics. This leads to the possibility that the apparently metaphorical interpretation of noun2 in the real AN examples is due to just such a modification. that is, that the real AN compounds which lack a noun2 reference have had their interpretation fixed by convention such that the defining characteristics of the head noun have been changed. Quite why this radical change to the head noun has been allowed in the real compounds but not the novel is open to question. It may just be due to chance selection of materials for this study. However it may be a function of the great difference in context provided for the interpretation of these compounds. When confronted with the novel compounds in this task, subjects had very little or no context influencing their interpretation. But for the real compounds the reverse is the case, because their interpretation of these compounds didn't really occur during the course of this study, all they did during this study was to recall an earlier interpretation, which was probably made in a very rich context. This rich context would have given them the support necessary to settle upon an interpretation which did not maintain some of the essential features of the head noun. This account of the difference between interpreting real and novel compounds therefore implies that the same processes occur in each case, but that without the support of context interpretations which violate the essence of the head noun will be very rare. With contextual support such interpretations can and do occur.

Differences from Wisniewski and Gentner

The overall difference in the amount of slot filling relations recorded in this study, compared to that of Wisniewski and Gentner might be a cause for concern. Some of this difference can be accounted for by the fact that the present study included real compounds, which Wisniewski and Gentner did not. These compounds did show an increased likelihood to have slot filling relations, 78% of real compound paraphrases were categorised as slot filling compounds, compared to 62% of novel compounds. However, this still leaves a substantial discrepancy.

It could be argued that this difference is due to the fact that not all of the paraphrases in this study were actually categorised for modification relationship. Those paraphrases which lacked a noun2 as head, or were judged not to conform to the 'noun1 modifies noun2' pattern were excluded from the final stage. However, since the discarded paraphrases were predominantly classified as 'noun1 does not modify noun2' (61%) or uninterpretable data (a further 6%) it seems very unlikely that they could contain a high proportion of slot filling relations. An alternative explanation for the discrepancy might be sought in the fact that this study failed to include mass nouns. However, mass nouns are, as Wisniewski and Gentner point out, likely to specify what an artefact is made of (Wisniewski and Gentner report this to occur in 59% of all mass-count compounds). This is clearly a slot filling relation, so at least in this type of compound using a mass noun it would be expected that slot filling might increase. Again, it is unfortunate that Wisniewski and Gentner do not report on the extent of slot filling in the other compounds containing mass terms (count-mass), and conjecture about it would only be confusing. Although it is not likely to be useful to speculate on this matter, it would seem surprising if the remaining mass term compounds were so heavily biased against slot filling as to account for the difference in slot filling reported in these two studies. The most likely explanation for the difference would seem to be simply that the criteria for what is counted as a slot filling relation were not precisely the same in both studies. In fact, the judges in the Wisniewski and Gentner study were not explicitly told to look for 'slot filling' relations, rather they were instructed as to what to include in each category largely by way of several examples. Judges in the present study were explicitly asked to look for 'slot filling' relations (since they were cognitive psychology postgraduates this kind of theory of conceptual representation was not new to them). It would seem quite possible that picking out slot filling relations by similarity to a few examples would give a more conservative measure than doing the same task when using more theoretically based criteria.

Summary

In summary then, this experiment has lead to several conclusions. Firstly, it replicated the result of Wisniewski and Gentner (1991) that slot filling is not the only method used to interpret noun noun compounds, but it is the single most common. The other important mechanism is property mapping. For AA, AN and NA compounds slot filling was far more common than property mapping, but for NN compounds this was reversed. This effect is attributed to differences between the representations of natural kind concepts, which tend to have several salient features, and artefact concepts which are defined principally by their functions. This explanation entails that property mapping is not a secondary mechanism called upon only when slot filling fails, and has much in common with current theories of metaphor interpretation (e.g. Glucksberg and Keysar, 1990).

It was also found that the difference in proportions of slot filling and property mapping interpretations between compound compositions was more extreme for real compounds than it was for novel compounds. This is explained in terms of novel compound interpretations being more successful, and so more likely to be adopted into the language, if they are in accordance with a general trend.

Counter to the findings of Wisniewski (1996), constituent similarity had no effect upon the proportion of property mapping interpretations given to a compound. It is suggested that this is because the effect reported by Wisniewski only occurs when constituents are extremely similar. In compounds composed of less similar constituents discrepancies of constituent similarity between compounds have little or no effect.

It was also found that real AN compounds had a surprisingly high proportion of interpretations which did not have noun2 as the reference. The noun2 of several of these compounds was given a metaphorical interpretation in many cases. This is attributed to the difficulty of modifying a natural kind by an artefact. This difficulty is overcome for real compounds by giving the noun2 a metaphorical reading, which is possible in the supportive context assumed to be present for real compounds. Since no such contextual support was available for the novel compounds such metaphorical interpretations were less likely.

Experiment 6 : Possible Causes of Property Mapping

Introduction

Experiment 5 examined the influence of several factors on the proportion of slot filling and property mapping interpretations given to noun noun compounds. It was found that compounds composed of two natural kind concepts were more likely to have property mapping interpretations than were other compound compositions.

Wisniewski (1996) examined the effect of similarity between the two constituent concepts, and concluded that compounds of similar concepts were more likely to have property mapping interpretations than were compounds of dissimilar concepts. He accounts for this in terms of **structural alignment**. Wisniewski suggests that when two constituents are similar, people mentally align their representations with each other, and so compare them for points of contrast. When a salient contrast is found, we then resolve it by mapping across a feature from the modifier to the head, so the compound inherits the value of the contrasting attribute from the modifier.

However, in apparent contradiction of Wisniewski's finding a post-test of the materials used in experiment 5 found that rated similarity did not have any such effect in those data. The explanation proposed for this discrepancy relies on differences in the procedures used for the selection of materials for the two experiments, and on the identification of another possible influence on the type of interpretation given to noun noun compounds - the number of highly salient features available. This factor, the number of salient features of the constituent concepts, along with the factor of constituent similarity, are the subject of this investigation. The hypothesis is that the number of salient features available and constituent similarity should both have an influence on the type of interpretation given to a compound.

Method

This study investigates the influence of two factors in particular on the interpretation of novel compounds - feature clashing and constituent similarity. In order to select appropriate materials for the experiment, two preparatory stages were used to gather a suitable range across each of these two factors.

1 - Obtaining defining features

Subjects

Thirty six unpaid volunteers participated in this experiment. They were all undergraduates studying psychology at the University of Durham. There was an approximately equal number of males and females, aged 18 - 30 years.

Materials

Forty eight natural kind nouns and forty eight artefact nouns were selected. These were the same nouns as those used in the compounds of experiment 5 (see appendix, table A5.1). These were used to compile two similar questionnaires (twenty four artefacts and twenty four natural kinds in each) in which each item was enclosed within an outline border along with a single dotted line. There were twelve such presentations per page of a five page booklet, the first page being subjects' instructions, as follows:

"This booklet contains this instruction page followed by four test pages.Each of the test pages contains twelve boxes similar to this :

{13N4} **snake**

....

I would like you to read the underlined word, and decide what you think is the **most defining characteristic** of the concept it names. Please write this characteristic on the dotted line below the word.

You shouldn't need to spend long on each item, but please do try to write the answer which you feel is most appropriate."

The test items were sorted into two different random orders.

Design

This part of the experiment had a single two level within subjects factor, Concept Type (artefact or natural kind).

Procedure

Subjects were asked to read each item and write on the dotted line what they judged to be the *most defining feature* of that concept. They were given as long as they need to complete this task, which was typically 12 - 15 minutes.

Treatment of the data

The defining features listed for each item were collated, and the data were used to select concept pairs in two different conditions, those likely to exhibit a clash

(Clash) of defining features, and those unlikely to exhibit such a clash (Nonclash).

To produce the pairs of the **clash** condition, concepts which had one or two defining properties listed very frequently were chosen. These were paired with one another in the cases in which the values of frequent properties were not the same. For example, the slot '**function**' was frequent for both **glove** (8 times) and **door** (10 times), and had a different value in each case ('for protecting a hand', and 'for allowing entry/exit'), so these were paired in the clash condition. For the **nonclash** condition items were paired which did not have the same highly salient slots. For example, **glove** was paired with **street** for which function was only mentioned once. Whereas the clash condition only contained concepts which were characterised by one or two highly salient features, the nonclash condition also included concepts which were defined by several moderately frequent features as well. Combinations of artefacts and natural kinds were made in all possible permutations, giving all four possible compounds types. Details of materials are given in the appendix, table A6.1.

2 - Obtaining similarity judgements.

The concept pairs in both the clash and nonclash conditions were then submitted to subjects to obtain ratings of the judged similarity between the concepts within each pair.

Subjects

Four unpaid volunteers participated in this experiment. They were all postgraduates studying psychology at the University of Durham.

Materials

Four questionnaires were produced, each containing the full set of paired concepts. Concept pairs were sorted into a random order on which the four questionnaires were based. In order to eliminate any bias from presentation the four questionnaires differed according to two binary criteria, pair order and presentation order. Pair order simply refers to which way round the similarity judgement was to be made (i.e. "how similar is A to B?" versus "how similar is B to A?"). Presentation order is the sequence in which the questions were presented, either forwards (question 1 first) or reversed (question 1 last).

Design

This part of the experiment was of a 4 (Compound Composition) x 2 (Clash v Nonclash) design, with repeated measures on both factors.

Procedure

Subjects were asked to work their way through the questionnaire, and answer each question by circling the appropriate number on a scale of one (extremely dissimilar) to 7 (extremely similar). Subjects were able to take the questionnaire away and complete it at a convenient time.

Treatment of the data

The data were used to assess the similarity of each concept pair, regardless of pair order, so for each pair of concepts there were four ratings. Concept pairs were sorted into two conditions, high similarity and low similarity. In the first instance, pairs with three ratings of 5 (similar) or higher were put into the high similarity condition, and those with 3 ratings of 2 (very dissimilar) or below were put into the low similarity condition. However, these criteria left some conditions very low or empty, so they were relaxed until there were several concept pairs in each condition. Some conditions were much less likely to have items for certain similarity ranges, so the numbers in each condition varied.

3 - Experimental questionnaires

Subjects

Eighteen unpaid volunteers participated in this experiment. They were all undergraduates studying psychology at the University of Durham. There was an approximately equal number of males and females, aged 18 - 30 years.

Materials

The noun pairs selected by the two preparatory stages of the study were made into compounds, the constituents being combined in each order. Table 4.6 shows the number of noun pairs, and the mean similarity rating in each condition. The compounds were then sorted into two random orders, and each presented above a dotted line, 10 such presentations per page. Nine questionnaires were of one random order, nine of the other. Of each set of nine, five questionnaires had the original page order, and the remaining four had the pages in reversed order, to counter any order effect. Materials and their similarity ratings are given in the appendix, table A6.1.

Design

A 4 (Compound Composition) x 2 (Clash / Nonclash) x 2 (Similarity) design was used, with repeated measures on all factors.

Procedure

Subjects were asked to write what they understood by each compound on the dotted lines. They were able to take the questionnaires away and complete them at their leisure.

Table 4.6 : The number of noun pairs and the mean similarity rating in each condition[9]

	High Similarity				Low Similarity				Mean	
	Clash		Non-clash		Clash		Non-clash			
	sim	n	sim	n	sim	n	sim	n	sim	n
AA	3.85	(10)	3.65	(10)	1.89	(14)	1.69	(26)	2.77	15
AN	2.75	(3)	2.75	(5)	1.75	(7)	1.50	(6)	2.19	5
NA	2.75	(3)	2.75	(5)	1.75	(7)	1.50	(6)	2.19	5
NN	4.89	(32)	4.58	(16)	3.25	(18)	3.5	(8)	4.06	19
Mean	3.56	(12)	3.43	(9)	2.16	(12)	2.05	(12)	2.80	(11)
Mean		3.50	(11)			2.11	(12)			

Results

Although not directly the subject of this investigation, some statistics associated with the data generated in the preparatory stages of this study are given here.

1 - Defining Features

Table 4.7 shows the mean number of different defining features for each concept.

Table 4.7 Mean number of different defining features per item

artefacts	natural kinds
5.9	8.0

---

[9] AN and NA conditions have the same similarity scores and numbers because the compounds in these conditions are simply reversals of each other, i.e. "AB" in the AN condition becomes "BA" in the NA condition.

2 - Experimental results

The definitions given by subjects for each of the compounds in the four conditions were categorised according to the type of relationship they implied between the two constituent nouns. The categories used were the same as those used in experiment 5. A 2 (Clash / Nonclash) x 2 (Similarity) x 4 (Compound Composition) analysis of variance with repeated measures on all factors by subjects, and on none by materials was conducted for the mean number of property mapping interpretations listed per item.

The main effects of Similarity (Min F' = 7.13, df = 1, 8, $p < 0.05$) and Compound Composition (Min F' = 8.744, df = 3, 14, $p < 0.01$) were both significant by Min F'.

The interactions Similarity x Compound Composition (F1 = 5.88, df = 3, 45, $p < 0.002$; F2 = 1.49, df = 3, 7, n.s.), Clash x Compound Composition (F1 = 6.71, df = 3, 45, $p < 0.001$; F2 = 1.81, df = 3, 7, n.s.) and Clash x Similarity (F1 = 5.73, df = 1, 15, $p < 0.03$; F2 = 2.97, df = 1, 7, n.s.) were significant on the F1 analysis only. No other effects were significant.

As can be seen from the data in Table 4.8 below, the main effect of Compound Composition is principally due to a higher proportion of property mapping interpretations in the NN compounds, although the effect is somewhat graded. The same table also shows that the Similarity effect is due to property mapping interpretations being more prevalent amongst high similarity compounds than low similarity compounds.

The Clash / Nonclash x Similarity interaction which approaches significance in the materials (F2) analysis is illustrated in Fig 4.6. The main effect of Constituent similarity is more marked for compounds in the nonclash condition than it is for compounds in the clash condition.

Table 4.8 : Mean Proportion of Property Mapping Interpretations by Compound Composition and Similarity and Clash/Nonclash

	Clash		NonClash		Mean
	High	Low	High	Low	
AA	.30	.20	.35	.16	.25
AN	.44	.35	.40	.10	.32
NA	.19	.29	.25	.14	.22
NN	.55	.36	.67	.45	.51
Mean	.37	.30	.42	.21	
Mean	.34		.32		

Fig 4.6 : Proportion of Property Mapping Interpretations for Clash/Nonclash x Similarity (sig. by F1 only)

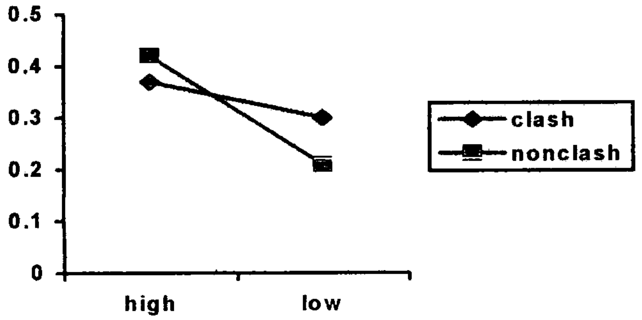
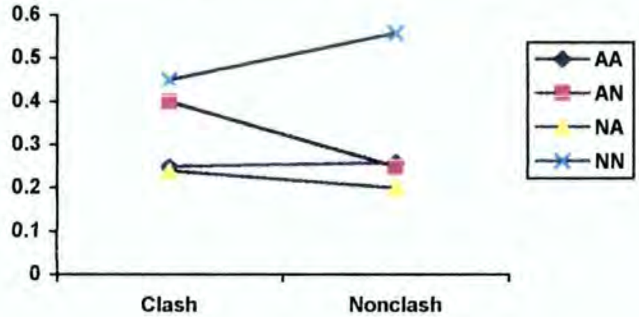


Fig 4.7 shows the Clash/Nonclash x Compound Composition interaction, which was also nonsignificant by F2. For NN compounds there are more property mapping interpretations in the nonclash condition. This difference is reversed for

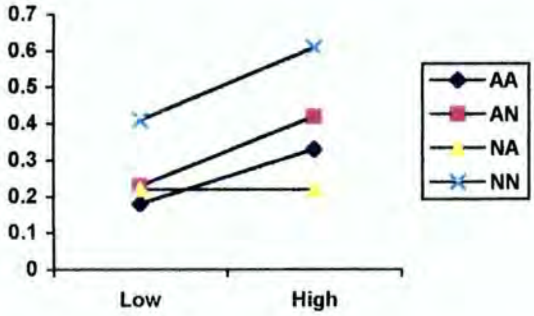
AN and NA compounds, and the clash and nonclash conditions are almost equal for AA compounds.

Fig 4.7 : Proportion of Property Mapping Interpretations for Clash/Nonclash x Compound Composition (sig. by F1 only)



The Similarity x Compound Composition interaction (nonsig. by F2) is illustrated in fig 4.8. The NA Compound Composition does not follow the main effect of Constituent Similarity, since property mapping is equally frequent in both high and low similarity NA compounds.

Fig 4.8 : Proportion of Property Mapping Interpretations for Constituent Similarity x Compound Composition (sig. by F1 only)



Discussion

The main effect of similarity is significant, property mapping being more frequent in high similarity than low similarity compounds. This result would seem to offer support to the suggestion that property mapping is facilitated by structural alignment. However, a cognitive mechanism involving structural alignment is not the simplest account which might explain such a result. It could simply be that if a property is "offered" by the modifier, then that property is most likely to be acceptable if the head concept is similar to the modifier. This too would predict that property mapping would be most common among compounds composed of similar constituent concepts. Although these two accounts are rather alike, it's important to distinguish between them. Structural alignment is postulated as a cognitive process. The alternative account suggests that the increase in property mapping is simply the inevitable consequence of an increase in constituent similarity. It is based on the increased probability of features being acceptable to both concepts, regardless of the cognitive mechanisms underlying property mapping.

Since the two explanations do have much in common, determining which is more appropriate is not easy. Consideration of the behaviour of the compounds with respect to compound composition in this study could help in this respect. The main effect of compound composition is significant, NN compounds exhibiting the most property mapping. This result repeats a finding of experiment 5. However, in the present study the result is somewhat confounded with similarity, since the rated similarity is also highest for NN compounds. But the effect of Compound Composition here cannot be written off on the basis that it is simply due to similarity differences between compound compositions, because the order of the effect does not match that of similarity differences, as shown in Table 4.9 below.

Table 4.9 : Order of Effects of Rated Similarity and Property Mapping

similarity:	NN	>	AA	>	AN	=	NA
property mapping:	NN	>	AN	>	AA	>	NA

Importantly, property mapping is more common in AN compounds than AA or NA compounds. This is relevant firstly in that it is counter to the ordering of rated similarity of constituents. Apart from this however, it is also **not** the case that property mapping is more common in homogeneous compounds than in heterogeneous compounds. Such evidence counts against the hypothesis that the important influence is whether or not constituents are of a similar nature. Clearly the picture is more complicated than that. In this respect then, the main effects of Similarity and Compound Composition can be taken as supporting the hypothesis that property mapping is facilitated by structural alignment.

But this seems a far from complete explanation of the data. The graded effect of Compound Composition suggests that other influences are present. Structural alignment (via similarity) accounts for why NN compounds exhibit more property mapping than AN compounds, and why AA compounds have more property mapping than NA compounds. The constituents of NN and AA compounds are more similar than AN and NA compounds. But it also seems that compounds with a natural kind head noun are more likely to be interpreted by property mapping than those with an artefact head. It was a finding similar to this in experiment 5 that lead to the hypothesis that the likelihood of feature clashes might be responsible for the difference. But the main effect of Feature Clash in the present study proved to be nonsignificant, and so that hypothesis would seem to be refuted. Having replicated the original result, however, some explanation of the effect is needed. The results for the Similarity x Clash/Nonclash interaction suggest that the presence of feature clashes does have some influence. This interaction was highly significant by F1, and the nonsignificant F2 result is quite possibly due to the low numbers of materials in some conditions (see Table 4.6). The data for this interaction (Table 4.8) suggest that the main effect of Constituent Similarity is largely confined to the nonclash condition. It therefore would seem reasonable to interpret these data as indicating that feature clashes do have an influence on property mapping. It seems that the presence of feature clashes does interfere with the effect of Constituent Similarity. Ideally this interpretation ought to be tested with a larger pool of materials. The other 2 way interactions are also significant by Fl but not F2. the Clash/Nonclash x Compound Composition interaction (fig 4.7) also suggests that feature clashes do have an important influence over compound interpretation. Property mapping is equal in both clash and nonclash conditions for AA compounds, but more common in the clash condition for AN and NA compounds, and more common in the nonclash condition for NN compounds. Perhaps for diverse constituents (AN and NA compounds) a property clash actually acts as a useful point of comparison, and so helps structural alignment. For more closely related constituents, it has little overall effect (AA) or makes structural alignment more difficult (NN compounds).

The Constituent Similarity x Compound Composition interaction (sig by Fl only) also suggests that Constituent Similarity is not the only important factor. NA compounds do not show the main effect of Constituent Similarity, property mapping being equally frequent in both high and low similarity conditions. However, with there being only a few NA compounds tested in each condition, and F2 proving nonsignificant, it would be inappropriate to read too much into this result.

