The exemplar-based approach to categorisation has recently become perhaps the most influential one within cognitive psychology. Still, it seems to be largely disregarded in other fields of cognitive science, such as philosophy or linguistics. In this article, I describe exemplar models against the background of classical and prototype theories, and try to explain the difficulties concerning wider application of exemplar models to categorisation studies in lexical semantics. I begin with a few general remarks on the nature of categorisation and the necessity to understand this notion in a broader theoretical frame. In a brief discussion of models of categorisation, I focus on the distinction between classical and ‘similarity’ models (the latter meaning prototype and exemplar models) and its far reaching implications, as well as explain the distinction between categorisation by prototypes and exemplars. By way of conclusion, I venture a suggestion that in thinking about categorisation in humans, different views should be treated as complementary rather than contradictory accounts.

* I use American spelling in the title, but British spelling elsewhere in the text.
Before exploring the controversy over the phenomenon of categorisation, I think it would be prudent to devote a few words to a still more basic aspect of this problem. The core term in this article, ‘categorisation’, is a rather capacious one and might itself be the primary source of potential confusion. This central notion can be understood in a number of related, though not thoroughly compatible ways, and hence it will be useful to relate it to a particular scientific tradition.

It seems that in the philosophical tradition, categorisation is routinely thought of as a process whose units – categories – neatly correspond to either words of language, or at least to some kind of semantically evaluable concepts. Still, there is another possibility to understand categorisation, a much broader and basic one, which, originating perhaps in cognitive psychology, seems to be the received one throughout cognitive science today. On this understanding, categorisation is a fundamental and ubiquitous process, whose nature needs not necessarily be mental, and categories need not necessarily have lexical or even conceptual correlates. Lakoff and Johnson (1999:17) state: “[e]very living being categorizes. Even the amoeba categorizes the things it encounters into food or nonfood, what it moves toward or moves away from. The amoeba has no choice as to whether to categorize; it just does. The same is true at every level of the animal world. Animals categorize for food, predators, possible mates, members of their own species, and so on.”

As one can see, not only language and thought are based on categorisation; the same pertains to our most basic sensorimotor behaviour. Harnad (2002) in particular capitalises upon the essentially sensorimotor foundations of this phenomenon. Categorisation is our intrinsic capacity:

---

1 A few terminological remarks are necessary. In this paper, categorisation in its core sense is understood, rather broadly, as the act or process of assigning some stimulus to a given category. The interpretation of the word ‘categorisation’ based on popular use admits conflicting intuitions that result in its questionable application to category formation. ‘Categorisation’ is indeed sometimes used in the sense of category formation even in specialised literature (e.g. Nęcka et al. 2006: 101–103, Aarts 2006). Such a conflation may result from speaking about ‘categorisation’ to denote a general area of theoretical interest, rather than a specific psychological process. In the context of an individual cognitive agent, e.g. a human child, this leads to the question of ontogenetic priority between categorisation and categories, as it seems to suggest, incorrectly, that (mental representations of) categories are formed by the process of categorisation. While category acquisition/learning/formation is clearly different from categorisation, it is also clearly related. On some accounts category formation might be the reverse of categorisation; for example – as noted by Laurence and Margolis (1999: 11) – if categorisation is assumed to consist in checking for the required criterial features as posited by the classical approach, category learning may “run in reverse”, consisting in assembling the features that form the category.
(...) to sort the blooming, buzzing confusion that reaches our sensorimotor surfaces into the relatively orderly taxonomic kinds marked out by our differential responses to it -- including everything from instrumental responses such as eating, fleeing from, or mating with some kinds of things and not others, to assigning a unique, arbitrary name to some kinds of things and not others (...).

It is easy to forget that our categorisation capacity is indeed a sensorimotor capacity. In the case of instrumental responses, (...) what we tend to forget is that these nonarbitrary but differential responses are actually acts of categorisation too, partitioning inputs into those you do this with and those you do that with.

In this tradition, categorisation is basically any act of grouping of inputs (be they perceptual or conceptual); the result of this reduction of inputs is a common memory representation, which may take the form of a concept, or a lexical label, but also of some kind of a specific way of behaviour (‘unitary sensorimotor response’), as in the examples quoted above. Categorisation is thus not only a foundation for cognition; it is a foundation on which all behaviour is dependent.