In summary, the main effect of Constituent Similarity replicates the finding of Wisniewski (1996). Given the further result that property mapping is not more frequent in each of the homogeneous compounds (AA and NN) than in the heterogeneous ones (AN and NA), the data supports a structural alignment theory rather than one of simple feature transferability.

The low F2 values for many effects are attributed to low numbers of materials in some conditions. The data do however suggest that feature clashes are influential in compound interpretation, since there appears to be some interaction with the main effect of Constituent Similarity, and with Compound Composition. A further study with balanced materials would be needed to examine these effects more properly.

Chapter five: The Significance of Each Constituent

Experiment 7 : Concept Identification

Introduction

In the 1970's, largely due to the work of Rosch and co-workers (e.g. Rosch 1977), the classical view of concept representation was seen to be inadequate in its account of how we mentally represent concepts. Thus the notion of concepts being defined by sets of `necessary and sufficient' features was replaced by the idea that concepts were somehow represented by `prototypes', and that category membership was related to the degree of similarity to that prototype.

This theory reflected the wealth of psychological data which demonstrated the prototypical behaviour of human concepts. But there are problems associated with such a view. Armstrong, Gleitman and Gleitman (1983) showed that some concepts which were clearly and explicitly `classical' in their definition, such as `triangle' and `odd number', also exhibited `prototypicality effects'. From this they argued that apparently prototypical behaviour should not be taken as indicative of an underlying prototypical structure. Smith and Medin (1981) have shown that standard probabilistic models make erroneous predictions about the levels of typicality of category members to a combined concept. This line of reasoning has been extended by Murphy (1990), who claims that much of the behaviour of concepts is too complicated to be modelled in such a restrictive fashion. This has lead to the idea that concepts may be represented in a complex structure. Murphy subscribes to the view that theoretical knowledge of the world at large is integral to concept representation. Less extreme than this notion is the idea that a concept has some central core combined with features which are typical of instances of that concept, and which can be used for identification / classification purposes. This idea was explored by Miller and Johnson-Laird

(1976), a more recent formulation exists in Keil (1987). The idea of `identification procedures' has also been suggested from a philosophical tradition, Putnam (1975).

The `core' in such a theory could be various things. It has been suggested that such a core is valid only psychologically, in that it corresponds to nothing in the real world - this is known as psychological essentialism (Medin and Ortony, 1989). Putnam also claimed that the core was in some sense illusory, in that people defer knowledge of such things to experts - who may or may not in fact know of essential defining characteristics. It has also been widely suggested that the nature of a concept's `core' depends upon the super-ordinate semantic category of the concept. Artefacts seem to be defined by their function (or something close to it), whereas living things are essentially defined by their genetic code and atoms by their atomic number (Keil, 1989).

The idea that artefacts and natural kinds could differ in their core content has not been widely empirically investigated. One recent paper which did examine this theory is that of Malt and Johnson, 1992. They attempted to test the claims that the function (or in fact the `intended function', i.e. what something is `manufactured and sold to do') of an artefact is both *necessary* and *sufficient*. To test the assumption that intended function is *sufficient* for categorising artefacts they presented subjects with passages, each of which contained two types of information, relating to the *features* and *intended function* of a particular object. Each passage was followed by a question such as "Is this thing a sweater?". The two conditions of the experiment were *usual* and *unusual*. In the usual condition the features were typical of the concept referred to in the subsequent question (in this example, made of wool, has buttons on the front and has open sleeves). In the unusual condition the features were highly atypical (made of rubber, buckles at the back, and closed sleeves). In both cases the second half of the passage described the usual intended function of the target object (provides warmth for the upper body, and worn over a shirt). Subjects were found to respond that the passage **did** describe the target object far more frequently in the *usual* condition than in the *unusual* condition. In fact, the unusual condition resulted in less than half of the responses being positive. Thus Malt and Johnson claimed that artefacts can be excluded from categories in spite of possessing the associated function, so it is not a sufficient criterion for category membership.

They used a similar paradigm in their second experiment, in which they varied the intended function between *normal*, *related*, *bizarre* and *denial*, and maintained the features as constantly typical of the target. This was designed to test the hypothesis that having the appropriate intended function is *necessary* for categorisation as a certain artefact. Findings for this study showed that while the normal function led to the greatest likelihood of categorisation as the target, all the other conditions still led to many subjects affirming that the object described was an example of the target concept. Malt and Johnson therefore conclude that intended function provides neither necessary nor sufficient information for the categorisation of artefacts, and as such is unlikely to be a concept core in the commonly understood sense.

There are however problems with this strong interpretation of the results. Results of two other studies (by Keil, 1989, and Rips, 1989) which have included similar sorts of tasks to this have found that subjects' judgements **were** determined by the intended functions of artefacts. Malt and Johnson suggest that due to problems of the design of Keil's study, and problems of interpretation of Rips' study, these results are actually compatible with theirs. However, there are a number of issues which might be made about Malt and Johnson's experiments, that question the interpretation of the results obtained. The experimental design seems to bias towards their findings. Importantly, in their test of *sufficiency* of function, Malt and Johnson always presented subjects with passages which contained feature information at the start of the passage, and then function information at the end. Since it is this kind of feature information which has been claimed to be utilised in concept identification, it is perhaps not surprising that subjects' responses are influenced by it. It is quite possible that subjects identify the concept from this information, and the following information about function is then simply accommodated within this concept.

Further, the experiment then asks the subject to make a categorisation decision about the object, but in doing so prompts the subject with the name of the target concept. Asking "Is this thing a sweater?" introduces a confounding influence. If the subject thought that it was a sweater, but had some doubt, this question might well support such doubts. If the subject thought that there was only a slight possibility that it was a sweater, the question might suggest that it really was one. Perhaps better would be to simply ask "What is this thing?", or "Choose the appropriate name for this from the following...". Another problem associated with how the question is asked is the possibility of over interpreting the question. Subjects might interpret it as "Is this a sweater by all criteria?" rather than "Can this be called a sweater?". Finally, for most subjects participating in a study like this, it would surely seem odd to consistently ignore the majority of the information given for each task, and to give only the simplest response in each case. Subjects might well expect that such behaviour would be regarded as either obtuse or deliberately disruptive.

The materials used in the second experiment are also questionable, particularly the `denial' condition. These intended functions are supposed to explicitly contradict the intended function of the target, but to this reader they plainly do not, see Table 5.1 below.

These various criticisms do not entirely dispel the conclusion that intended function is not absolutely necessary and sufficient for the categorisation of artefacts. The fact remains that, for whatever reason people have been shown to use so called characteristic features for categorisation even when information about function is made explicit.

Table 5.1: Example `denial' functions from experiment 2 (Malt and Johnson, 1992)

BOAT: ...for collecting samples of marine flora and fauna under sterile conditions and is totally mechanised so no people are allowed on board under any circumstances.

COUCH: ...a seating area for participants in mood experiments in psychology labs and is very uncomfortable so it induces anger and frustration.

STOVE: ...to heat foods in a laboratory for various tests, but in so doing releases toxins that make the foods inedible.

Similar finding have been reported by Hampton (1995). Using a variety of concepts, he found that subjects' categorisations were influenced by supposedly characteristic features both when the supposedly defining features for a concept were explicitly present, and when they were explicitly absent. Over a series of four experiments, Hampton altered the experimental conditions slightly in each case (mainly alterations to subjects' instructions and questions), and in so doing increased the relative importance of the `defining' features over the `characteristic' features. But the influence of the characteristic features could not be eliminated. The sensitivity of this effect to experimental detail suggests that the effect itself could be an experimental artefact, and may be practically impossible to eliminate in such studies. As in the above criticism of Malt's experiments, it is unreasonable to expect subjects' categorisations to be insensitive to the peculiar demands of the experimental task.

This sort of interpretation is apt to provoke criticism from those who would see it as counter-empirical. Hampton writes,

> "Philosophers and psychologists may agree that natural kinds such as WATER or LEMON should be defined in terms of some essential and hidden structural property. The subjects in these experiments and in Malt's studies do not apparently share this view. The question then is whether the psychology of concepts should be modelling concepts per se, or people's categorisation behaviour." (Hampton, 1995, p703)

However, it is in the nature of such studies to influence the categorisation behaviour of subjects. The failure to observe subjects treating some features as defining and others as characteristic may be a limitation of this kind of experiment rather than an indication that people do not behave in that way.

Function can still be seen as some kind of a core for artefacts. In the light of the studies by Malt and Johnson and Hampton this core cannot be seen as constituting a rigorously necessary and sufficient condition for category membership. Such judgements are subject to influences of context. However, these results do not show that concepts lack a core of meaning which subjects would tend to rely on in a suitable context. These results can be interpreted as the behaviour of subjects who do have function at the centre of their artefact concepts, but who are also subject to contextual influences which may override the rigorous and logical application of this categorisation scheme. It is far from unheard of that a subject might make a decision more influenced by contextual cues than by logical principles.

In the light of these problems with the Malt and Johnson study, and considering other investigations with apparently contradictory findings, it would seem premature to reject the `core + identification features' type of model, or to deny the role of function / intended function in the core of artefacts. The work so far discussed has investigated the nature of the proposed concept core, and possible differences between semantic categories with respect to this core. Little work has been carried out on the identification procedure, or how this might be subject to differences between semantic types.

This experiment concerns people's identification of concepts. Descriptions of different types of concept are presented to subjects, who are asked to identify the target concept. The descriptions are feature listings in which the features have all been rated as to how descriptive they are of the target. Subjects were also asked to indicate (by ticking them) which of these features they actually used in coming to their decision.

The types of concept vary in several ways. Firstly, concepts are either artefacts or natural kinds. As has been discussed, it has been widely suggested that these two semantic categories may differ structurally, principally with respect to the proposed concept core. There is evidence to show that manipulation of this core information can alter how an object is categorised. This experiment will determine whether one of these types of concept is easier to recognise from a list of features than the other. Malt and Smith (1984) also showed that natural kind concepts tend to have a more limited and defined set of features than do artefacts. It will therefore also be interesting to compare the number of features used by subjects to identify these two types of concept.

As well as these simple concepts, subjects are also asked to identify complex concepts, in the form of noun noun compounds. Both real and novel compounds are used as targets. It is predicted that, it will be much easier to identify real compounds than novel ones, mainly because the former are familiar parts of the language and so less ambiguous. The compounds are composed of the simple artefacts (A) and natural kinds (N) already used in the experiment, so as to form each possible combination (AA, AN, NA, NN). Comparison of these different compounds will also be of interest, since in other experiments in this thesis, they were found to have some different properties.

The present study also includes the factor of feature set (fs1, fs2, fs3), in which features are divided according to which part of the compound (noun1, noun2, or the compound itself) they were originally generated for. Examination of the features used by subjects with respect to this factor should reveal information about the kind of features which are of most use in identifying compounds, and so give an insight into the identification procedure employed.

This experiment should therefore be able to provide a range of information about concept identification, and in doing so enable a more informed judgement on the validity of the core + features theories of concept representation. Data may also relate to possible differences between artefact and natural kind concepts and concept combination.

Method

Subjects

144 unpaid volunteers participated in this experiment. They were all undergraduates studying psychology at the University of Durham. There was an approximately equal number of males and females, aged 18 - 30 years.

Materials

The same experimental items used in experiments 2, 3 and 4 were used here (see appendix, table A2.1). The format and design of the questionnaires used in this experiment is described below. The questionnaire set contained 24 different questionnaires, and six replicas of this set were used in this study, totalling 144 questionnaires.

Questionnaire format

Each questionnaire contained seven different pages, the first being instructions and then six test pages. The order of these test pages was randomly determined. Of the six sets of questionnaires, three were in one random order, and the other three were in the reverse of this order. Each test page of these questionnaires had a blank dotted line at the top for the subject to enter his/her response. This was followed by a list of thirty features, which were grouped into categories, as can be seen in the example test page shown in fig 5.1. At the foot of each page were several dotted lines, on which the subjects could record any further information relating to the derivation of their response.

Fig 5.1 : Example test page from a questionnaire

{1a1} ....

Necessary

☐ heats

Very important

☐ uses gas or electricity
☐ temperature controlled by a thermostat
☐ hot
☐ cooks

Fairly important

☐ racks inside
☐ kitchen preparation
☐ huge cube/box shape
☐ hatch opens at the front
☐ enclosed

Typically true

☐ insulating material

Not usually true

☐ padded and insulated
☐ keeps hand warm

Very unlikely

☐ used to carry hot things safely
☐ soft
☐ smaller than 20cm
☐ shape of human hand
☐ no separate finger holes
☐ five fingers
☐ 5 cylindrical extensions

Impossible

☐ worn on the hands
☐ the two hands are joined by a strip of material
☐ strip of strong heat-resistant cloth
☐ stop hands being burnt
☐ protects hand
☐ material
☐ made of linen
☐ fits on a hand
☐ cloth may be plain or patterned
☐ 2 pockets at each end for hands

....

....

....

....

Questionnaire design

The features used, and the categories to which they were assigned were derived from experiments 2 and 3 in which subjects had been required to produce features and give descriptiveness ratings for them with respect to given concepts. Each set of thirty features relates to a concept triple, comprised of three Compound Parts i.e. a noun noun compound, and its constituent nouns, noun1 and noun2. Thus each set of thirty features is itself composed of three subsets of 10 features, referred to as Feature Sets each Feature Set being the 10 most commonly produced features when subjects were asked to list the features of the noun1, noun2 and compound. The example test page in fig 5.1 is for the concept "oven", and the features are derived from "oven", "glove" and from the compound "oven glove"

In experiment 3 subjects were required to rate these lists of thirty features as to how descriptive of each of the three associated concepts each of the features was, on a scale which allowed conversion to numeric ratings 0 to 8, given in Table 5.2 below.

Table 5.2 : Categories which subjects used to rate features in experiments 1 and 2

numeric score	code entered by subjects	Category Description
8	N	necessarily true for all possible examples
6	A	a very important part of the definition
5	B	a fairly important part of the definition
4	C	typically true but not defining
3	D	not usually true
2	E	very unlikely, but possible
0	I	impossible for any example

Experiment 3 collected four ratings from different subjects for each of these features with respect to each item. These ratings were used to determine the feature categories for this study. Features were assigned to the appropriate categories based on a function of the average (i.e. mathematical mean) and the mode (i.e. the most frequent) ratings given to them in experiment 3. Details of this are given in Table 5.3 below.

Table 5.3 : Rules for assignment of features to categories

- If three scores are the same, and the difference between this value and that of the fourth score is greater than 2, then the outlying score was discarded.

- The mean of all remaining scores was calculated (x).
  - If $x<1$, a value of 0 was recorded
  - If $1 \leq x<2$, a value of 2 was recorded
  - If $x>7$, a value of 8 was recorded
  - If $6<x\leq 7$, a value of 6 was recorded

- x was then rounded to the nearest integer, and this value converted back to the corresponding category assignment.

Since this study used the data obtained from experiments 2 and 3, the materials varied along the same factors. There were therefore two values of Extensional Status - real and novel. Each compound was also one of four possible Compound Compositions, either artefact artefact, artefact natural kind, natural kind artefact or natural kind natural kind. Each questionnaire contained items from only one level of each of these factors. Questionnaires reflecting all possible permutations of the factors Extensional Status (2 levels), Compound Composition (4 levels) and Compound Part (3 levels) were produced, resulting in a total of 24 different questionnaires.

Design

This study was of a 2 (Extensional Status) x 4 (Compound Composition) x 3 (Compound Part) x 3 (Feature Set) design. Repeated measures were used on the last two factors by materials, and on the final factor only by subjects.

Procedure

Subjects were given the questionnaires, and asked to read the instructions carefully. The instructions informed subjects that each of the following pages contained features which had been categorised to form a kind of `profile' of concepts, and it was their task to identify what each of these concepts was. The instructions actually varied slightly according to the test condition, in that subjects were told whether the target was a one or two word concept, and also they were told that the concept was novel in the novel/compound condition. A short example was also given (not a genuine test item), and this too varied so as to accord with the test condition. Subjects were asked to indicate what concept they thought the features described by writing it at the top of the page. They were also asked to indicate which of the features had been useful to them in coming to this decision by placing a tick next to them. Finally, they were asked to make any further comments about how they arrived at their decisions on the few blank lines at the foot of each page. Subjects were then left to complete the questionnaire unsupervised, and the experimenter returned to collect it later.

Results

Correct concept identification

Concept identification was recorded on two different measures, *precise* and *liberal*. If the exact wording of the target was used by the subject, then this was correct by the *precise* criteria. The *liberal* measure includes all of the precisely correct responses in addition to those responses which are either semantically very close to the precisely correct answer (e.g. "falcon" when "hawk" was the target, and "pen" when "pencil" was the target) or it had the correct head noun. Responses were judged to have the correct head noun if the response was a compound which had the target noun as its noun2 (e.g. "oven glove" when "glove" was the target).

Table 5.4 below shows the number of correct responses in each category for noun1 and noun2, by Compound Composition. Each Compound Composition was represented by 6 items, and each of these was tested with 6 subjects, so for each condition N = 36.

Table 5.4 : Correct responses for Noun1 and Noun2 by Compound Composition and Extensional Status

		Noun1		Noun2	
		precise	liberal	precise	liberal
AA	real	15	28	27	34
	novel	21	23	23	30
AN	real	31	33	31	35
	novel	30	31	32	35
NA	real	35	36	2	34
	novel	13	19	27	32
NN	real	30	30	18	28
	novel	22	26	31	32

Overall, there appears to be little difference between the number of correct responses recorded for artefacts and natural kinds, as shown in Table 5.5.

Table 5.5 : Comparison of Number of Correct responses between artefacts and natural kinds

artefacts		natural kinds	
precise	liberal	precise	liberal
176 (61%)	244 (85%)	212 (74%)	241 (84%)

Similar data was recorded for the compounds. In these cases, the liberal count includes only the precise plus semantically close responses. Two other categories of response were recorded in this condition. If the response was only a single word, and if that word was the noun2 of the target compound (e.g. "kettle" when the target was "fish kettle"), then this was recorded as *head only*. If the response gave the correct two nouns, but in reverse order to the target (e.g. "pineapple whale" if the target was "whale pineapple"), this was recorded as a *reversed* response. These data are given in Table 5.6. To give an idea of how many responses were in the broad semantic area of the target answer, *reverse*, *liberal* and *head* scores have been summed in the *overall* column.

Table 5.6 : Correct Responses for the Compound by Compound Composition and Extensional Status

		reverse	precise	liberal	head	overall
AA	real	0	24	24	6	30
	novel	2	6	8	9	19
AN	real	0	32	33	2	35
	novel	6	6	6	12	24
NA	real	0	17	20	10	30
	novel	11	3	7	8	26
NN	real	0	13	13	15	28
	novel	8	5	7	3	18
Total	real	0	86 (60%)	90	33	123 (85%)
	novel	27	20 (14%)	28	32	87 (60%)

Two 4 (Compound Composition) x 1 χ^2 tests were carried out on these data in the real condition, one for the precise and one for the liberal data. In both precise and liberal counts χ^2 was significant by α at the 0.05 level. More correct responses were given in the AN compounds than in the others.

Precise measure : $\chi^2 = 9.29,\ 0.05 > p > 0.02$

Liberal measure : $\chi^2 = 9.72,\ 0.05 > p > 0.02$

Because of the small number of correct responses in the novel condition no such test was carried out for those data.

Number of features used to correctly identify concepts

The number of features used by subjects to correctly identify the concepts was also recorded. These data were analysed in a 4 (Compound Composition) x 2 (Extensional Status) x 3 (Part) x 3 (Feature Set) analysis of variance, with repeated measures on the final factor by subjects (F1), and repeated measures on the last two factors by materials (F2). Similar analyses were also conducted for the number of features used in identifying concepts for all results (i.e. including those cases in which the concept was *incorrectly* identified). The details of these analyses are given in the appendix (tables A7.3 - A7.5). all of the significant findings from the *correct* results analysis were also significant in the *all results* analysis, and operated in the same direction. Because this analysis only included data from those cases in which the subject correctly identified the target, some data points from the original data set were treated as missing. This meant that 10 cases were excluded from the subjects analysis (leaving 134 valid cases), and 107 cases were excluded from the items analysis (leaving 37 valid cases).

Part was a significant factor (min F'= 6.606, df = 2, 163, $p<0.01$), as was Feature Set (min F'= 10.062, df = 2, 141, $p<0.01$). No other main effects were significant.

Part x Feature Set was a significant interaction (min F' = 90.857, df = 4, 273, $p<0.01$). No other interactions were significant by min F', but Part x Compound Composition (F1 = 2.28, df = 6, 110, $p < 0.05$; F2 = 2.99, df = 6, 58, $p<0.05$), and Part x Compound Composition x Feature Set (F1 = 2.32, df = 12, 220, $p<0.01$; F2 = 3.02, df = 12, 116, $p<0.01$) were both significant by both F1 and F2 analyses independently.

Table 5.7 : Mean number of features used to reach correct answer by Part x Feature Set x Compound Composition

	AA			AN			NA			NN			mean
	fs1	fs2	fs3	fs1	fs2	fs3	fs1	fs2	fs3	fs1	fs2	fs3	
noun1	4.21	0.19	0.62	4.90	0.31	0.81	4.19	0.08	0.74	3.94	0.15	0.57	
fs sum		5.02			6.02			5.01			4.66		5.18
noun2	0.51	3.94	1.35	0.13	3.55	1.29	0.13	3.63	0.70	0.77	4.85	1.43	
fs sum		5.80			4.97			4.46			7.05		5.57
comp	0.77	2.14	3.20	1.67	1.74	2.37	2.89	3.35	3.86	2.77	4.44	2.24	
fs sum		6.11			5.78			10.10			9.45		7.86
mean	1.83	2.09	1.72	2.20	1.87	1.49	2.40	2.35	1.77	2.49	3.15	1.41	
mean		5.64			5.56			6.52			7.05		6.19

Concept type of the constituents (artefact or natural kind) was not specifically addressed in this experiment, so no analysis is conducted for this comparison. However, the mean number of features used to correctly identify the constituents was calculated, and it was found that approximately the same number of features was used in each case: Artefacts = 5.33, Natural Kinds = 5.43.

Inspection of Table 5.7 reveals that considerably more features were used in correctly identifying the compound than either noun1 or noun2. Also, features from feature sets 1 and 2 are more frequently used than those from feature set 3 (fs1 = 2.27, fs2 = 2.39, fs3 = 1.60).

The Compound Part x Feature Set interaction is shown in fig 5.2. This shows that for each Compound Part the corresponding Feature Set provides the largest number of features.

Fig 5.2 : Mean number of features used to reach correct answer for Part x Feature Set

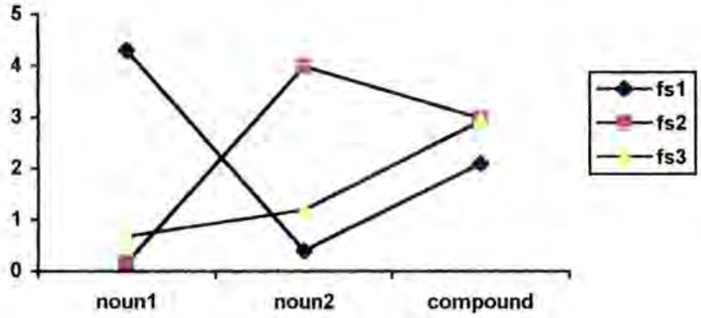


Fig 5.3 shows the Compound Part x Compound Composition interaction. This appears to be due to the large difference in the number of features used to correctly identify the AA and AN compounds and the number used for the NA and NN compounds. Subjects used more features for the NA and NN compounds.

The 3 way interaction of Compound Part x Feature Set x Compound Composition is shown in fig 5.4. Feature Set clearly modifies the finding that NA and NN compounds required the greatest number of features to be correctly interpreted. The effect holds for fs1 and fs 2 (although here NN requires more than NA) but not for fs3. In fact fewer fs3 features were used for NN compounds than for any other Compound Composition.

Fig 5.3 : Mean number of features used to reach correct answer for Part x Compound Composition

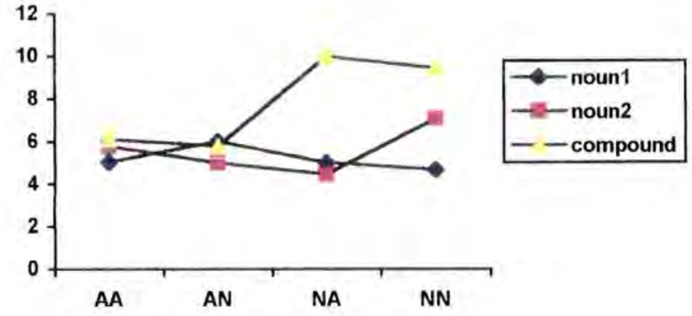
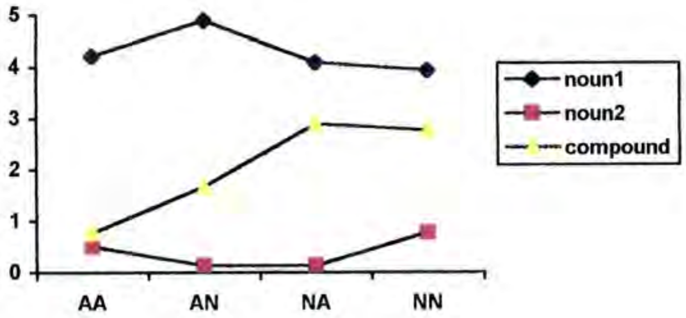
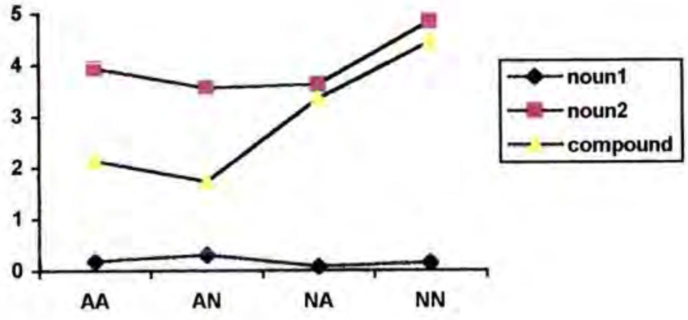


Fig 5.4 : Mean number of features used to correctly identify each concept for Compound Part x Feature Set x Compound Composition

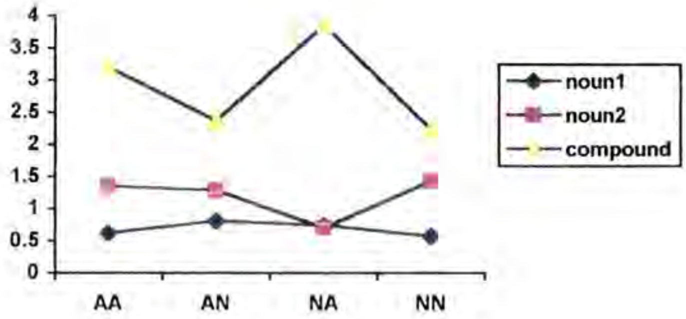
FS1



FS2



FS3



Discussion

This experiment addressed issues relating to the identification of a range of concept types. Two measures involved were how many concepts were successfully identified, and how many features were used to identify them. The results of these two measures are discussed, first with respect to simple concepts, and then compounds.

The comparison between artefact and natural kind concepts shows that there is no clear difference in how easy it is to identify these concepts (correct identification (liberal count): artefacts = 85%, natural kinds = 84%). So, by this measure at least, the two kinds of concept behave similarly. However, this result should not be interpreted as implying that no structural difference exists between artefacts and natural kinds. Theories which postulate difference between the cores of artefacts and natural kinds do not predict a difference in ease of identification. In fact, it has been suggested that identifiability is determined more by the ordinate level of category (i.e. superordinate, basic level and subordinate) than by semantic classification (Rosch, 1978). The materials in this study were not controlled for their level, but were mostly basic level concepts.

It was also found that the same number of features was used to correctly identify artefacts and natural kinds (5.33 and 5.43 respectively). The lack of any difference here is rather more of a challenge to the idea that artefacts and natural kinds are differently represented. It might be expected that artefacts could be identified by a single feature - their function, whereas natural kinds would be identified by clusters of features. Also, in experiment 6 of this thesis it was found that subjects listed fewer features as "defining" for artefacts (mean = 5.9) than for natural kinds (mean = 8.0). The result in the present study however suggests that either no such difference exists between artefacts and natural kinds, or that if it does exist then the process of concept identification is insensitive to the difference. A mean number of features used of around 5.4 suggests that subjects used an identification technique which relied upon a handful of indicative features rather than regarding a single property as absolutely diagnostic.

But it is also possible to offer an account of this result while retaining the idea that artefacts and natural kinds have different essences. Importantly, the subjects' task is this experiment is one of concept identification. The information provided about each item may be peculiarly informative, for example it might explicitly state the intended function of some item. This sort of information is often difficult to come by in realistic encounters with the world. As such, subjects may choose to ignore it, and rely on cues which are more usually available in real-life concept identification, such as shape, size, texture etc. This account therefore suggests that whatever differences may exist between artefact and natural kind concepts at a representational level, in terms of identification they tend to rely upon a similar numbers of features.

Of the compound concepts, the real compounds were much more likely to be identified, as predicted. There are several factors which may contribute to this.

First, it is likely the case that the feature list or `profile' given for each item was inherently more coherent, and so more helpful, for real compounds than for novel ones. This is a consequence of the way the profiles were produced. In the generation of these profiles, subjects were asked to list, and to rate for typicality, features of both real and novel compounds. Since the novel compounds would have some degree of ambiguity, and the real compounds would be virtually unambiguous, it is inevitable that the profiles of real compounds will be more consistent and coherent. Secondly, the real compounds are terms which describe things which definitely do exist, and which subjects are likely to be familiar with. This is unlikely to be the case in the novel condition. It is also important that each real compound is the privileged label for a particular reference. Thus, if the reference has been successfully understood from the given profile, then the subject need only recall the appropriate linguistic label. In the case of novel compounds, subjects who did manage to apprehend the correct reference of the profile had no single phrase to label it with, so they are then faced with the task of thinking up an appropriate label. This, then, is a difference in how easy it is to be precise about the identification.

If this difference in the ease of precision were the only effect in operation here, then the difference should be eliminated by including all those responses which were somehow imprecise, but were in the correct semantic area. These totals (Table 5.6, `overall' figures) show that the difference in ease of identification between real and novel compounds persists even when using the most general criteria. This suggests that precision is not the critical factor, and that the other two influences alluded to above (profile coherence and familiarity with the target concept) do account for the difference in identification between real and novel compounds.

Compounds were also categorised according to the semantic categories of their constituent nouns. Too few of the novel cases were correctly identified to observe any effect, but in the real condition there did prove a difference between Compound Compositions. The fact that there is no such difference between simple artefacts and natural kinds suggests that the difference between compounds is the result of interaction between the constituents, rather than an effect simply inherited from the head nouns. This is borne out by the fact that differences between the four compound types are not in accordance with the type of noun2; number identified: AN > AA > NA > NN. The compounds in fact seem to divide into the AN and AA compounds, which are quite well identified (correct in 32 and 24 cases respectively), and the NA and NN compounds which are only successfully identified in about half as many cases (correct in 17 and 13 cases respectively).

Examining these results along with those concerning the number of features used by subjects to identify compounds, a pattern begins to emerge. Subjects used just over half as many features in successfully identifying the AN and AA compounds (6.10 and 5.79) as they did to identify the NA and NN compounds (10.01 and 9.45), as illustrated in fig 5.3. This suggests that subjects need to attend to more features to identify NA and NN compounds than to identify AA and AN compounds. It appears that this extra required effort means that fewer NA and NN compounds are correctly identified. The questions remain as to which extra features are needed in the NA and NN cases, and why. Data for the compounds by the factor Feature Set (illustrated in figure 5.4) shows that the most marked increase is from feature sets 1 and 2 (and also from fs3 for in the NA condition). The extra information required to identify the NA and NN compounds is therefore mainly compositional (fs1 and fs2) rather than noncompositional (fs3). It is not possible, from the data of this experiment, to come to a firm conclusion about the underlying cause of this effect. However, it is the case that the type of modifier varies along with the likelihood of successful identification and the number of features used. Compounds with an artefact modifier are better identified and require fewer features than those with a natural kind modifier. This suggests that the role of the modifier within the compound may be central to the effect.

In summary, no difference was found between how well simple artefact and natural kind concepts are identified, nor in how many features were used to correctly identify them. This might suggest that there is no difference in the number of defining features in each case, but might also be due to subjects using an identification strategy which is relatively inefficient in the context of this experiment. It was also found that real compounds were better identified than novel ones. Three possible reasons were considered; the coherence of the feature profiles, subject familiarity with the target compound, and the linguistic precision of compounds. The latter of these can be ruled out as the sole cause of the effect, since the difference persists even when less precise responses are counted. Finally, among real compounds it seems that two Compound Compositions, NA and NN, are harder to correctly identify than the other two, in that fewer are correct, and of these more features were needed to identify them. The common factor between these two less well identified Compound Compositions is that they have natural kind modifiers. It is possible that this has some effect on the ease of identification of the compounds.

Experiment 8 : Constituent Position

Introduction

This experiment addresses an issue about compound nouns on which a number of authors seem to have made assumptions, but which has been the subject of little empirical research. That is the question of which, if either, of the constituent positions in a compound is more influential in determining the overall relation between the nouns. There would appear to be three possibilities. Either noun1 is of principle importance in determining the overall relationship, or noun2 is, or there is an interaction between the two. Clearly this latter situation is true to some extent, since different relations do exist between pairs of compounds with the same noun1s or noun2s; consider **elephant shrew** and **elephant gun**, and **bulldog clip** and **paper clip**, each pair contains one property mapping relation and one slot filling relation. More interesting is the question of whether there is any significant difference in the relative importance of the two positions in determining the overall relation.

There are theoretical arguments to be made in support of each position being more important than the other. Noun1 is in the leftmost position, and as such is the first element of the compound to be encountered, it could therefore have a primary influence on the compound's structure. From a linguistic point of view, noun1 is also the modifier, in that it supplies the reference of the modification. This being the case, it might also be expected to specify the type of modification. There is some empirical evidence in support of this. Gagné and Shoben (unpublished manuscript, 1995) compared the influences of noun1 and noun2. Using data from an earlier study (Medin and Shoben, 1988) they identified types of slot filling relationships that certain nouns entered into as noun1s or noun2s either frequently or infrequently. From these data, they composed compounds based on whether the frequency was high (H) or low (L) for each of the nouns, such that they had HH, HL and LH combinations (not enough data was available to generate LL compounds).

Subjects were presented with these compounds, and asked to judge if they were meaningful, with the response time being recorded. Clearly HH compounds would be expected to be accepted rapidly. Gagné and Shoben further reasoned that if noun1 was critical to the compound's relation, then reaction times for HH and HL compounds would be similar, and shorter than those for LH compounds. Conversely, if noun2 were critical, then HH and LH would be fastest, and HL slower. If neither was important alone and only the combination was critical, then HH compounds should be judged faster than LH or HL. They found that HH and HL compounds had very similar judgement times for correct responses, both being significantly faster than LH compounds. From this they conclude that noun1 is critical in determining the relation in nonpredicating compounds. However, this interpretation maybe going a bit too far. Clearly the data implies that the frequency of relation is more important with respect to the modifier than to the head in terms of how rapidly we make sense of compounds. But this is not the same as claiming that the modifier is the critical element in determining the meaning of compounds.

In a further experiment, Gagné and Shoben include a comparison of **nonword - word** compounds with **word - nonword** compounds in a lexical decision task. They found that compounds with a nonword modifier are more rapidly rejected than those with a nonword head. They interpret this as being due simply to subjects encountering the modifier first, and so being able to make an early decision to reject compounds with a nonword modifier. This result and explanation hint at an alternative interpretation of the data from their initial experiment. The subjects' first hypothesis as to the compound's relation will probably be based on what is a likely relation for the first encountered noun (i.e. the modifier) to enter into, and for HH and HL compounds this is likely to be the correct choice. For LH compounds this initial hypothesis will be incorrect, and if the subject is to interpret the compound correctly, the hypothesis will have to be rejected, and an alternative sought. This process of rejecting the initial supposed relation and generating a new alternative will inevitably mean an increase in the time taken to judge whether or not the compound is meaningful. There is, however, no good reason to suppose that the modifier has a greater influence in determining the ultimate interpretation of the compound just because it has the initial input in the sequence of processing. Gagné and Shoben therefore seem to have demonstrated that the modifier does provide us with an initial relation to consider for a compound, but it is not clear that this carries any more weight than the influence of the head noun in determining the ultimate relation when a meaning for the compound is resolved.

In fact, there are theoretical reasons to suppose that noun2 could be more influential in determining the ultimate interpretation of compounds. Since noun2 acts as the head of compounds, it defines the central reference of the compound, in a sense defining just what kind of thing it is that the compound refers to. In characterising the compound in this way, it might reasonably be supposed that the head would impose the more severe limitations on a compound's interpretation. Again, there is some empirical evidence to support this view. Shoben (1993) has shown that the interpretation of a novel compound can be influenced by prior encounters of contextualised compounds with the same head noun. Subjects read short passages which contained one of two novel compounds, which had different noun1s, but a common noun2. Each context supported a different relation in the compounds, e.g. **cat rash** - a rash caused by a cat (causal relation), **hamster rash** - a rash on a hamster (locative relation). Afterwards, the subject had to define a third compound with the same noun2, e.g. **horse rash**. It was found that an original bias towards one relation for a given compound was shifted towards the relation previously encountered in a compound with the same noun2. So although **horse rash** was interpreted by most people to contain a locative relation, more subjects used this relation if they had previously seen the **hamster rash** passage (locative) than if they had seen the **cat rash** passage (causal).

This demonstrates that we can be influenced by analogy to compounds with the same noun2 when we interpret novel compounds. Shoben describes the process as a kind of priming of a certain thematic role[10] for a head. Clearly, there is some influence here. It is not possible to say however whether this kind of priming or interpretation by analogy can occur over periods greater than the few minutes involved in such an experiment. It is plausible that such a mechanism might be limited to operating within the limits of the average discourse, rather than spanning days and weeks. On the other hand, it's quite possible that a novel compound will commonly trigger a search for a compound with the same head whose meaning is known, and may have been learnt some considerable time ago. Indeed some vivid recent coinages seem to require just this mechanism, e.g. **trolley rage** is a supermarket equivalent of **road rage**, while a **net potato** is the internet version of a **couch potato**. But such compounds may well be the exceptions rather than the rule, their rhetorical impact may result from their analogical interpretation being a somewhat unusual strategy.

Others have also investigated the possibility of compound interpretation using analogy. Ryder (1994) was able to some extent to predict the interpretation of novel compounds using analogy to lexicalised compounds. However, the analogies drawn could be between whole compounds (i.e. both noun1 and noun2

---

[10] Shoben uses the term "thematic role" to refer to the set of roles a noun might participate in (e.g. causal, locative, etc.). In other contexts, thematic roles are specifically associated with the arguments of a verb. Thus a given noun may fill the agent thematic role in one sentence, the patient role in another and so on. This is not how Shoben uses the term.

simultaneously), or just a single noun in common, or, at the most abstract level just between compounds with similar nouns. For example, at the whole compound level, Ryder suggests that we may interpret the novel compound **water sock** in accordance with a linguistic template ***fluid* sock**, derived from the real compound **wind sock**. At the intermediate level **water sock** might be interpreted by analogy to **water bed**. At the most abstract level **water sock** could be interpreted by analogy to **snow suit** or **rain coat**.

Because the analogies were so broad, similar results would seem to be attainable by predictions based on semantic plausibility rather than analogy to linguistic templates. Van Jaarsveld et al (1994) used a lexical decision type task to look at analogy processes online. They found evidence of priming between lexicalised and novel compounds, for both noun1s and noun2s indiscriminately in experiment 1, and for compounds with common noun2s and related noun1s in experiment 2. However, this simple priming effect did not interact with productivity in experiment 1 (i.e. how common noun1 and noun2 are in lexicalised compounds) or with the frequency of the lexicalised analogue in experiment 2 (van Jaarsveld et al suggest that a frequently occurring constituent ought to be a salient analogue). This lack of interaction leads the authors reject the idea that compound interpretation is analogy based, since the effect of priming is not subject to the size of the search (experiment 1) nor to the apparent salience of a suitable analogue (experiment 2).

Apart from these studies, little empirical work has been conducted either in the area of interpretation of compounds via analogy, or into the question of whether noun1 or noun2 is more significant in determining the relation of a novel compound. Most importantly, these studies tend to have examined the role of only one constituent, or of both in a design where independent effects are undetectable. Only the Gagné and Shoben study has actually compared the roles of the two constituents in the same design, but as discrete factors. The issues of analogy and the possible priority of one constituent over the other seem to be somewhat tied. If one noun is more important than the other, it seems plausible that it effects this influence by analogy to similar lexicalised compounds.

This experiment uses a simple paradigm designed to compare noun1 and noun2, to see if either has a greater effect on the interpretation given to a novel compound than the other. Subjects are simply asked to paraphrase a number of novel compounds. Each compound is part of a triple, in which the central (original) compound shares a noun1 with one member of the triple, and a noun2 with the other. This will allow a comparison between the two types of altered compound, to see whether those with the same noun1 as the central compound, or those with the same noun2 tend to have the same relation.

Method

Subjects

Twenty-four unpaid volunteers participated in this experiment. They were all undergraduates studying psychology at the University of Durham. There was an approximately equal number of males and females, aged 18 - 30 years.

Materials

There were three sets of questionnaires, with eight questionnaires in each set. Each of the questionnaires in any one set contained the same items. Items were in triples (**original, same noun1** and **same noun2**; e.g. **bee bag, worm bag** and **worm razor**), and no items from the same triple were in the same questionnaire, they were split between the three questionnaire sets.

Compounds were composed of natural kind (N) and artefact (A) nouns in each of the possible combination types (AA, AN, NA and NN), ten items in each triple in each combination type. These combination types were equally represented in each questionnaire. The sets of ten natural kind nouns were subcomposed of seven animals (e.g. **baboon**, **lobster**), two living non-animals (e.g. **fern**, **mushroom**), and one inanimate natural kind (e.g. **star**, **mountain**). A full list of materials is in the appendix, table A8.1.

Each questionnaire contained four pages of ten compounds, in half of the questionnaires the pages were in one order, in the other half the order of pages was reversed.

Design

The experiment was of a 3 (Noun in Common) x 4 (Compound Composition) design, with repeated measures on the first factor by subjects, and on the last factor by items.

Procedure

Subjects were given the questionnaires described above, and asked to complete and return them in their own time. Each questionnaire took around 15 minutes to complete. There was an instruction page fixed to the front of each questionnaire, containing the following instructions:

"On each of the following pages are several different phrases, which are underlined and enclosed in separate boxes, along with some dotted lines. Each phrase is made up of two words, and names a type of object. I want you to write a *definition* of each of these, that is, write briefly and in your own words, another way of saying what you understand by this phrase. Please write this on the dotted lines directly below the relevant phrase.

The phrases might not refer to real things, and might not be familiar to you. If this is the case, then try and imagine what the object might be like if it did exist, and write a definition of it as though it did actually exist.

You need not take long over this, but please try to give your own opinion accurately and clearly. Please do not leave any blank, if you feel that some of the phrases are not possible to provide a definition for, then just write down your best attempt."

Results

Each of the definitions generated by the subjects was categorised according to the relation that it asserted between the two nouns of the compound. There were two main classes of relationship; **property mapping** and **slot filling**. Some definitions were otherwise categorised, details of these other categories are given in Table 5.8 below. In some cases a subject would offer multiple definitions of one compound. If this was so, only the first definition was considered.

Those definitions that implied a slot filling relationship were further categorised according to the nature of that relationship. This finer categorisation was based upon the taxonomy suggest by Shoben (1991). However, some kinds of slot filling relation were recorded which were not listed by Shoben. Of these several were deemed frequent enough to warrant recognition in their own right, these are given in Table 6.1 below. Other of Shoben's categories were slightly adapted to better describe the data. Slot filling relations not listed by Shoben, and appearing only infrequently in the data were simply classified as **novel slot filling**.

In order not to cause the categories to proliferate any further, inclusion into existing categories was really quite liberal. For example, phrases meaning located in, on, at and living in were all included in the **located in** category.

Table 5.8 : Categorisation of definitions, with examples from the data.

		Category	e.g. Compound	E.g. Definition
Property Mapping				
	†	property mapping	bullet virus	"a virus that strikes suddenly"
Slot Filling				
		n2 is about n1	taxi magazine	"a magazine describing taxis"
	†	n2 belongs to n1	baboon piano	"a piano belonging to a baboon"
	†	n2 eats n1	mushroom wolf	" vegetarian wolf..."
	†	n2 is eaten by n1	badger grass	"what badgers eat to aid digestion"
		n2 has n1	hippopotamus beer	"make of beer with hippo logo"[11a]
		n2 is located in n1	igloo scorpion	"insect found in igloos"
		n2 is location of n1	fossil box	"box containing fossils"
		n2 makes n1	basket canary	"a canary which makes a basket like nest"
		n2 is made of n1	camel whisk	"made of camel hair, to be used for beating carpets..."
	†	n2 sells n1	jug shop	"a shop that sells jugs of different types"
		n2 uses n1	tractor penguin	"a lego piece shaped like a penguin which will fit onto the tractor"[11b]
		n2 is used by n1	ant fort	"a fort governed by ants"
	†	n2 is used for n1	bee bag	"a bag to put bees in"
	*	novel slot filling	beach cat	"a cat that likes the beach"
Other				
	†	hybrid	fox buffalo	"a cross-species, about the size of a deer"
	*†	missing	-	
	*†	noun1 is the head	cup dolphin	"a type of cup"
	†	structure mapping	tractor locust	"locust with back legs much bigger than the front legs"
		superordinate class	worm razor	"shaving with a worm"
	*†	unclear relationship	deer mole	"a mole"
	*†	interpreted as a verb	shed toad	"type of toad which sheds its skin like a snake"

* Nonspecific categorisations, excluded from analysis.

† Categories not included in Shoben's 1989 taxonomy.

[11a&b] These definitions also involve `construal'.

The main purpose of the present study was to see if compounds with a noun1 in common were more or less likely to have similar definitions than compounds which have a noun2 in common. Therefore, each definition for the Noun1 in Common and Noun2 in Common compounds was compared to the definitions of its counterpart in the Original compound condition. The number of instances in which the common noun compound's definition category was the same as the original compound's definition category was recorded. Since there were eight definitions given to each compound (one provided by each of the subjects who saw that item), the maximum number of matches possible was eight. The only categories excluded from this count were those which did not genuinely pertain to a specific type of relationship between the compound constituents, such as the categories **missing data** or **unclear data** (see items marked * in Table 5.8 for a full list). It would be misleading to count correspondences of these nonspecific categorisations as definition matches.

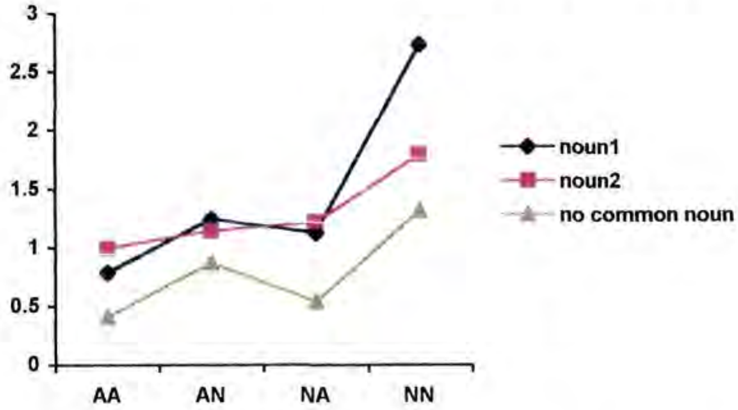
In order to see if either of the two common noun conditions showed more matches than would arise by chance, a further level was added to the common noun factor. This extra level - 'No Noun in Common' - is the number of definition type matches between the Noun1 in Common compounds and their respective Noun2 in Common compounds. These compounds do not share any common nouns with one another. All that they do have in common is that the constituent nouns are of the same superordinate types (i.e. natural kinds or artefacts). This third level therefore provides a baseline control condition.

The mean number of valid matches for each item is shown in Table 5.9. These data were submitted to a 3 (Noun in Common) x 4 (Compound Composition) analysis of variance, with repeated measures on both factors by subjects and on the final factor by items. The main effect of Compound Composition was significant by both subjects and items analyses (Min F' = 5.57, df = 3, 80, $p < 0.01$). The factor Common Noun was significant by F1 and F2 independently, but not by minF' (F1 = 10.28, df = 2, 46, $p < 0.0005$; F2 = 3.27 df = 2, 72, $p < 0.044$). The interaction Compound Composition x Common Noun was also significant (Min F' = 2.50, df = 6, 152, $p < 0.05$).

Table 5.9 : Mean number of matches per item per subject by Compound Composition and Noun in Common

	AA	AN	NA	NN	mean
noun1 in common	0.79	1.25	1.13	2.74	1.48
noun2 in common	1.00	1.15	1.22	1.80	1.29
no noun in common	0.41	0.87	0.54	1.32	0.79
Mean	0.73	1.09	0.96	1.95	1.19

Fig 5.5 : Interaction between Compound Composition x Noun in Common



Inspection of figure 5.5 shows that the main effect of Compound Composition is mainly due to the NN compounds, but is somewhat graded, with AA having fewest matches, NN having most and AN and NN being intermediate between them. The figure also reveals that the main effect of Noun in Common appears to be due to the No Noun in Common level (the control condition) having fewer matches than either of the other two levels, since the Noun1 in Common and Noun2 in Common conditions are very similar.

The one exception to Noun1 in Common and Noun2 in Common levels being almost identical is the case of NN compounds. The interaction Compound Composition x Noun in Common results from the increase in matches for NN compounds being more extreme in the Noun1 in Common condition than in the other conditions. It is possible that both of the main effects might depend entirely upon the NN condition. In order to test this a further analysis was conducted from which NN compounds were excluded. In all other respects this second analysis was similar to the initial one.

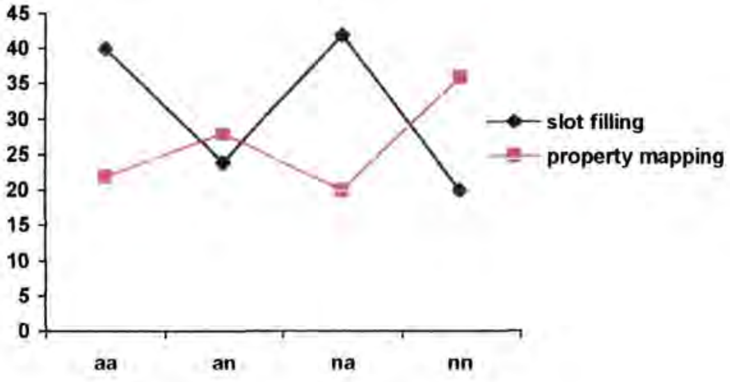
This second analysis found that the Noun in Common factor was significant by subjects ($F1 = 5.42$, $df = 2, 46$, $p < 0.05$), but marginal by items ($F2 = 2.55$, $df = 2, 54$, $p < 0.087$). No other effects approached significance. It therefore appears that the significant main effect of Compound Composition and the interaction of Compound Composition x Noun in Common were due to the NN compounds. The main effect of Noun in Common probably obtains across all Compound Compositions.

Table 5.10 shows the distribution of property mapping and slot filling definitions in these data (illustrated in fig 5.6). Clearly there is some effect of compound composition operating. For AA and NA compounds slot filling is about twice as common as property mapping, for NN compounds this is reversed, and for AN compounds the numbers of slot filling and property mapping definitions are approximately equal.

Table 5.10 : Numbers of slot filling and property mapping definitions in each condition, out of 80.

	AA		AN		NA		NN		Mean	
	SF	PM	SF	PM	SF	PM	SF	PM	SF	PM
original	37	27	31	24	36	24	10	43	29	30
noun1 common	45	18	21	29	49	17	31	31	37	24
noun2 common	38	20	19	31	41	19	20	44	30	29
mean	40	22	24	28	42	20	20	36	32	28

Fig 5.6 : Numbers of slot filling and property mapping definitions in each condition, out of 80.



Discussion

The first analysis shows an interaction Noun in Common x Compound Composition, and main effects of both factors. In the second analysis, from which the NN compounds were omitted, only the main effect of Noun in Common remains (and this is only marginal in the items analysis). This implies that the original Compound Composition x Noun in Common interaction, and the main effect of Compound Composition were both due to the NN compounds.

The main effect of Compound Composition can be explained in terms of the preferred type of interpretation for NN compounds. In experiments 5 and 6, it was found that property mapping was preferred for NN compounds, whereas slot filling was preferred in other Compound Compositions. Figure 5.6 shows that the same is true in this experiment. Since all property mapping interpretations were counted as matches to each other (unlike the slot filling compounds which were subcategorised and counted as matches accordingly) , it is unsurprising that NN compounds, for which property mapping is most frequent, also show the most matches. Such a result could therefore be regarded as an experimental artefact. It could perhaps be eliminated by subdividing the property mapping interpretations into categories relating to the specific property concerned (e.g. colour, size, etc.). However, there is no real theoretical justification for such a division. Many of the slot filling categories could themselves also be subdivided, for example "n2 located in n1" included a variety of relations, such as "located in", "located on", "located at" and "lives in" etc.

The NN compounds have already been identified as the source of the interaction Noun in Common x Compound Composition (there is no interaction when the NN compounds are omitted). Although all three levels of the Noun in Common factor exhibit an increase in interpretation matches for NN compounds, the increase is most marked in the Noun1 in Common condition, as shown in figure 5.5. Having established that the increase in interpretation matches among NN compounds is probably due to the high frequency of property mapping it is interesting to note this has its greatest effect for compound pairs which share a noun1. It seems credible that the noun1 is more important than the noun2 in determining whether or not a property mapping interpretation is appropriate.

The central aim of this experiment was to investigate the relative importance of the two constituent positions in determining the overall relation between the two nouns of a compound. Figure 5.5 shows that, with the exception of NN compounds, the Noun1 in Common and Noun2 in Common levels are almost identical. There is virtually no difference between them. The main effect of Noun in Common, which persists (albeit marginally on F2) when NN compounds are omitted, is due to both these conditions having more interpretation matches than the No Noun in Common condition. So although neither constituent position can be said to be more influential than the other, it seems that each does have some effect. Indeed, the baseline provided by the No Noun in Common condition may itself be a little higher than a true baseline, since it doesn't provide a comparison between entirely unrelated compounds, but compares compounds of the same Compound Compositions (e.g. NA compounds were compared only with other NA compounds). Such a comparison corresponds to the most abstract level of compound analogy hypothesised by Ryder (1994) discussed above, and so interpretation matches even for this condition may be above a true chance level.

Given that both Common noun1 and Common Noun2 levels do show more matches than the control condition, what is the best account of this phenomenon? The consistency of interpretation between compounds sharing a common constituent might be due to our interpreting novel compounds by analogy with other more familiar compounds, as suggested by Ryder (1994). However, the same result could be accounted for by constituent concepts being limited with respect to what relationships they are likely to enter into by their nature. To borrow an example from Ryder (1994), there are many compounds with the noun1 `fire' - fire man, fire engine, fire hydrant, fire station etc. In each case the relation of the noun2 to `fire' is the same, a general paraphrase being something like 'a noun2 used to combat fires'. Thus when we encounter a novel compound, 'fire X' we can recall these analogous examples and interpret it accordingly. Alternatively it could be that the predictability comes not from the external relations of the word, but from the meaning of the word itself. Fire tends to be a dangerous phenomenon, something which society guards against and fights. It is therefore very appropriate for us to interpret a novel compound "fire X" in accordance with the schema of fire fighting. In the terms of Shoben, (1993) nouns tend to have a limited number of plausible thematic roles. This study does not provide data which can distinguish between these accounts.

In summary, the main effect of Compound Composition is best regarded as an experimental artefact. This is due to the way property mapping interpretations are treated in the data, and the fact that NN compounds are more likely to be interpreted by property mapping than are other Compound Compositions. That this effect is most marked in the Noun1 in Common condition may be significant. This interaction provides some support for the intuition that noun1 might be more important than noun2 in determining that a compound be interpreted by property mapping. Inspection of fig 5.5, and the persistent main effect of Noun in Common demonstrate that while neither constituent position is more important than the other in determining the overall relation between the nouns of a compound, each does have some effect. However, it is not possible to determine whether such findings result from analogical interpretation, as suggested by Ryder (1994), or by semantic limitations.

Chapter six : Final Discussion

Compositionality

Experiments 1 to 4 of this thesis examined the compositionality of noun noun compounds. Subjects listed the features of compound noun concepts, and also listed features of the simple concepts of which the compounds were composed. Compositionality was measured by comparing the feature lists (experiment 2), and by comparing typicality ratings of these features (experiments 1, 3 and 4) for each of the compounds and its constituents. The absolute differences between ratings of these features with respect to constituents and compounds was taken as a measure of noncompositionality. Compounds were either **real** or **novel**, and were composed of artefacts (A) and natural kinds (N) in each possible combination type (**AA**, **AN**, **NA** and **NN**).

Extensional Feedback (EF) : Extensional feedback is the process of modifying a representation of a concept such that it includes attributes common among instances of the concept. For example **telephone boxes** are commonly red, and so this becomes a feature in our representation of them. Clearly such a process introduces noncompositional elements into conceptual representations. Also evident is that extensional feedback can only occur if concepts have instances of the concept available to provide such information.[12] It was therefore predicted that real compounds would exhibit more noncompositionality than novel compounds due to the additional possibility of extensional feedback.

---

[12] Some nonextensional terms may be instantiated by particular representations, and so can exhibit something like extensional feedback. For example, the unicorn is said to be a horse with a single horn, and in pictures unicorns are usually white. This attribute of colour may be included in the mental representation by a process very like extensional feedback.

Maintenance of Coherence Inferences : The other hypothesised source of noncompositionality is inferences about the compound in order that it is coherent. For example, we might infer that a **pet shark** is smaller than other sharks, because of the size restrictions of living in someone's home. Such inferences tend to modify the meaning of the head of the concept (i.e. shark in this example). Because of possible differences in how we represent artefact and natural kind concepts, it was predicted that the extent to which people make inferences such as these would depend upon what sort of concepts (artefacts or natural kinds) the compound was composed of.

The results showed the following: First, novel AN compounds showed more noncompositionality than novel AA or real AA compounds, particularly for noun2 (experiment 1). Second, fewer features common to the constituents and the compound were generated for AA and NA compounds than for NN and AN compounds. This was most marked for NN compounds and for noun2 (experiment 2). Third, features of homogeneous compounds (AA and NN) were rated as more similar to the features of their constituents, than were those of heterogeneous compounds (AN and NA). As would be expected, features were most descriptive of the compound part (noun1, noun2 or compound) which they were originally listed as describing. However, this was not true of AN compounds, for which features for noun2 were most highly rated (experiment 3).

As discussed in the experimental chapters, the most obvious explanation for the lack of difference between real and novel conditions, that EF does not happen, is probably wrong. The evidence that EF is a real phenomenon is certainly quite persuasive. That wooden spoons are large is a fact we have learnt from experience. The most probable alternative is that EF does occur in the real compounds, but also that a similar effect also occurs in novel compounds, thus eliminating any difference in the amount of compositionality. A number of different researchers, using different models, have alluded to mechanisms which could account for an increase in noncompositionality in novel compounds. Murphy (1990), in his concept specialisation model suggests that combining concepts involves reference to our general knowledge of the world. This knowledge is used to infer things about the compound concept which couldn't be directly inherited from its constituent parts. For example, we might expect an `apartment dog' to have certain physical and behavioural characteristics. However, the problem with this account in explaining the current data is that this mechanism would seem to operate on **all** compounds, real and novel. But the data show no evidence of the additional noncompositionality we would expect to be caused by EF in real compounds.

Another candidate mechanism for increasing noncompositionality is construal. Wisniewski (1996) highlights the process of construal. In his data, he found that subjects would commonly interpret a compound not as being a combination of the two constituents, but as the combination of one constituent with a semantically related concept of the other constituent. For example "tiger chair" was reported as being a chair with a tiger skin seat, an "artist collector" is one who collects the works of an artist. But construal again would seem to operate equally powerfully with real and novel compounds, and so as with the use of world knowledge cannot account fully for the data of experiments 1 to 4.

Ryder (1992) proposes that we use a mechanism that she calls "accommodation". It is possible that a novel compound will not be coherent with our prior beliefs about its constituents. Indeed in using a compound we are in some sense asserting that the referent differs in some significant way from the usual meaning of its head. Ryder gives the example of a wine drinking dog. To understand the phrase, we probably need to alter, slightly, our understanding of dogs. That dogs can drink wine is unlikely to be a preconception held by many people. Clearly this mechanism would increase the apparent noncompositionality of novel compounds, since the compound has meaning which was not true of its constituents. Significantly, the situation with respect to real compounds would be somewhat different. Because we have encountered real compounds previously, what was once novel to us has by now been incorporated into our understanding of the concepts concerned. For example, on encountering the notion of a wine drinking dog for a second time, we are already aware that it is a possible feature of dogs that they drink wine. As such, accommodation is more in keeping with the present data, providing a mechanism by which noncompositionality is increased in novel compounds, parallel to EF in real compounds.

The final possibility comes from Hampton (1991). He found that when people were asked to describe certain "impossible combinations" (for example "a teapot that is a computer", or "a fish that is a vehicle") they included many diverse adaptations of the constituent concepts. It seems that people are prepared to elaborate extensively in this kind of task, introducing noncompositional features. In experiments 1 to 4 of this thesis, the novel combinations were random, and so naturally tended to be pretty unlikely / impossible. By contrast, the compounds of the real condition, by virtue of the fact that they are in common use, are likely to be more sensible combinations, and as such will not need such extensive elaboration.

Experiments 1 to 4 also looked at compositionality with respect to the ontological category of a compound's constituents. It has been suggested that artefact and natural kind concepts might differ significantly in the way they are represented. Importantly, one such proposed difference is the accessibility of the core or essential characteristics of the concept (e.g. Schwartz, 1978). Although no strong predictions were formulated about the effect compound composition might have on compositionality, differences in the accessibility of the essence of a concept might reasonably be expected to have a bearing on the process of modification. With this in mind, some effect of compound composition would not be a surprise. Experiment 1 showed higher noncompositionality in novel AN compounds than in either real AA or novel AA compounds. This result was interpreted as indicating that subjects struggled to combine the concepts, and so had to make a large number of inferences about what such a combined concept would be like. However, the result was not replicated in experiments 3 and 4. In experiments 3 and 4, the data show no overall effect of compound composition with respect to compositionality. However, a closer examination reveals that there are potentially important differences between compound compositions in how much each constituent concept has in common with the compound. Again, the novel AN compounds stood out. However, their behaviour varied in different experiments. In experiment 1, the noun1 actually had more in common with the compound than did noun2. This was counter to the pattern of other compound compositions, and clearly not what would be expected in light of noun2 being the head. In experiments, 3 and 4 however, the novel AN compounds behaved differently. In these experiments, AN novel compounds did follow the predicted pattern of noun2 having more in common with the compound than did noun1. In fact in these experiments the novel AN compounds were unusual in that the difference in noncompositionality between noun1 and noun2 was in the same direction as the other compound compositions, but more extreme.

Possible reasons for findings differing in these experiments include changes in the rating scale, different methods of compiling the feature lists, and different experimental variables. More significant is the fact that the novel AN compounds appear to have behaved anomolously in each case. This does suggest that there is something about the process of modification in compounding which is sensitive to differences between concept types. However, the results of experiments 2, 3 and 4 are probably more reliable, since feature lists were more controlled and a full range of types of combination was tested. (In fact, in experiment 2, AN and NN compounds behaved similarly. This might be explained by reference to differences between slot filling and property mapping interpretations. This possibility is addressed along with further discussion of these issues below).

Perhaps the somewhat inscrutable nature of natural kinds has a polarising effect on the way we modify them. In some cases the difficulty of assessing how the head can safely be modified will lead to it remaining virtually intact, with minimum contribution from the modifier. In other cases, as with Hampton's subjects, combining two quite different concepts could mean radical elaboration, and many changes to the head. Such a situation would account for the different results obtained in these experiments (under differing conditions), but with the AN compounds tending to behave differently to other compound types in each case.

The Implications of Noncompositionality : The compositionality or otherwise of combined concepts has become a troublesome issue within the current debate about concepts. Clearly concepts ought to be compositional, otherwise unreasonable constraints are placed on the human cognitive capacity. Some argue that, since current theories of concepts involve elements of noncompositionality, they are utterly wrong, (Fodor and Lepore, 1996).

Rips (1995) attempts to finesse the problem, by trivialising the examples of noncompositional features found in the literature. Most emergent features are, he claims, due to EF, we are told "There is something suspicious however about emergent properties based on extensional feedback" (Rips, 1995, p92). Because such features are based on prior experience, Rips argues it would be asking too much of a theory of conceptual combination to be able to generate them. He does go on to acknowledge that some emergent properties are not down to EF. He gives the example of "smoky apples" being thought to be "bad tasting", although this is not a feature of either smoky things nor of apples generally. Neither is it likely that subjects have experience of the flavour of smoky apples. But Rips claims that such emergent properties as this are, in principle, predictable or computable. We simply need rather more sophisticated models and knowledge bases. He does however concede that theoretical computability may not in itself be fully satisfactory. Time and search constraints would have to be met for any theory to be psychologically plausible. Essentially, Rips is able to claim that noncompositionality is not a problem if it is either due to EF, or it is minimal in its extent.

If, as suggested by the quantitative data in experiments 1 to 4, and by Hampton's (1990) qualitative investigation, novel compounds exhibit more extensive and elaborate emergent features, Rips arguments begin to look thin. Fodor's criticisms of current theories of concepts would thus be all the more persuasive. Perhaps the current problem is to quantify and characterise noncompositionality appropriately.

Headedness

It has commonly been assumed, often tacitly, that compounds in English are right-headed, that is that noun1 acts as a modifier to noun2 as the head (e.g. Levi, 1977; Wisniewski and Gentner, 1991). However, although this can be seen to hold for many compounds, there appears to be no empirical evidence as to its universality. Some evidence on this matter can be derived from several of the experiments in this thesis. In experiments 1 - 4 it was found that noun2 had more meaning in common with the compound than did noun1. This is interpreted as indicating that people interpreted the compounds as being right-headed. A more explicit and quantitative measure was made in experiment 5. Subjects' paraphrases of compounds were examined, and it was found that in 81% of cases the phrase referred to a kind of noun2. In 10% of cases neither noun provided the reference, and in 5 % of the cases noun1 provided the reference. So compounds would seem to be right-headed a large majority of the time, but exceptions are not impossible.

Concept Identification

Experiment 7 looked at concept identification. Following the findings with respect to noncompositionality and compound composition, it was anticipated that a concept identification task might cast more light on the ways that different types of concept combine. The first part of the experiment compared simple artefact and natural kind concepts. Both were found to be equally identifiable (85% of artefacts and 84% of natural kinds identified correctly, by a liberal count), and the same number of features (5.3 and 5.4 features per item respectively) was used to identify each.

These findings are a little surprising. They run counter to the intuition that the function of an artefact would lead to a quick and easy identification[13]. As discussed in chapter 5, part of the reason may lie in the use of non-basic level

---

[13] Malt (1992) has provided some evidence that function may not be a simple core for artefact concepts. This has been addressed further in chapter 5.

concepts, which probably require more specification. It is also possible that subjects simply failed to make use of such information in the most efficient way possible. The function of something is often not explicitly available. Unless we see something in use, only such surface features as colour, shape and weight may be available to us. For this reason, we may tend to rely on these less diagnostic cues such as appearance, and this may be the strategy which subjects needlessly adhered to in this experiment.

Of the compounds, 63% of those in the real condition were correctly identified, and only 19% of the novel ones (both by liberal counts). Even considering the ambiguity built into the feature profiles, very few of the novel compounds were correctly identified. Many compounds were generated for the novel compound concept profiles, and few were "correct". The poor performance of subjects on this identification task has two possible interpretations. Firstly, it could be taken as an indication that the feature profiles themselves provide a poor representation of novel compound concepts. The other possible reason is simply that subjects were more familiar with the concepts referred to by real compounds, and so recognised them more easily.

With the real compounds, there were more correct responses, and there did seem to be a pattern. AN and AA compounds were better identified, and fewer features were deemed necessary in this identification. Perhaps the common factor of an artefact modifier is influential in this respect. Whether or not the modifier is privileged in defining the relationship between the constituents of a compound is open to question. Gagné and Shoben (1995) provide some evidence that this may be the case, but results from experiment 8 suggest that if this is so, it is not always the case, perhaps varying with the details of the task.

Slot Filling and Property Mapping

Experiments 5 and 6 of this thesis address the differences between two alternative ways of combining concepts in a compound: slot filling and property mapping. Until quite recently only slot filling interpretations have been seriously looked at by researchers, (e.g. Levi, 1977; Shoben, 1991). Wisniewski and Gentner (1991) introduced the term 'property mapping', and made some comparison between the two kinds of process. They asked subjects to paraphrase novel noun noun compounds, and found that although slot filling seemed to be the most common interpretation type, property mapping interpretations were not uncommon. Experiment 5 replicates this finding. Of interpretations in which noun1 modified noun2, 67% were slot filling and 26% property mapping. It was further found that although slot filling was preferred to property mapping for AA, AN and NA compounds, this balance was reversed for NN compounds. Counts of property mapping were included in experiments 5, 6 and 8. In each case NN compounds had markedly more property mapping interpretations than other compound types. Much less marked, but also consistent in these three experiments, AN compounds showed more property mapping than AA compounds, which in turn showed more than NA compounds. However, these differences were small, so only the high frequency of property mapping among NN compounds can be confidently expected to generalise.[14]

In the real compound condition of experiment 5 NN compounds again showed more property mapping than other compound types, but the difference was more pronounced. The real / novel comparison is consistent with an analogical model

---

[14] This ordering of property mapping frequency (NN>AN>AA>NA) is the same as the order of frequency of feature matches between noun2 and the compound in experiment 2. It is plausible that the two findings are related, since property mapping alters the meaning of the head noun only on a single feature, leaving the noun2 and compound with all other features in common.

of compound interpretation, in that the pattern for novel compounds is rather like a noisy echo of that for real compounds. However, the same data could be explained in other terms. The process of lexicalisation of novel compounds could itself account for such a pattern developing. If there is a some underlying semantic influence which biases novel NN compounds to be interpreted by property mapping, and other novel compounds to be interpreted by slot filling, then compounds which do not fit with the general pattern will probably be slightly less good at conveying their meaning, and so these are less likely to be adopted into common use. This would lead to the kind of polarisation of compound compositions seen in the data for real compounds.

Wisniewski (1996) found that compounds with similar constituents were more likely to have property mapping interpretations than those with less similar constituents. He interprets this as implying that property mapping interpretations are arrived at by "structural alignment", in which concepts are compared for any salient differences which are then resolved by mapping the modifier's value of the contrasting property across to the head noun. However, a post-test of similarity for the materials in experiment 5 found that the NN constituents were no more similar to each other than were the AA constituents, and yet NN compounds had many more property mapping interpretations. Neither was there any correlation between property mapping interpretations and constituent similarity. The increase in property mapping in NN compounds was therefore interpreted as being due to two factors. Firstly, because the constituent were of the same type, properties of the modifier were naturally more likely to be appropriate to the head than would be the case for heterogeneous compounds (AN and NA). Secondly, natural kinds seem to be characterised by having several salient features, whereas an artefact's most salient feature is usually its function. This means that artefacts will often have clashing (i.e. mutually contradictory) salient properties, and so mapping the salient property of one to another will be inhibited. Thus NN compounds would be more likely to have a property mapping interpretation than other compounds.

Experiment 6 looked at possible reasons for the variation in interpretation types across Compound Compositions. Three factors were included in this experiment; Compound Composition (AA, AN, NA or NN), Constituent Similarity (High or Low), and Feature Clash (Clash or Nonclash, referring to whether or not the constituents defining properties conflicted with each other).

Compounds composed of highly similar constituents were found to be more likely to be interpreted by property mapping than those composed of dissimilar constituents. But this effect was confounded with the significant effect of Compound Composition, for which NN compounds again showed more property mapping than other compounds. High levels of similarity seem to be more frequent in pairs of natural kinds than in other concept combinations, and this factor proved difficult to isolate. However, since the main effect of Constituent Similarity was not limited to NN compounds, the results do support the theory of structural alignment.

Because Constituent similarity tended to vary with Compound Composition it proved difficult to obtain many materials in some conditions, so interactions that were significant on F1 were nonsignificant on the F2 analysis. However, the data strongly suggest three interactions; Constituent Similarity x Feature Clash, Feature Clash x Compound Composition and Constituent Similarity x Compound Composition. The Constituent Similarity x Feature Clash interaction is due to similarity having a much reduced influence in the clash condition. This suggests that a clash of salient properties may inhibit property mapping among compounds with similar constituents.

The remaining interactions are difficult to interpret due to the confounding of similarity and compound composition. However, they do suggest that Constituent Similarity is only one of several influential factors in determining that a compound should have a property mapping interpretation.

Property mapping interpretations of noun noun compounds have not been the subject of much attention in the literature thus far, but are worth finding out more about. They are in a unique position, having the potential to offer insights into several other phenomena such as slot filling compounds, adjective noun compounds and metaphor.

Relative importance of noun1 and noun2

Various authors have investigated the possibility that one noun position in a compound or the other may be of primary importance in deciding the type of modification used in interpreting the compound (e.g. Van Jaarsveld et al (1994), Gagné and Shoben (1995)). As already mentioned, the results of experiment 5 suggest that noun1 may be particularly important in concept identification. In experiment 8 subjects were asked to paraphrase **original** compounds (e.g. worm bag), **same noun1** compounds (e.g. worm razor) and **same noun2** compounds (bee bag). These paraphrases were then compared to see which pair were most likely to have the same modification type. They were also compared to cases where there were no constituents in common to assess whether the presence of a common noun increased the modification type matches above chance level.

It was found that **Noun1 in Common** and **Noun2 in Common** comparisons were equal, both being more similar than the control comparison with no common constituents. It was therefore concluded that both nouns do have an influence in determining the modification type, but that neither position was more influential than the other. This could be seen as supportive of a theory of compound interpretation by analogy, such as that of Ryder (1994). However, semantic information could also bring about such preferences.

There was also an interaction between Compound Composition and Noun in Common in experiment 8. The Noun1 in Common comparison showed a marked increase in matches for the NN compounds. This was attributed to the high frequency of property mapping among these compounds. This suggests that noun1 may have a primary role in determining that a compound has a property mapping interpretation. Other studies have involved such complicating factors as influential contexts (Shoben, 1993), prior exposure to analogues (Shoben, 1993; Van Jaarsveld et al, 1994), or time pressure (Gagné and Shoben, 1995). The results from experiment 8 can best be seen as an indication that as a baseline both positions are equally important. Noun1 may be more important in property mapping compounds, but also other factors may be significant in different circumstances.

Findings across experiments

Two themes in particular run throughout these experiments; a comparison of real and novel compounds, and a comparison of AA, AN, NA and NN compound compositions. Some overall conclusions about these factors are drawn here.

Real / Novel Comparison

In experiments 1 - 4 there was no overall difference between real and novel compounds. Both were found to exhibit similar degrees of noncompositionality.

However, there was a difference between real and novel compounds with respect to how much each of the constituent concepts has in common with the compound. In each case noun2 is more similar to the compound than noun1, but the difference is greater for real compounds than for novel ones. In experiment 7 real compounds were found to be considerably easier to identify correctly from a feature profile than were novel compounds. In experiments 5, 6 and 8 it was found that property mapping was more common in NN compounds than in other compound compositions. In experiment 5 this difference was more marked for real compounds than for novel ones. No real compounds were included in experiments 6 and 8.

With respect to the issue of compositionality, it proved very difficult to measure quantitatively the extra noncompositionality associated with real compounds that would be expected due to extensional feedback. There are several possible explanations, but the matter is perhaps worth pursuing in light of the possible conclusion that, ultimately, it may prove difficult to propose an account of compounding which is compositional.

The finding from experiments 1 - 4, that the pattern of feature inheritance is similar for real and novel compounds, but more marked in the real cases, is similar to the finding in experiment 5 with respect to the pattern of property mapping and slot filling among real and novel compounds. In both cases, the differences in the real compounds are also found in novel compounds, only less marked. This suggests some interesting relationship between real and novel compounds, but could be interpreted in two different ways. Firstly, such similarity could be due to the fact that real compounds were themselves once novel. Perhaps one factor which increased the likelihood of their being adopted into the lexicon was that they conformed to the standard. The alternative possibility is to view the process the other way around, to see the interpretation of novel compounds as influenced by existing real compounds, in a manner something like Ryder's analogy model (1994). These data do not provide a way of deciding between these two explanations.

Experiment 7 found that real compounds were more likely to be correctly identified than novel ones. Three possible causes of this were considered. First, that real compounds are easier to identify since subjects are simply more familiar with them - they are more salient, and we know more about them. Second, that the feature profiles that subjects were presented with were more coherent for real compounds due to the way they were generated. Finally, that real compounds have a unique linguistic label associated with them, so it is easier for subjects to be precisely correct in their identification. This latter influence was shown to have some small effect in the increase in correct identifications when semantically close responses were included.

Compound Composition

The ontological category of compound constituents looks like it ought to be important. Although usually ignored, other researchers have occasionally included this factor in experimental designs, but either not reported results for it (Wisniewski and Gentner, 1991) or failed to comment on apparent differences (Wisniewski, 1996).

Results in experiments 1 to 4 here suggest that there are implications of compound composition for the modification process. In each of these experiments, the novel AN compounds were found to be anomalous with respect to the amount of meaning each constituent had in common with the compound. In experiment 1 novel AN compounds had higher noncompositionality than the other compounds tested, particularly with respect to noun2. In experiment2 it was found that the noun2s of AN and NN compounds more features in common with the compounds than other conditions. In experiments 3 and 4 noncompositionality with respect to noun1 was higher for novel AN compounds than the others.

Although the differences from other compound compositions were not the same in each case, it does appear that AN compounds may be a special case with respect to compositionality. Variations in their behaviour between experiments are attributed to differences in the task, and in experimental design.

In the identification of simple concepts in experiment 7 no difference was found between artefacts and natural kinds. There was however some difference between Compound Compositions in the real condition. More AA and AN compounds were correctly identified, and fewer features were needed to do so than for NA and NN compounds. In experiment 5, property mapping was found to be more frequent in NN compounds than in others. This finding was replicated in experiments 6 and 8. Although the differences within each individual experiment were small, the rank order of Compound Compositions in terms of property mapping frequency was the same in each of these experiments: NN>AN>AA>NA. Results from experiment 6 suggest that no single factor can account for property mapping, since Constituent Similarity, Feature Clashes and Compound Composition all interact. Finally, the results of experiment 8 suggest that noun1 may play a primary role in determining property mapping among NN compounds in particular.

Looking at Compound Composition from these several different points of view, it looks like it is a factor of some importance. Differences between Compound

Compositions seem more dependent upon the details of the experiment in some cases (e.g. with respect to compositionality) than others (e.g. property mapping / slot filling). Also the relationships between these findings are hard to determine. For example, whether a compound is interpreted by property mapping or slot filling ought to have an effect on its compositionality, but no clear pattern emerges from these results. Other factors can also vary along with Compound Composition, and it has proved difficult to control for them, such as Constituent Similarity in experiment 6. Overall, however the results from these experiments suggest that the ontological category of a compound's constituents can have a significant influence over several aspects of the compound.

Summary and Conclusions

Overall, these experiments have resulted in a number of key findings. Real and novel compounds show noncompositionality, both to a similar extent, and it has been demonstrated that noun2 tends to be the head in such combinations. Also, people find it difficult to identify compounds from a feature profile of the concept, particularly in the case of novel compounds. The most common type of modification is slot filling, but property mapping does occur in a large number of cases, and is actually more common than slot filling for NN compounds. This difference is more marked for real compounds than for novel ones. Compounds are also more likely to be interpreted by property mapping if the constituents are similar than if they are quite different, but if the salient features of the constituents clash, the effect of similarity diminishes. It was also found that neither constituent position is necessarily dominant in terms of predicting the preferred type of modification used in interpreting compounds, but the details of specific tasks probably have an influence over this. Also, noun1 seems to be of particular importance in the interpretation of NN compounds, perhaps due to the high frequency of property mapping among NN compounds.

These results illustrate the importance of considering differences between compounds (whether they are real or novel, and what semantic categories they are composed of), and differences between experimental tasks subjects are asked to undertake.

These various results can be related to existing theories of concepts and conceptual combination. Experiments 1 - 4 looked at noncompositionality of compounds by combining compound constituents in a manner based on that used by Hampton (1987). The predicted differences in compositionality were not found. This is particularly troublesome for the comparison between real and novel compounds, in which the effect of extensional feedback ought to be evident. In this respect the hypothesis fails. However, Hampton's model was proposed for explicitly conjunctive combinations and so may not apply to compounds.

Experiment 7, showed that people seem to find it difficult to identify compound concepts from feature lists, particularly novel compounds. Perhaps such a representation (e.g. Smith et al, 1984; Wisniewski and Gentner, 1991) is simply not rich enough. A more elaborate representation which involved more causal relationships, and reference to real world knowledge might lead to better compound identification (e.g. Murphy 1990). Structural alignment (Wisniewski, 1996) could be implemented in any of these theories, and experiments 6 and 7 taken together provide support for it as a mechanism for facilitating property mapping. However, the results also suggest that property mapping can occur in compounds with dissimilar constituents in which structural alignment is unlikely. Finally, several results are consistent with a theory of compound interpretation by analogy (Ryder, 1992). In experiment 6, it was found that the pattern of results for novel compounds followed that of real compounds, but was less marked. Such a finding might well be accounted for by analogical interpretation. In experiment 8, it was found that compounds which share a common constituent (whether it is noun1 or noun2) are more likely than other similar compounds to have the same sort of interpretation. Again, this is as would be predicted by theories of interpretation by analogy. Such results however might also be accounted for by reference to purely semantic information (van Jaarsveld et al, 1994), or by thematic roles (e.g. Shoben, 1993).

Appendix

Table A1.1 : Materials used in Experiment1

real AA	novel AA	novel AN
bicycle clip	bicycle spoon	bicycle moth
bread bin	bread sack	bread mushroom
cable television	cable umbrella	cable tree
carpet shampoo	carpet hook	carpet leopard
coat hook	coat shampoo	coat frog
desk chair	desk box	desk dog
drawing pin	drawing boot	drawing leaf
football boot	football curtain	football monkey
goal post	goal gun	goal sea
library shelf	library pin	library stag
makeup bag	makeup television	makeup snake
match box	match chair	match rose
oven glove	oven slate	oven eagle
paint brush	paint handle	paint grass
pen lid	pen polish	pen lion
phone card	phone belt	phone cat
refuse sack	refuse shelf	refuse peacock
roof slate	roof glove	roof elephant
seat belt	seat card	seat bird
shoe polish	shoe lid	shoe bull
shower curtain	shower bin	shower apple
soup spoon	soup clip	soup owl
staple gun	staple post	staple horse
umbrella handle	umbrella brush	umbrella animal

Table A1.2 : Mean Raw Scores, Experiment1

	subject group	noun1	noun2	compound
real AA	1	0.676	1.478	1.020
	2	0.555	1.432	0.636
novel AA	1	0.676	1.509	0.680
	2	0.555	1.345	0.451
novel AN	1	0.676	2.410	0.762
	2	0.555	2.254	0.685

Table A1.3 : Analysis of Variance of Raw Scores, Experiment1

Source	Sums of Squares	d.f.	Mean squares	F	p
Between Subjects	43.07	61	0.71		
Compound Type	9.13	2	4.56	6.46	< 0.003
Within Subjects : involving Subject Group	22.05	61	0.36		
Subject Group	1.69	1	1.69	4.69	< 0.034
Type x Group	0.00	2	0.00	0.00	ns
Within Subjects : involving Compound Part	124.13	122	1.02		
Compound Part	96.21	2	48.11	47.28	< 0.001
Type x Part	13.21	4	3.30	3.24	< 0.014
Within Subjects : involving Group x Part	28.01	122	0.23		
Group x Part	0.11	2	0.06	0.25	ns
Type x Group x Part	0.62	4	0.15	0.67	ns

Table A1.4 : Mean Signed Difference Scores, Experiment1

	subject group	diff1	diff2
real AA	1	0.316	0.362
	2	0.118	-0.120
novel AA	1	-0.227	-0.172
	2	-0.187	-0.287
novel AN	1	-0.290	-0.807
	2	-0.135	-0.720

Table A1.5 : Analysis of Variance of Mean Signed Difference Scores, Experiment1

Source	Sums of Squares	d.f.	Mean squares	F	p
Between Subjects	190.07	61	3.12		
Compound Type	17.09	2	8.54	2.74	ns
Within Subjects : inolving Subject Group	54.43	61	0.89		
Subject Group	0.65	1	0.65	0.73	ns
Type x Group	1.95	2	0.98	1.09	ns
Within Subjects : inolving Difference Part	91.19	61	1.49		
Difference Part	2.75	1	2.75	1.84	ns
Type x Difference Part	2.20	2	1.10	0.74	ns
Within Subjects : inolving Subject Group x Difference Part	29.76	61	0.49		
Group x Difference Part	0.71	1	0.71	1.46	ns
Type x Group x Difference Part	0.29	2	0.14	0.30	ns

Table A1.6 : Mean Absolute Difference Scores, Experiment1

	subject group	abs1	abs2
real AA	1	2.461	1.716
	2	2.162	1.651
novel AA	1	2.137	1.765
	2	2.038	1.801
novel AN	1	2.144	2.224
	2	2.013	2.158

Table A1.7 : Analysis of Variance of Mean Absolute Difference Scores, Experiment1

Source	Sums of Squares	d.f.	Mean squares	F	p
Between Subjects	58.23	61	0.95		
Compound Type	1.50	2	0.75	0.79	ns
Within Subjects : involving Subject Group	21.96	61	0.36		
Subject Group	0.34	1	0.34	0.93	ns
Type x Group	0.27	2	0.14	0.37	ns
Within Subjects : involving Absdiff Part	42.64	61	0.70		
Absdiff Part	3.69	1	3.69	5.28	< 0.025
Type x Absdiff Part	5.20	2	2.60	3.72	< 0.030
Within Subjects : involving Subject Group x Absdiff Part	22.31	61	0.37		
Group x Absdiff Part	0.42	1	0.42	1.15	ns
Type x Group x Absdiff Part	0.02	2	0.01	0.03	ns

Table A2.1 : Materials used in Experiments 2, 3, 4 and 7, with Word Frequencies from Kucera and Francis (1967)

	Real	Novel
AA (word frequency)	oven glove (7, 9)	apron filter (7, 9)
	car door (274, 312)	street church (255, 348)
	coat hanger (43, 0)	weapon clamp (42, 0)
	pen knife (18, 76)	candle bottle (18, 76)
	tire lever (22, 14)	pitch gum (22, 14)
	brick wall (18, 160)	drain floor (18, 158)
AN	dairy cow (19, 29)	airport corn (19, 34)
	toilet duck (13, 9)	tub moss (13, 9)
	house fly (591, 33)	school rice (492, 33)
	computer mouse (13, 10)	carpet slug (13, 10)
	clay pigeon (100, 3)	bridge beaver (98, 3)
	mug tree (1, 59)	ladle bear (1, 57)
NA	dog dish (75, 16)	animal jar (68, 16)
	bird cage (31, 9)	tea inn (28, 9)
	fish kettle (35, 3)	fruit mop (35, 3)
	butterfly net (2, 34)	crow pencil (2, 34)
	horse shoe (117, 14)	plant arrow (125, 14)
	daisy wheel (0, 56)	giraffe motor (0, 56)
NN	stag beetle (8, 0)	pig parsnip (8, 0)
	crab apple (0, 9)	whale pineapple (0, 9)
	cat fish (23, 35)	sheep chicken (23, 37)
	tiger moth (7, 1)	algae zebra (7, 1)
	sparrow hawk (0, 14)	lizard bull (0, 14)
	eagle owl (5, 2)	otter shrimp (5, 2)

Table A2.2 : Mean Exact Type Matches with Compound, Experiment2.

	noun1		noun2	
	real	novel	real	novel
AA	1.3333	0.6667	2.1667	0.8333
AN	0.1667	0.0000	3.1667	4.3333
NA	1.5000	1.1667	1.5000	0.8333
NN	2.0000	3.1667	5.1667	4.1667

Table A2.3 : Analysis of Variance of Mean Exact Type Matches with Compound, Experiment2.

Source	Sums of Squares	d.f.	Mean squares	F	p
Between Subjects	150.92	40	3.77		
Compound Composition	90.53	3	30.18	8.00	< 0.001
Extensional Status	1.26	1	1.26	0.33	ns
Composition x Extensional Status	7.78	3	2.59	0.69	ns
Within Subjects : involving Compound Part	179.58	40	4.49		
Compound Part	55.51	1	55.51	12.36	< 0.001
Composition x Part	52.86	3	17.62	3.92	< 0.015
Extensional Status x Part	1.26	1	1.26	0.28	ns
Extensional Status x Composition x Part	9.28	3	3.09	0.69	ns

Table A2.4 : Mean Exact Token Matches, Experiment2.

	noun1		noun2	
	real	novel	real	novel
AA	1.6667	1.3333	2.1667	1.0000
AN	0.1667	0.0000	3.5000	7.0000
NA	1.5000	2.0000	1.8333	0.8333
NN	6.1667	5.3333	7.6667	6.1667

Table A2.5 : Analysis of Variance of Mean Exact Token Matches with Compound, Experiment2.

Source	Sums of Squares	d.f.	Mean squares	F	p
Between Subjects	352.00	40	8.80		
Compound Composition	262.25	3	87.42	9.93	< 0.001
Extensional Status	1.50	1	1.50	0.17	ns
Composition x Extensional Status	19.58	3	6.53	0.74	ns
Within Subjects : involving Compound Part	483.33	40	12.08		
Compound Part	84.37	1	84.37	6.98	< 0.012
Composition x Part	119.54	3	39.85	3.30	< 0.03
Extensional Status x Part	2.04	1	2.04	0.17	ns
Extensional Status x Composition x Part	42.71	3	14.24	1.18	ns

Table A2.6 : Mean Liberal Type Matches, Experiment2.

	noun1		noun2	
	real	novel	real	novel
AA	2.3333	1.3333	5.0000	1.5000
AN	1.0000	0.0000	5.0000	6.6667
NA	1.8333	2.5000	3.3333	3.0000
NN	2.8333	4.8333	6.8333	5.8333

Table A2.7 : Analysis of Variance of Mean Liberal Type Matches with Compound, Experiment2

Source	Sums of Squares	d.f.	Mean squares	F	p
Between Subjects	226.25	40	5.66		
Compound Composition	99.78	3	33.26	5.88	< 0.002
Extensional Status	2.34	1	2.34	0.41	ns
Composition x Extensional Status	30.36	3	10.12	1.79	ns
Within Subjects : involving Compound Part	234.25	40	5.86		
Compound Part	157.59	1	157.59	26.91	< 0.001
Composition x Part	68.61	3	22.87	3.91	< 0.015
Extensional Status x Part	5.51	1	5.51	0.94	ns
Extensional Status x Composition x Part	29.53	3	9.84	1.68	ns

Table A2.8 : Mean Liberal Token Matches, Experiment2.

	noun1		noun2	
	real	novel	real	novel
AA	2.8333	2.3333	7.1667	2.3333
AN	1.1667	0.0000	6.5000	11.000
NA	2.0000	4.1667	6.1667	4.5000
NN	6.1667	10.000	12.8333	9.6667

Table A2.9 : Analysis of Variance of Mean Liberal Token Matches with Compound, Experiment2.

Source	Sums of Squares	d.f.	Mean squares	F	p
Between Subjects	6.84.08	40	17.10		
Compound Composition	553.78	3	184.59	10.79	< 0.001
Extensional Status	0.26	1	0.26	0.02	ns
Composition x Extensional Status	60.11	3	20.04	1.17	ns
Within Subjects : involving Compound Part	900.75	40	22.52		
Compound Part	372.09	1	372.09	16.52	< 0.001
Composition x Part	146.78	3	48.93	2.17	ns
Extensional Status x Part	33.84	1	33.84	1.50	ns
Extensional Status x Composition x Part	138.03	3	46.01	2.04	ns

Table A3.1 : Mean Raw Scores, Experiment3

Real									
		n1			n2			c	
	fs1	fs2	fs3	fs1	fs2	fs3	fs1	fs2	fs3
AA	5.125	2.810	2.949	2.544	5.121	3.788	2.585	4.881	4.814
AN	5.275	1.013	2.367	1.621	5.383	3.688	2.371	3.067	5.054
NA	5.738	1.413	1.875	1.329	4.886	3.354	2.400	3.958	4.400
NN	5.700	2.071	2.838	2.963	5.421	4.104	3.025	5.521	4.521
	5.459	1.826	2.507	2.114	5.203	3.733	2.595	4.357	4.697
		9.792			11.05			58.246	
Novel									
AA	4.775	2.638	2.713	2.495	5.058	3.553	3.293	4.378	4.133
AN	5.216	1.058	1.274	0.979	5.195	4.188	1.638	4.742	4.167
NA	5.388	1.458	3.001	1.196	4.821	2.804	3.425	4.121	3.733
NN	5.321	2.117	3.633	1.813	5.650	3.354	3.421	4.254	4.113
	5.175	1.818	2.655	1.621	5.181	3.475	2.944	4.374	4.037
		9.648			10.277			11.355	

Table A3.2 : Analysis of Variance of Mean Raw Scores by Subjects, Experiment3

Source	Sums of Squares	d.f.	Mean squares	F1	p
Between Subjects	93.32	43	2.17		
Compound Composition	55.83	3	18.61	8.58	< 0.001
Within Subjects : involving Extensional Status	30.63	43	0.71		
Extensional Status	3.57	1	3.57	5.01	< 0.030
Composition x Extensional Status	3.58	3	1.19	1.67	ns
Within Subjects : involving Compound Part	65.12	86	0.76		
Compound Part	52.11	2	26.06	34.41	< 0.001
Composition x Part	15.74	6	2.62	3.47	< 0.004
Within Subjects : involving Feature Set	42.45	86	0.49		
Feature Set	33.08	2	16.54	33.51	< 0.001
Composition x Feature Set	11.43	6	1.90	3.86	< 0.002
Within Subjects : involving Extensional Status x Part	55.03	86	0.64		
Extensional Status x Part	1.55	2	0.78	1.21	ns
Composition x Extensional Status x Part	9.02	6	1.50	2.35	< 0.038
Within Subjects : involving Extensional Status x Feature Set	48.53	86	0.56		
Extensional Status x Feature Set	2.81	2	1.41	2.49	ns
Composition x Extensional Status x Feature Set	11.90	6	1.98	3.51	< 0.004
Within Subjects : involving Part x Feature Set	132.80	172	0.77		
Part x Featrue Set	1294.91	4	323.73	419.27	< 0.001
Composition x Part x Featrue Set	60.94	12	5.08	6.58	< 0.001
Within Subjects : involving Extensional Status x Part x Feature Set	74.88	172	0.44		
Extensional Status x Part x Feature Set	14.80	4	3.70	8.50	< 0.001
Composition x Extensional Status x Part x Feature Set	42.33	12	3.53	8.10	< 0.001

Table A3.3 : Analysis of Variance of Mean Raw Scores by Materials, Experiment3

Source	Sums of Squares	d.f.	Mean squares	F2	p
Between Subjects	44.40	40	1.11		
Extensional Status	1.96	1	1.96	1.77	ns
Compound Composition	33.95	3	11.32	10.19	< 0.001
Extensional Status x Compound Composition	1.68	3	0.56	0.51	ns
Within Subjects : involving Compound Part	34.00	80	0.43		
Compound Part	25.41	2	12.71	29.89	< 0.001
Extensional Status x Part	0.86	2	0.43	1.02	ns
Composition x Part	7.84	6	1.31	3.07	< 0.009
Extensional Status x Composition x Part	3.89	6	0.65	1.53	ns
Within Subjects : involving Feature Set	64.70	80	0.81		
Feature Set	16.38	2	8.19	10.13	< 0.001
Extensional Status x Feature Set	1.15	2	0.58	0.71	ns
Composition x Feature Set	6.53	6	1.09	1.35	ns
Extensional Status x Composition x Feature Set	6.48	6	1.08	1.35	ns
Within Subjects : involving Compound Part x Feature Set	104.02	160	0.65		
Compound Part x Feature Set	655.05	4	163.76	251.90	< 0.001
Extensional Status x Part x Feature Set	7.69	4	1.92	2.96	< 0.022
Composition x Part x Feature Set	31.96	12	2.66	4.10	< 0.001
Extensional Status x Composition x Part x Feature Set	22.47	12	1.87	2.88	< 0.001

Table A3.4 : Mean Signed Difference Scores (compound - constituents), Experiment3

Real						
		diff1			diff2	
	fs1	fs2	fs3	fs1	fs2	fs3
AA	-0.366	2.071	1.865	0.040	-0.240	1.026
AN	-0.829	2.054	2.688	0.750	-2.317	1.367
NA	-0.604	2.546	2.525	1.071	-0.928	1.046
NN	-1.683	3.450	1.683	0.063	0.100	0.417
	-0.871	2.530	2.190	0.481	-0.846	0.964
		3.849			0.599	
Novel						
AA	-0.143	1.740	1.420	0.799	-0.681	0.611
AN	-0.913	3.683	2.893	0.658	-0.453	0.363
NA	-0.896	2.663	0.732	2.229	-0.700	0.763
NN	-0.608	2.138	0.479	1.608	-1.396	0.758
	-0.639	2.556	1.381	1.324	-0.807	0.624
		3.298			1.141	

Table A3.5 : Analysis of Variance of Mean Signed Difference Scores by Subjects, Experiment3

Source	Sums of Squares	d.f.	Mean squares	F1	p
Between Subjects	125.25	43	2.91		
Compound Composition	3.68	3	1.23	0.42	ns
Within Subjects : involving Extensional Status	112.11	43	2.61		
Extensional Status	0.34	1	0.34	0.13	ns
Composition x Extensional Status	9.99	3	3.33	1.28	ns
Within Subjects : involving Difference Part	23.36	43	0.54		
Difference Part	13.77	1	13.77	25.35	< 0.001
Composition x Part	14.52	3	4.84	8.91	< 0.001
Within Subjects : involving Feature Set	187.86	86	2.18		
Feature Set	459.36	2	229.68	105.15	< 0.001
Composition x Feature Set	29.43	6	4.90	2.25	< 0.046
Within Subjects : involving Extensional Status x Part	17.66	43	0.41		
Extensional Status x Part	1.44	1	1.44	3.50	ns
Composition x Extensional Status x Part	5.69	3	1.90	4.62	< 0.007
Within Subjects : involving Extensional Status x Feature Set	165.16	86	1.92		
Extensional Status x Feature Set	41.28	2	20.64	10.75	< 0.001
Composition x Extensional Status x Feature Set	71.91	6	11.98	6.24	< 0.001
Within Subjects : involving Part x Feature Set	70.19	86	0.82		
Part x Featrue Set	1141.79	2	570.90	699.53	< 0.001
Composition x Part x Featrue Set	51.13	6	8.52	10.44	< 0.001
Within Subjects : involving Extensional Status x Part x Feature Set	19.82	86	0.23		
Extensional Status x Part x Feature Set	1.04	2	0.52	2.25	ns
Composition x Extensional Status x Part x Feature Set	18.36	6	3.06	13.28	< 0.001

Table A3.6 : Analysis of Variance of Mean Signed Difference Scores by Materials, Experiment3

Source	Sums of Squares	d.f.	Mean squares	F2	p
Between Subjects	101.86	40	2.55		
Extensional Status	0.00	1	0.00	0.00	ns
Compound Composition	3.81	3	1.27	0.50	ns
Extensional Status x Compound Composition	4.31	3	1.44	0.56	ns
Within Subjects : involving Difference Part	95.58	40	2.39		
Difference Part	58.52	1	58.52	24.49	< 0.001
Extensional Status x Part	2.39	1	2.39	1.00	ns
Composition x Part	10.26	3	3.42	1.43	ns
Extensional Status x Composition x Part	4.59	3	1.53	0.64	ns
Within Subjects : involving Feature Set	230.04	80	2.88		
Feature Set	72.97	2	36.49	12.69	< 0.001
Extensional Status x Feature Set	14.87	2	7.44	2.59	ns
Composition x Feature Set	15.35	6	2.56	0.89	ns
Extensional Status x Composition x Feature Set	34.60	6	5.77	2.01	ns
Within Subjects : involving Compound Part x Feature Set	147.31	80	1.84		
Compound Part x Feature Set	303.53	2	151.77	82.42	< 0.001
Extensional Status x Part x Feature Set	1.17	2	0.59	0.32	ns
Composition x Part x Feature Set	17.63	6	2.94	1.60	ns
Extensional Status x Composition x Part x Feature Set	5.13	6	0.85	0.46	ns

Table A3.7 : Mean Absolute Difference Scores, Experiment3

Real						
		absdiff1			absdiff2	
	fs1	fs2	fs3	fs1	fs2	fs3
AA	2.774	2.221	2.049	0.762	0.999	1.249
AN	3.046	2.413	2.913	1.100	2.742	1.975
NA	3.471	2.779	2.833	1.471	1.461	1.363
NN	2.875	3.541	2.033	0.804	0.867	0.867
	3.041	2.739	2.457	1.034	1.517	1.363
		8.237			3.914	
Novel						
AA	1.663	1.993	1.754	1.290	1.097	1.167
AN	3.621	3.683	3.176	1.044	1.045	1.138
NA	2.138	2.679	1.688	2.121	1.158	1.421
NN	2.250	2.379	1.771	1.850	1.746	1.250
	2.418	2.684	2.097	1.576	1.262	1.244
		7.199			4.082	

Table A3.8 : Analysis of Variance of Mean Absolute Difference Scores by Subjects, Experiment3

Source	Sums of Squares	d.f.	Mean squares	F1	p
Between Subjects	64.44	43	1.50		
Compound Composition	55.13	3	18.38	12.26	< 0.001
Within Subjects : involving Extensional Status	30.46	43	0.71		
Extensional Status	6.93	1	6.93	9.79	< 0.003
Composition x Extensional Status	4.38	3	1.46	2.06	ns
Within Subjects : involving Absdiff Part	61.60	43	1.43		
Absdiff Part	210.95	1	210.95	147.25	< 0.001
Composition x Part	10.76	3	3.59	2.50	ns
Within Subjects : involving Feature Set	45.45	86	0.53		
Feature Set	32.39	2	16.20	30.65	< 0.001
Composition x Feature Set	22.38	6	3.73	7.06	< 0.001
Within Subjects : involving Extensional Status x Part	49.94	43	1.16		
Extensional Status x Part	4.64	1	4.64	3.99	ns
Composition x Extensional Status x Part	50.70	3	16.90	14.55	< 0.001
Within Subjects : involving Extensional Status x Feature Set	33.69	86	0.39		
Extensional Status x Feature Set	6.10	2	3.05	7.78	< 0.001
Composition x Extensional Status x Feature Set	3.78	6	0.63	1.61	ns
Within Subjects : involving Part x Feature Set	108.77	86	1.26		
Part x Featrue Set	9.49	2	4.74	3.75	< 0.027
Composition x Part x Featrue Set	8.63	6	1.44	1.14	ns
Within Subjects : involving Extensional Status x Part x Feature Set	109.59	86	1.27		
Extensional Status x Part x Feature Set	4.19	2	2.10	1.65	ns
Composition x Extensional Status x Part x Feature Set	20.07	6	3.35	2.63	< 0.022

Table A3.9 : Analysis of Variance of Mean Absolute Difference Scores by Materials, Experiment3

Source	Sums of Squares	d.f.	Mean squares	F2	p
Between Subjects	32.00	40	0.80		
Extensional Status	1.52	1	1.52	1.90	ns
Compound Composition	21.08	3	7.03	8.78	< 0.001
Extensional Status x Compound Composition	1.59	3	0.53	0.66	ns
Within Subjects : involving Absdiff Part	96.84	40	2.42		
Absdiff Part	110.68	1	110.68	45.72	< 0.001
Extensional Status x Part	2.91	1	2.91	1.20	ns
Composition x Part	4.37	3	1.46	0.60	ns
Extensional Status x Composition x Part	24.46	3	8.15	3.37	< 0.028
Within Subjects : involving Feature Set	29.65	80	0.37		
Feature Set	3.85	2	1.92	5.19	< 0.008
Extensional Status x Feature Set	0.48	2	0.24	0.64	ns
Composition x Feature Set	5.24	6	0.87	2.36	< 0.038
Extensional Status x Composition x Feature Set	1.46	6	0.25	0.66	ns
Within Subjects : involving Compound Part x Feature Set	51.21	80	0.64		
Compound Part x Feature Set	2.68	2	1.34	2.09	ns
Extensional Status x Part x Feature Set	5.84	2	2.92	4.56	< 0.013
Composition x Part x Feature Set	4.94	6	0.82	1.29	ns
Extensional Status x Composition x Part x Feature Set	5.89	6	0.98	1.53	ns

Table A4.1 : Raw Mean Scores, Experiment4

Word Order 1									
		n1			n2			c	
	fs1	fs2	fs3	fs1	fs2	fs3	fs1	fs2	fs3
AA	4.775	2.638	2.713	2.495	5.058	3.553	3.293	4.378	4.133
AN	5.216	1.058	1.274	0.979	5.195	4.188	1.638	4.742	4.167
NA	5.388	1.458	3.001	1.196	4.821	2.804	3.425	4.121	3.733
NN	5.321	2.117	3.633	1.813	5.650	3.354	3.421	4.254	4.113
	5.175	1.818	2.655	1.621	5.181	3.475	2.944	4.374	4.037
		9.648			10.277			11.355	
Word Order 2									
AA	5.151	2.650	3.517	2.845	5.446	3.893	4.785	4.633	4.276
AN	5.583	1.182	2.549	1.520	5.671	3.638	3.342	4.754	4.133
NA	5.496	1.054	2.267	1.079	5.279	2.783	3.604	4.158	3.979
NN	5.725	2.129	3.063	2.499	5.833	3.561	4.113	5.067	4.240
	5.489	1.754	2.849	1.986	5.557	3.469	3.961	4.653	4.157
		10.092			11.012			12.771	

Table A4.2 : Analysis of Variance of Raw Mean Scores by Subjects, Experiment4

Source	Sums of Squares	d.f.	Mean squares	F1	p
Between Subjects	117.99	87	7.36		
Compound Composition	60.05	3	20.02	14.76	< 0.001
Word Order	20.43	1	20.43	15.06	< 0.001
Composition x Word Order	9.23	3	3.08	2.27	ns
Within Subjects : involving Compound Part	96.82	174	0.56		
Compound Part	77.74	2	38.87	69.85	< 0.001
Composition x Part	15.63	6	2.61	4.68	< 0.001
Word Order x Part	3.21	2	1.61	2.89	ns
Composition x Word Order x Part	6.46	6	1.08	1.94	ns
Within Subjects : involving Feature Set	101.85	174	0.59		
Feature Set	33.90	2	16.95	28.96	< 0.001
Composition x Feature Set	9.35	6	1.56	2.66	< 0.017
Word Order x Feature Set	9.07	2	4.53	7.74	< 0.001
Composition x Word Ordser x Feature Set	2.92	6	0.49	0.83	ns
Within Subjects : involving Part x Feature Set	168.63	348	0.48		
Part x Feature Set	1262.64	4	315.66	651.41	< 0.001
Composition x Part x Feature Set	84.12	12	7.01	14.47	< 0.001
Word Order x Part x Feature Set	5.52	4	1.38	2.85	< 0.024
Composition x Word Order x Part x Feature Set	18.75	12	1.56	3.22	< 0.001

Table A4.3 : Analysis of Variance of Raw Mean Scores by Materials, Experiment4

Source	Sums of Squares	d.f.	Mean squares	F2	p
Between Subjects	28.87	40	0.72		
Compound Composition	32.46	3	10.82	14.99	< 0.001
Word Order	8.98	1	8.98	12.45	< 0.001
Compound Composition x Word Order	4.09	3	1.36	1.89	ns
Within Subjects : involving Compound Part	28.51	80	0.36		
Compound Part	39.58	2	19.79	55.53	< 0.001
Composition x Part	8.12	6	1.35	3.80	< 0.002
Word Order x Part	2.00	2	1.00	2.80	ns
Composition x Word Order x Part	2.48	6	0.41	1.16	ns
Within Subjects : involving Feature Set	31.81	80	0.40		
Feature Set	16.30	2	8.15	20.50	< 0.001
Composition x Feature Set	4.77	6	0.79	2.00	ns
Word Order x Feature Set	4.30	2	2.15	5.40	< 0.006
Composition x Word Order x Feature Set	1.85	6	0.31	0.78	ns
Within Subjects : involving Compound Part x Feature Set	83.13	160	0.52		
Compound Part x Feature Set	639.41	4	159.85	307.69	< 0.001
Extensional Status x Part x Feature Set	41.52	12	3.46	6.66	< 0.001
Word Order x Part x Feature Set	3.22	4	0.80	1.55	ns
Composition x Word Order x Part x Feature Set	9.73	12	0.81	1.56	ns

Table A4.4 : Mean Signed Difference Scores, Experiment4

Word Order 1						
		diff1			diff2	
	fs1	fs2	fs3	fs1	fs2	fs3
AA	-0.143	1.740	1.420	0.799	-0.681	0.611
AN	-0.913	3.683	2.893	0.658	-0.453	0.363
NA	-0.896	2.663	0.732	2.229	-0.700	0.763
NN	-0.608	2.138	0.479	1.608	-1.396	0.758
	-0.639	2.556	1.381	1.324	-0.807	0.624
		3.298			1.141	
Word Order 2						
AA	-0.367	1.983	0.760	1.940	-0.813	0.383
AN	-2.242	3.572	1.585	1.822	-0.917	0.496
NA	-1.892	3.104	1.713	2.525	-1.121	1.196
NN	-1.613	2.938	1.178	1.614	-0.767	0.679
	-1.528	2.899	1.309	1.975	-0.905	0.689
		2.680			1.759	

Table A4.5 : Analysis of Variance of Signed Difference Scores by Subjects, Experiment4

Source	Sums of Squares	d.f.	Mean squares	F1	p
Between Subjects	205.55	87	2.36		
Compound Composition	9.96	3	3.32	1.41	ns
Word Order	9.10	1	9.10	3.85	ns
Composition x Word Order	3.78	3	1.26	0.53	ns
Within Subjects : involving Difference Part	28.31	87	0.33		
Difference Part	8.71	1	8.71	26.76	< 0.001
Composition x Part	12.31	3	4.10	12.62	< 0.001
Word Order x Part	0.18	1	0.18	0.55	ns
Composition x Word Order x Part	5.20	3	1.73	5.33	< 0.002
Within Subjects : involving Feature Set	317.69	174	1.83		
Feature Set	153.11	2	76.55	41.93	< 0.001
Composition x Feature Set	56.93	6	9.49	5.20	< 0.001
Word Order x Feature Set	10.53	2	5.27	2.88	ns
Composition x Word Ordser x Feature Set	24.46	6	4.08	2.23	< 0.042
Within Subjects : involving Part x Feature Set	62.74	174	0.36		
Part x Feature Set	1211.60	2	605.80	1680.2	< 0.001
Composition x Part x Feature Set	65.15	6	10.86	30.12	< 0.001
Word Order x Part x Feature Set	2.01	2	1.01	2.79	ns
Composition x Word Order x Part x Feature Set	10.60	6	1.77	4.90	< 0.001

Table A4.6 : Analysis of Variance of Signed Difference Scores by Materials, Experiment4

Source	Sums of Squares	d.f.	Mean squares	F2	p
Between Subjects	64.30	40	1.61		
Compound Composition	4.95	3	1.65	1.03	ns
Word Order	0.00	1	0.00	0.00	ns
Compound Composition x Word Order	2.66	3	0.89	0.55	ns
Within Subjects : involving Difference Part	46.49	40	1.16		
Difference Part	18.95	1	18.95	16.30	< 0.001
Composition x Part	10.05	3	3.35	2.88	< 0.048
Word Order x Part	3.06	1	3.06	2.64	ns
Composition x Word Order x Part	4.38	3	1.46	1.26	ns
Within Subjects : involving Feature Set	164.40	80	2.05		
Feature Set	30.26	2	15.13	7.36	< 0.001
Composition x Feature Set	19.42	6	3.24	1.57	ns
Word Order x Feature Set	0.70	2	0.35	0.17	ns
Composition x Word Order x Feature Set	10.57	6	1.76	0.86	ns
Within Subjects : involving Difference Part x Feature Set	81.52	80	1.02		
Difference Part x Feature Set	479.89	2	239.94	235.48	< 0.001
Extensional Status x Part x Feature Set	23.82	6	3.97	3.90	< 0.002
Word Order x Part x Feature Set	12.45	2	6.23	6.11	< 0.003
Composition x Word Order x Part x Feature Set	2.58	6	0.43	0.42	ns

Table A4.7 : Mean Absolute Difference Scores, Experiment4

Word Order 1						
		absdiff1			absdiff2	
	fs1	fs2	fs3	fs1	fs2	fs3
AA	1.663	1.993	1.754	1.290	1.097	1.167
AN	3.621	3.683	3.176	1.044	1.045	1.138
NA	2.138	2.679	1.688	2.121	1.158	1.421
NN	2.250	2.379	1.771	1.850	1.746	1.250
	2.418	2.684	2.097	1.576	1.262	1.244
		7.199			4.082	
Word Order 2						
AA	1.253	2.161	1.571	2.149	1.351	1.195
AN	2.311	3.686	2.685	2.006	1.272	1.238
NA	2.083	3.188	2.246	2.558	1.396	1.704
NN	2.029	3.096	2.186	2.067	1.367	1.565
	1.919	3.033	2.172	2.195	1.347	1.426
		7.124			4.968	

Table A4.8 : Analysis of Variance of Absolute Difference Scores by Subjects, Experiment4

Source	Sums of Squares	d.f.	Mean squares	F1	p
Between Subjects	67.53	87	0.78		
Compound Composition	50.74	3	16.91	21.79	< 0.001
Word Order	0.02	1	0.02	0.03	ns
Composition x Word Order	5.40	3	1.80	2.32	ns
Within Subjects : involving Absdiff Part	139.11	87	1.60		
Absdiff Part	104.59	1	104.59	65.41	< 0.001
Composition x Part	51.84	3	17.28	10.81	< 0.001
Word Order x Part	4.89	1	4.89	3.06	ns
Composition x Word Order x Part	16.85	3	5.62	3.51	< 0.019
Within Subjects : involving Feature Set	60.83	174	0.35		
Feature Set	71.22	2	35.61	101.85	< 0.001
Composition x Feature Set	5.28	6	0.88	2.52	< 0.023
Word Order x Feature Set	2.79	2	1.39	3.98	< 0.020
Composition x Word Ordser x Feature Set	5.57	6	0.93	2.65	< 0.017
Within Subjects ; involving Part x Feature Set	209.22	174	1.20		
Part x Feature Set	51.76	2	25.88	21.53	< 0.001
Composition x Part x Feature Set	6.96	6	1.16	0.96	ns
Word Order x Part x Feature Set	10.62	2	5.31	4.42	< 0.013
Composition x Word Order x Part x Feature Set	2.06	6	0.34	.29	ns

Table A4.9 : Analysis of Variance of Absolute Difference Scores by Materials, Experiment4

Source	Sums of Squares	d.f.	Mean squares	F2	p
Between Subjects	16.38	40	0.41		
Compound Composition	17.94	3	5.98	14.61	< 0.001
Word Order	1.31	1	1.31	3.21	ns
Compound Composition x Word Order	1.58	3	0.53	1.29	ns
Within Subjects : involving Absdiff Part	76.19	40	1.90		
Absdiff Part	55.63	1	55.63	29.21	< 0.001
Composition x Part	26.07	3	8.69	4.56	< 0.008
Word Order x Part	1.84	1	1.84	0.97	ns
Composition x Word Order x Part	4.44	3	1.48	0.78	ns
Within Subjects : involving Feature Set	22.10	80	0.28		
Feature Set	6.67	2	3.34	12.08	< 0.001
Composition x Feature Set	1.06	6	0.18	0.64	ns
Word Order x Feature Set	0.30	2	0.15	0.54	ns
Composition x Word Order x Feature Set	0.99	6	0.17	0.60	ns
Within Subjects : involving Difference Part x Feature Set	46.78	80	0.58		
Difference Part x Feature Set	19.62	2	9.81	16.78	< 0.001
Extensional Status x Part x Feature Set	2.11	6	0.35	0.60	ns
Word Order x Part x Feature Set	6.13	2	3.06	5.24	< 0.007
Composition x Word Order x Part x Feature Set	1.1	6	0.18	0.32	ns

Table A5.1 : Materials, Experiment5 and Part1 of Experiment6

real	novel (order1)	novel (order2)
oven glove	apron filter	filter apron
car door	street church	church street
coat hanger	weapon clamp	clamp weapon
pen knife	candle bottle	bottle candle
tyre lever	pitch gum	gum pitch
brick wall	drain floor	floor drain
dairy cow	airport corn	corn airport
toilet duck	tub moss	moss tub
house fly	school rice	rice school
computer mouse	carpet slug	slug carpet
clay pigeon	bridge beaver	beaver bridge
mug tree	ladle bear	bear ladle
dog dish	animal jar	jar animal
bird cage	tea inn	inn tea
fish kettle	fruit mop	mop fruit
butterfly net	crow pencil	pencil crow
horse shoe	plant arrow	arrow plant
daisy wheel	giraffe motor	motor giraffe
stag beetle	pig parsnip	parsnip pig
crab apple	whale pineapple	pineapple whale
cat fish	sheep chicken	chicken sheep
tiger moth	algae zebra	zebra algae
sparrow hawk	lizard bull	bull lizard
eagle owl	otter shrimp	shrimp otter

Table A6.1 : Similarity ratings for Concept Pairs, by Clash/Nonclash and Similarity, Experiment6

		clash			
		high similarity			
		Similarity ratings			
		AA			
street	church	5	1	4	4
glove	coat	6	1	5	5
door	lever	5	5	4	3
knife	arrow	5	1	5	3
tyre	clamp	4	4	3	4
		AN			
gum	algae	4	1	5	2
lever	horse	3	1	4	2
coat	horse	3	4	2	2
		NN			
moss	corn	5	2	5	4
lizard	bear	5	2	4	5
moss	daisy	5	2	5	5
lizard	sheep	4	4	4	5
lizard	bird	5	5	5	3
beetle	slug	6	7	5	6
moss	parsnip	5	7	4	5
plant	corn	5	2	7	6
fish	bear	5	7	4	5
algae	moss	6	7	6	6
plant	parsnip	6	3	7	6
bear	sheep	6	5	5	5
sheep	bird	4	7	4	5
sheep	chicken	5	4	5	3
corn	parsnip	6	3	5	5
daisy	parsnip	5	5	4	5
		low similarity			
		AA			
glove	door	2	1	4	1
knife	floor	2	1	2	2
pencil	church	2	1	3	2
glove	lever	3	1	3	1
glove	weapon	3	1	2	1
floor	pencil	3	1	3	1
door	apron	2	1	3	1
		AN			
weapon	horse	2	1	2	2
floor	pineapple	2	1	3	1
arrow	pineapple	2	2	2	1
apron	horse	3	1	2	1
door	horse	2	1	2	2

Table A6.1 continued...

kettle	horse	2	1	3	1
pencil	pineapple	2	1	3	1
	NN				
beetle	algae	4	2	3	4
beetle	moss	4	2	3	4
moss	crow	4	7	2	3
plant	crow	4	2	3	1
algae	corn	5	2	4	2
algae	crow	3	4	4	2
algae	parsnip	4	3	4	2
corn	crow	4	3	4	3
crow	parsnip	4	2	4	2
	No clash				
	High Similarity				
	AA				
weapon	arrow	6	1	7	2
dish	clay	5	2	4	6
kettle	dish	4	1	4	3
lever	clamp	5	1	4	4
lever	pencil	6	1	4	3
	AN				
arrow	fly	4	1	4	2
lever	bull	3	2	4	2
dish	animal	2	2	3	4
wheel	animal	3	2	4	2
weapon	bull	4	1	3	3
	NN				
dog	giraffe	5	6	5	4
dog	stag	5	6	5	5
dog	tiger	6	6	5	5
dog	zebra	5	4	5	6
giraffe	tiger	5	5	5	6
giraffe	zebra	6	7	5	5
stag	tiger	5	4	5	5
stag	zebra	6	3	5	5
	Low similarity				
	AA				
glove	tyre	2	1	3	1
glove	wall	2	1	3	1
church	tyre	3	1	2	1
church	gum	2	1	2	1
coat	pencil	2	1	3	1
lever	church	2	1	2	1
apron	street	2	1	3	1
apron	airport	2	1	2	1
apron	tyre	2	1	3	1
apron	clamp	2	1	3	1

Table A6.1 continued...

apron	wheel	2	1	3	1
kettle	street	2	1	3	1
church	kettle	3	1	2	1
AN					
street	duck	2	1	2	1
airport	duck	2	1	2	1
airport	bear	2	1	2	1
kettle	moth	2	1	2	1
tyre	cat	2	1	2	1
glove	bull	2	1	2	1
NN					
mouse	corn	4	4	3	5
mouse	rice	4	1	3	2
mouse	slug	4	3	4	2
giraffe	stag	5	3	5	4

Table A6.2 : Analysis of Variance of the Mean Number of Property Mapping Interpretations by Subjects, Experiment6

Source	Sums of Squares	d.f.	Mean squares	F1	p
Between Subjects	0.56	15	0.04		
Constant	27.02	1	27.02	726.8	< 0.001
Within Subjects : involving Clash	0.31	15	0.02		
Clash	0.02	1	0.02	1.17	ns
Within Subjects : involving Similarity	0.52	15	0.03		
Similarity	1.2	1	1.20	34.45	< 0.001
Within Subjects : involving Compound Composition	2.32	45	0.05		
Compound Composition	3.31	3	1.10	21.39	< 0.001
Within Subjects : involving Clash x Similarity	0.76	15	0.05		
Clash x Similarity	0.29	1	0.29	5.73	< 0.030
Within Subjects : involving Clash x Compound Composition	1.09	45	0.02		
Clash x Compound Composition	0.49	3	0.16	6.71	< 0.001
Within Subjects : involving Similarity x Compound Composition	0.98	45	0.02		
Similarity x Compound Composition	0.38	3	0.13	5.88	< 0.002
Within Subjects : involving Clash x Similarity x Compound Composition	1.05	45	0.02		
Clash x Similarity x Compound Composition	0.10	3	0.03	1.42	ns

Table A6.3 : Analysis of Variance of the Mean Number of Property Mapping Interpretations by Materials, Experiment6

Source	Sums of Squares	d.f.	Mean squares	F2	p
Main Effects	788.581	5	157.716	13.557	< 0.001
Clash	7.771	1	7.771	0.668	ns
Similarity	104.611	1	104.611	8.992	< 0.003
Compound Composition	516.068	3	172.023	14.787	< 0.001
2-Way Interactions	171.540	7	24.506	2.107	< 0.046
Clash x Similarity	34.569	1	34.569	2.972	ns
Clash x Compound Composition	63.165	3	21.055	1.810	ns
Similarity x Compound Composition	52.137	3	17.379	1.494	ns
3-Way Interactions	8.862	3	2.954	0.254	ns
Clash x Similarity x Compound Composition	8.862	3	2.954	0.254	ns

Table A7.1 : Mean Number of Features used to Correctly Identify Compounds, Experiment7

Real									
		n1			n2			comp	
	fs1	fs2	fs3	fs1	fs2	fs3	fs1	fs2	fs3
AA	4.25	0.15	0.76	0.78	3.50	1.23	0.09	2.53	2.44
AN	5.81	0.22	1.24	0.20	3.72	1.28	1.47	1.42	2.47
NA	3.25	0.17	0.19	0.12	2.78	1.43	2.47	1.57	5.50
NN	4.36	0.18	0.39	1.38	5.20	0.59	2.62	3.31	1.77
	4.42	0.18	0.65	0.59	3.74	1.69	1.69	4.09	3.10
sum		5.25			6.02			8.88	
Novel									
AA	4.17	0.24	0.48	0.24	4.38	1.43	1.79	3.21	4.33
AN	4.00	0.39	0.37	0.06	3.39	1.16	2.08	2.08	2.17
NA	5.14	0.00	1.28	0.14	4.48	0.81	3.40	3.40	1.90
NN	3.53	0.12	0.76	0.26	4.56	1.21	2.96	4.88	2.83
	4.21	0.19	0.72	0.18	4.20	1.15	2.64	3.47	2.79
sum		5.12			5.53			8.90	

Table A7.2 : Analysis of Variance of the Mean Number of Features used to Correctly Identify Compounds by Subjects, Experiment7

Source	Sums of Squares	d.f.	Mean squares	F1	p
Between Subjects	387.04	110	3.52		
Part	63.32	2	31.66	8.998	< 0.001
Compound Composition	14.58	3	4.86	1.381	ns
Extensional Status	1.70	1	1.70	0.484	ns
Part x Composition	48.19	6	8.03	2.283	< 0.041
Part x Extensional Status	5.58	2	2.79	0.792	ns
Composition x Extensional Status	10.89	3	3.63	1.032	ns
Part x Composition x Extensional Status	40.43	6	6.74	1.915	ns
Within Subjects : involving Feature Set	201.86	220	0.92		
Feature Set	47.11	2	23.55	25.67	< 0.001
Part x Feature Set	781.57	4	195.39	212.95	< 0.001
Composition x Feature Set	24.42	6	4.07	4.44	< 0.001
Extensional Status x Feature Set	3.99	2	2.00	2.17	ns
Part x Composition x Feature Set	25.56	12	2.13	2.32	< 0.008
Part x Extensional Status x Feature Set	8.40	4	2.10	2.29	ns
Composition x Extensional Status x Feature Set	15.77	6	2.63	2.87	< 0.010
Part x Composition x Extensional Status x Feature Set	28.08	12	2.34	2.55	< 0.004

Table A7.3 : Analysis of Variance of the Mean Number of Features used to Correctly Identify Compounds by Materials, Experiment7

Source	Sums of Squares	d.f.	Mean squares	F2	p
Between Subjects	26.16	29	0.90		
Compound Composition	17.65	3	5.88	6.52	< 0.002
Extensional Status	0.82	1	0.82	0.90	ns
Composition x Extensional Status	13.29	3	4.43	4.91	< 0.007
Within Subjects : involving Part	53.32	58	0.92		
Part	45.67	2	22.84	24.84	< 0.001
Composition x Part	16.47	6	2.75	2.99	< 0.013
Extensional Status x Part	2.86	2	1.43	1.55	ns
Composition x Extensional Status x Part	23.50	6	3.92	4.26	< 0.001
Within Subjects : involving Feature Set	71.51	58	1.23		
Feature Set	40.82	2	20.41	16.55	< 0.001
Composition x Feature Set	11.21	6	1.87	1.52	ns
Extensional Status x Feature Set	2.51	2	1.26	1.02	ns
Composition x Extensional Status x Feature Set	5.29	6	0.88	0.71	ns
Within Subjects : involving Part x Feature Set	105.43	116	0.91		
Part x Feature Set	576.11	4	144.03	158.47	< 0.001
Composition x Part x Feature Set	32.91	12	2.74	3.02	< 0.001
Extensional Status x Part x Feature Set	3.83	4	0.96	1.05	ns
Composition x Extensional Status x Part x Feature Set	13.46	12	1.12	1.23	ns

Table A7.4 : Mean Number of Features used to Identify All Compounds, Experiment7

Real	n1			n2			comp		
	fs1	fs2	fs3	fs1	fs2	fs3	fs1	fs2	fs3
AA	4.00	0.44	0.64	0.61	3.61	1.31	0.83	1.47	1.86
AN	5.89	0.25	1.31	0.28	3.25	1.72	1.44	1.50	2.42
NA	3.25	0.17	0.19	0.11	2.69	0.61	2.28	3.31	4.36
NN	4.22	0.22	0.53	1.31	4.78	1.83	2.06	4.31	1.64
	4.34	0.27	0.67	0.58	3.58	1.37	1.47	2.65	2.57
sum		5.28			5.53			6.69	
Novel									
AA	4.08	0.28	0.67	0.28	3.72	1.78	1.89	2.81	2.94
AN	3.97	0.39	0.36	0.56	3.25	1.22	1.89	3.11	2.36
NA	5.03	0.17	1.39	0.17	4.36	0.86	2.28	3.22	1.81
NN	3.31	0.11	0.94	0.22	4.56	1.14	2.44	2.72	1.64
	4.10	0.24	0.84	0.18	3.97	1.25	2.13	2.97	2.19
sum		5.18			5.40			7.29	

Table A7.5 : Analysis of Variance of the Mean Number of Features used to Identify All Compounds by Subjects, Experiment7

Source	Sums of Squares	d.f.	Mean squares	F1	p
Between Subjects	323.82	120	2.70		
Part	28.93	2	14.46	5.36	< 0.006
Compound Composition	5.48	3	1.83	0.68	ns
Extensional Status	0.18	1	0.18	0.07	ns
Part x Composition	34.81	6	5.80	2.15	ns
Part x Extensional Status	1.35	2	0.68	0.25	ns
Composition x Extensional Status	13.62	3	4.54	1.68	ns
Part x Composition x Extensional Status	40.16	6	6.69	2.48	< 0.027
Within Subjects : involving Feature Set	134.45	240	0.56		
Feature Set	51.98	2	25.99	46.39	< 0.001
Part x Feature Set	716.68	4	179.17	319.83	< 0.001
Composition x Feature Set	15.66	6	2.61	4.66	< 0.001
Extensional Status x Feature Set	2.06	2	1.03	1.84	ns
Part x Composition x Feature Set	18.46	12	1.54	2.75	< 0.002
Part x Extensional Status x Feature Set	9.55	4	2.39	4.26	< 0.002
Composition x Extensional Status x Feature Set	12.33	6	2.06	3.67	< 0.002
Part x Composition x Extensional Status x Feature Set	21.64	12	1.80	3.22	< 0.001

Table A7.6 : Analysis of Variance of the Mean Number of Features used to Identify All Compounds by Materials, Experiment7

Source	Sums of Squares	d.f.	Mean squares	F2	p
Between Subjects	19.68	40	0.49		
Compound Composition	5.48	3	1.83	3.71	< 0.019
Extensional Status	0.18	1	0.18	0.37	ns
Composition x Extensional Status	13.62	3	4.54	9.22	< 0.001
Within Subjects : involving Part	31.92	80	0.40		
Part	28.93	2	14.46	36.26	< 0.001
Composition x Part	34.81	6	5.80	14.54	< 0.001
Extensional Status x Part	1.35	2	0.68	1.69	ns
Composition x Extensional Status x Part	40.16	6	6.69	16.78	< 0.001
Within Subjects : involving Feature Set	78.68	80	0.98		
Feature Set	51.98	2	25.99	26.42	< 0.001
Composition x Feature Set	15.66	6	2.61	2.65	< 0.021
Extensional Status x Feature Set	2.06	2	1.03	1.05	ns
Composition x Extensional Status x Feature Set	12.33	6	2.06	2.09	ns
Within Subjects : involving Part x Feature Set	137.10	160	0.86		
Part x Feature Set	716.68	4	179.17	209.09	< 0.001
Composition x Part x Feature Set	18.46	12	1.54	1.80	ns
Extensional Status x Part x Feature Set	9.55	4	2.39	2.79	< 0.028
Composition x Extensional Status x Part x Feature Set	21.64	12	1.80	2.10	< 0.019

Table A8.1 : Materials used in Experiment8

Original		Common Noun1		Common Noun2	
		AA			
pen	anvil	pen	cabinet	chapel	anvil
window	television	window	blanket	paint	television
key	sign	key	plough	bungalow	sign
guitar	tent	guitar	road	pill	tent
plug	hotel	plug	plate	piston	hotel
radio	violin	radio	fork	zip	violin
bomb	trowel	bomb	statue	jacket	trowel
flag	bell	flag	railway	pan	bell
jug	shop	jug	rocket	cafe	shop
taxi	magazine	taxi	beam	hospital	magazine
		AN			
bullet	cactus	bullet	virus	banjo	cactus
coin	canary	coin	salmon	basket	canary
cup	dolphin	cup	daffodil	phone	dolphin
shed	eel	shed	toad	belt	eel
tractor	penguin	tractor	locust	bed	penguin
brake	monkey	brake	ape	rug	monkey
drum	antelope	drum	barnacle	prison	antelope
bus	rhubarb	bus	star	badge	rhubarb
igloo	forest	igloo	scorpion	camera	forest
corridor	caterpillar	corridor	lion	van	caterpillar
		NA			
baboon	whisk	baboon	piano	camel	whisk
lobster	school	lobster	bar	bush	school
ant	fort	ant	spade	ivy	fort
porcupine	map	porcupine	saw	kangaroo	map
fossil	anorak	fossil	box	armadillo	anorak
rat	pocket	rat	chain	otter	pocket
carp	uniform	carp	tunnel	python	uniform
cabbage	drill	cabbage	canoe	cliff	drill
fern	beer	fern	pillow	hippopotamus	beer
worm	bag	worm	razor	bee	bag
		NN			
fox	buffalo	fox	hare	oak	buffalo
deer	potato	deer	mole	mosquito	potato
fungus	puma	fungus	bird	rhinoceros	puma
badger	octopus	badger	grass	hamster	octopus
weed	shark	weed	ostrich	crocodile	shark
weasel	wolf	weasel	cloud	mushroom	wolf
cow	snail	cow	pike	turtle	snail
beach	leek	beach	cat	vulture	leek
parrot	shrew	parrot	spider	volcano	shrew
peacock	ocean	peacock	vine	leopard	ocean

Table A8.2 : Analysis of Variance of the Mean Number of Similar Interpretation Types by Subjects, Experiment8

Source	Sums of Squares	d.f.	Mean squares	F1	p
Between Subjects	70.44	23	3.06		
Constant	404.80	1	404.80	132.17	< 0.001
Within Subjects : involving Common Noun	55.25	46	1.20		
Common Noun	24.70	2	12.35	10.28	< 0.001
Within Subjects : involving Compound Composition	99.28	69	1.44		
Compound Composition	61.27	3	20.42	14.19	< 0.001
Within Subjects : involving Common Noun x Compound Composition	113.44	138	0.82		
Common Noun x Compound Composition	12.96	6	2.16	2.63	< 0.019

Table A8.3 : Analysis of Variance of the Mean Number of Similar Interpretation Types by Materials, Experiment8

Source	Sums of Squares	d.f.	Mean squares	F2	p
Between Subjects	2588.43	36	71.90		
Compound Composition	1977.89	3	659.3	9.17	< 0.001
Within Subjects : involving Common Noun	5772.87	72	80.18		
Common Noun	524.60	2	262.30	3.27	< 0.044
Common Noun x Compound Composition	294.53	6	49.09	0.61	ns

Table A8.4 : Analysis of Variance of the Mean Number of Similar Interpretation Types by Subjects, Excluding NN Compounds, Experiment8

Source	Sums of Squares	d.f.	Mean squares	F1	p
Between Subjects	35.40	23	1.54		
Constant	186.78	1	186.78	121.35	< 0.001
Within Subjects : involving Common Noun	48.03	46	1.04		
Common Noun	11.33	2	5.66	5.42	< 0.008
Within Subjects : involving Compound Composition	48.24	46	1.05		
Compound Composition	4.80	2	2.40	2.29	ns
Within Subjects : involving Common Noun x Compound Composition	53.10	92	0.58		
Common Noun x Compound Composition	1.30	4	0.33	0.56	ns

Table A8.5 : Analysis of Variance of the Mean Number of Similar Interpretation Types by Materials, Excluding NN Compounds, Experiment8

Source	Sums of Squares	d.f.	Mean squares	F2	p
Between Subjects	1800.30	27	66.68		
Compound Composition	132.82	2	66.41	1.00	ns
Within Subjects : involving Common Noun	2962.40	54	54.86		
Common Noun	280.16	2	140.08	2.55	0.087
Common Noun x Compound Composition	48.11	4	12.03	0.22	9.27

Bibliography

Bibliography

Anglin, J. (1976). Les Premiers termes de référence de l'enfant. In S. Ehrlich and E. Tulving (Eds.), La memoire sémantique. Paris: Bulletin de Psychologie.

Armstrong, S., E., Gleitman, L., R. and Gleitman, H. (1983).What some concepts might not be. Cognition, 13, 263-308.

Barsalou, L. W. (1992). Cognitive Psychology: An Overview for Cognitive Psychologists. Hillsdale, NJ: LEA.

Barton, M. E., and Komatsu, L. K. (1989). Defining features of natural kinds and artifacts. Journal of Psycholinguistic Research, 18, 5, 433-447

Bates, E. and MacWhinney, B. (1987). Competition, variation and language learning. In Brian MacWhinney (Ed.), Mechanisms of Language Acquisition, 157-194. Hillsdale NJ: LEA.

Battig, W. F. and 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 Monograph, 80(3, Pt.2).

Braisby, N., Franks, B., and Hampton, J. A. (1996). Essentialism, word use, and concepts. Cognition, 59, 247-274.

Cacciari, C., and Tabossi, P. (1988).The comprehension of idioms. Journal of Memory and Language, 27, 668-683.

Caramazza A., Hillis, A., Leek, E. C. and Miozzo, M. (1994). The organization of lexical knowledge in the brain: evidence from category- and modality-specific deficits. In Lawrence A. Hirschfeld and Susan A. Gelman, Mapping the Mind: Domain Specificity in Cognition and Culture.

Clark, R. (1970). Concerning the logic of predicate modifiers. Noûs, 4, 311-336.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 12, 335-359.

Cohen, B. and Murphy, G. L. (1984). Models of concepts. Cognitive Science, 8, 27-58.

Costello, F. and Keane, M. T. (1992). Conceptual Combination : A theoretical overview, The Irish Journal of Psychology, 13, 125 - 140.

Dimbleby, David. (1975). Richard Dimbleby : A biography. London : Hodder and Stoughton.

Downing, P. (1977). On the creation and use of English compound nouns. Language, 53, 4, 810-842.

Eysenck, M. W and Keane, M. T. (1990). Cognitive Psychology: A Student's Handbook, Hove and Hillsdale: Lawrence Erlbaum Associates.

Fodor, J., and Lepore, E., (1996). The red herring and the pet fish: why concepts still can't be prototypes. Cognition, 58, 253-370.

Gagné, C. L. and Shoben, E. J. (1995). The influence of thematic relations on the comprehension of nonpredicating combinations. Unpublished manuscript.

Gerrig, R. J. and Murphy, G. L. (1992). Contextual influences on the comprehension of complex concepts. Language and Cognitive Processes, 7(3/4), 205-230.

Gibbs, R. W. (1995). Idiomaticity and Human Cognition. In Martin Everaert, Erik-jan van der Linden, André Schenk and Rob Schreuder (Eds.), Idioms: Structural and Psychological Perspectives. LEA. Hillsdale, New Jersey; Hove UK

Gibbs, R. W. (1984). Literal meaning and psychological theory. Cognitive Science, 8, 275-304.

Glucksberg, S. and Keysar, B. (1990). Understanding metaphorical comparisons: beyond Similarity. Psychological Review, 97, 3-18.

Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. Memory and Cognition, 15(1), 55-71.

Hampton, J. A. (1988). Overextension of conjunctive concepts: evidence for a unitary model of concept typicality and class inclusion. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14(1), 12-32.

Hampton, J. A. (1991). The combination of prototype concepts. In P. J. Schwanenflugel (Ed.), The Psychology of Word Meanings. Hillsdale, NJ: LEA.

Hampton, J. A.(1995). Testing the prototype theory of concepts. Journal of Memory and Language, 34, 686-708.

Jackendoff, R. (1995). The boundaries of the lexicon. In Martin Everaert, Erik-jan van der Linden, André Schenk and Rob Schreuder (Eds.), Idioms: Structural and Psychological Perspectives. LEA. Hillsdale, New Jersey; Hove UK.

Keil, F. (1989). Concepts, kinds and cognitive development. Cambridge, MA: MIT Press.

Keil, F. (1987). Conceptual development and category structure. In U. Neisser, Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorisation. Cambridge, UK: Cambridge University Press.

Kripke, S. (1972a). Naming and necessity. In D. Davidson, and G. Harman (Eds.), Semantics of Natural Language. Dordrecht, Holland: Reidel.

Kucera, N. and Francis, W. N. (1967). A computational analysis of present-day American English. Providence, RI: Brown University Press.

Labov, W. (1972). Sociolinguistic Patterns. Philadelphia: University of Pennsylvania Press.

Lakoff, G. (1972). Hedges: a study of meaning criteria and the logic of fuzzy concepts. Papers from the eighth regional meeting, Chicago Linguistics Society. Chicago: University of Chicago Linguistics Department.

Levi, J. N. (1978). The Syntax and Semantics of Complex Nominals. New York: Academic.

Malt, B. C. (1990). Features and beliefs in the mental representation of categories. Journal of Memory and Language, 29, 289-315.

Malt, B. C. (1994). Water is not H_2O. Cognitive psychology, 27, 41-70.

Malt, B. C. and Johnson, E. C. (1992). Do artifact concepts have cores? Journal of Memory and Language, 31, 195-217,

Malt, B. C. and Smith, E. E. (1984). Correlated properties in natural categories. Journal of Verbal Learning and Verbal Behavior, 23, 250-269.

McCloskey, M. and Glucksberg, S. (1978). Natural Categories: well defined or fuzzy sets. Memory and Cognition, 6, 462-472.

Medin, D. L. and Ortony, A. (1989). Psychological essentialism. In Stella Vosniadou and Andrew Ortony, Similarity and Analogical Reasoning. Cambridge: Cambridge University Press.

Medin, D., L. and Shoben, E., J. (1988). Context and structure in conceptual combination. Cognitive Science, 20, 158-190.

Medin, D. L. and Smith, E. E. (1984). Concepts and concept formation. Annual Review of Psychology, 35, 113-138.

Miller, G. A. and Johnson-Laird, P. N. (1976). Language and Perception. Cambridge, Mass.: Harvard University Press.

Murphy, G. L. (1984). Establishing and accessing referents in discourse. Memory and Cognition, 12, 489-497, Murphy, G., L. (1988).Comprehending complex concepts. Cognitive Science, 12, 529-562.

Murphy, G., L. (1990). Noun phrase interpretation and conceptual combination. Journal of Memory and Language, 29, 259-288.

Osherson, D. N. and Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. Cognition, 9, 35-58.

Putnam, H. (1975). The meaning of "meaning". In H. Putnam, Mind, Language and Reality: Philosophical Papers, vol. 2. Cambridge: Cambridge University Press.

Rey, G. (1984). Concepts and stereotypes: a critical discussion of categories and concepts. Cognition,15, 1-3, 237-262.

Rips, L. J. (1989). Similarity, typicality and categorization. In Stella Vosniadou and Andrew Ortony, Similarity and Analogical Reasoning. Cambridge: Cambridge University Press.

Rips, L. J. (1995). The current status of research on concept combination. Mind and Language, 10(1/2), 72-104.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), Cognitive Development and the Acquisition of Language. New York: Academic Press.

Rosch, E. (1975a). Cognitive reference points. Cognitive Psychology, 7, 532-547.

Rosch, E. (1975b). Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 104, 192-233.

Rosch, E. (1977). Human Categorization. In N. Warren (Ed.), Advances in Cross-Cultural Psychology (Vol. 1). London: Academic Press.

Rosch, E. (1978). Principles of categorization. In Allan Collins and Edward E. Smith (Eds.), 1988. Readings In Cognitive Science: Perspectives from Psychology and Artificial Intelligence. San Mateo, California: Morgan Kaufman.

Rosch, E. and Mervis, C. B. (1975). Family resemblances: studies ion the internal structure of categories. Cognitive Psychology, 7, 573-605.

Rosch, E., Simpson, C. and Miller, R. S. (1976). Structural bases of typicality effects. Journal of Experimental Psychology: Human Perception and Performance, 2, 491-502.

Ryder, M. E. (1994). Ordered Chaos: the Interpretation of English Noun-Noun Compounds. University of California Press.

Schwartz, S. P. (1978). Putnam on artifacts. The Philosophical Review, LXXXVII, 4, 566-574.

Shoben, E. (1991). Predicating and nonpredicating combinations. In P. J. Schwanenflugel (Ed.), The Psychology of Word Meanings. Hillsdale, NJ: LEA.

Shoben, E. J. (1993). Non-predicating conceptual combinations. The Psychology of Learning and Motivation, 29, 391-409.

Smith, E. E. and Medin, D. L. (1981). Categories and Concepts. Cambridge, Mass.: Harvard University Press.

Smith, E. E., Osherson, D. N., Rips, L. J. and Keane, M. (1988). combining prototypes: a selective modification model. Cognitive Science, 12, 485-527.

Swinney, D. and Cutler, A. (1979). The access and processing of idiomatic expressions. Journal of Verbal Learning and Verbal Behavior, 18, 523-534.

Titone, D. A. and Connine, C. M. (1994). Comprehension of idiomatic expressions: effects of predictability and literality. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(5), 1126-1138.

Tversky, A. (1977). Features of Similarity. Psychological Review, 89, 123-154.

van Jaarsveld, H. J., Coolen, R. and Schreuder, R. (1994). The role of analogy in the interpretation of novel compounds. Journal of Psycholinguistic Research, 23(2), 111-137.

Wierzbicka, A. (1994). The structure of the lexicon, In Lawrence A. Hirschfeld and Susan A. Gelman, Mapping the Mind: Domain Specificity in Cognition and Culture.

Wisniewski, E. J. (1996). Construal and similarity in conceptual combination. Journal of Memory and Language, 35, 434-453.

Wisniewski, E. J. and Gentner, D. (1991). On the combinatorial semantics of noun pairs: minor and major adjustments to meaning, in G. B. Simpson (Ed.) Understanding Word and Sentence. Elsevier Science Publishers B. V. North-Holland.

Wittgenstein, L. (1953). Philosophical Investigations, (Trans. 1958, G. E. M. Anscombe). Basil Blackwell, Oxford.

Zadeh, l. A. (1965). Fuzzy Sets, Information and control, 8, 338-153.

Zadeh, l. A. (1982). A note on prototype theory and fuzzy sets, Cognition, 12, 291-297.