Following Ogden and Richards (1923, cited after Schulte, 1995:46), one can describe a conceptual category (ideally, corresponding to a lexeme) in a triadic relation illustrated by the semiotic triangle (Fig. 1 below). The linguistic sign – lexical label – is what individuates the category; and, as it is commonly assumed, it stands in the relation of reference to (some element of) extralinguistic reality. However, this link is not direct, but is mediated by a mental representation, or a concept, in the head of the language user. This points to the essentially conceptual character of lexical-semantic categorisation.
My point, much in line with the claims of the ‘embodiment’ tradition (e.g. Lakoff and Johnson, 1999, Varela et. al., 1993) is that the higher cognitive processes stem from, and have much in common with, the domain of perception. Goldstone and Barsalou (1998) demonstrate a number of ways in which conceptual processes have their roots in, and draw on the same computational resources as, perceptual processes, including the application of overall similarity to the creation of many common categories, an issue to which I will return shortly. The authors emphasise the notion of a continuum between perceptual and conceptual processes. Harnad (1990) shows that all symbolic categories agents use for communication must ultimately be grounded in sensorimotor categories, but I think that the list of similarities between perceptual and conceptual categorisation goes even further; I will elaborate on this point later in the text.

**Categorisation: classical versus ‘similarity-based’**

It is not my intention here to present a comprehensive overview of the theories of categorisation (for a review from the cognitive standpoint see e.g. Medin and Smith, 1984; Schulte, 1995: 49-72; Taylor, 1995: 21-75; Medin, Ross and Markman, 2001: 371-95, and especially Murphy 2003). I would like, however, to focus on the distinction into the so-called classical theories and ‘similarity’ theories, that is, prototype and exemplar theories. The term ‘classical’ here refers generically to the line of thought, going as far back as Aristotle, that pervaded and dominated this field of study until relatively recently.

On the classical view, categories are defined in terms of conjunctions of necessary and sufficient features. Taylor (1995: 79-80) summarises the main assumptions of the classical view of categorisation:

a) all members of a category have equal status
b) all non-members of a category have equal status
c) there is a fixed set of necessary and sufficient conditions defining membership to each category
d) all necessary and sufficient features defining a category have equal status
e) category boundaries are fixed
We could perhaps complement this set with an additional assumption of a more general philosophical kind; one on which this approach seems to be founded:

A) categories reflect the pre-existing, objective structure of the world

As is well known, the classical approach came in for severe criticism, the principal sources of which were Ludwig Wittgenstein’s (1958) reflections on the phenomenon of *family resemblance*, as well as the growing body of empirical evidence (Labov, 1973, and Rosch, 1973, 1975; after Taylor, 1995). Classical models were shown, in particular, to be unable to find the defining properties of categories, nor to account for *typicality effects*, classification of borderline cases, and application of nonnecessary properties in categorisation decisions (cf. Medin and Smith, 1984: 115-117). This challenge gave birth to an alternative approach to categorisation, i.e. ‘similarity’ views. Most importantly, they were compatible with a rather different philosophical stance:

A’) categories do not mirror ‘objective reality’; cognitive agent, rather than ‘discovering’ pre-existing categories, ‘negotiates’ or constructs them out of originally unstructured reality

More extensive acknowledgement of the role of the cognizer in the process of cognition bore implications evident in the ‘similarity’ approach to categorisation.

a’) membership in a category is gradable
b’) similarity ratings extend to cover non-members of a category, so that even the entities that technically fall outside the category borders can be rated as more, or less, similar to the category prototype or to stored exemplars
c’) no single feature can be necessary for category membership
d’) features by which we categorise are weighted, that is, some are more important than others
e’) category boundaries are fuzzy; categorisation decisions are a matter of judgement; they might vary across individuals and depend not only on the object of categorisation, but also on external, contextual factors,

---
2 It is worth noting that the impact of Wittgenstein’s reflections has been disproportionate to their brevity.
Categorisation: prototypes

Subscribing to the above set of statements, the prototype view of categorisation has for many years been the chief alternative, or a successor, to the classical view. The basic idea behind prototype models was that people go about a categorisation task by assessing the similarity of a new instance to the prototype of the category. There are at least two major points to be addressed at this stage.

First of all, ‘similarity’ is an immensely vague variable, which will become evident in the discussion of exemplar models. After all, for any arbitrarily chosen pair of things the elements of the pair can be estimated as similar (or dissimilar) in one respect or another, and “a zebra and a barber pole could be seen as more similar than a zebra and a horse, if the feature striped is given sufficient weight” (Medin, Ross and Markman, 2001:387). Such problems can be avoided by introducing the notion of typicality. Typicality is similarity only in respect of the key attributes, that is, those attributes that form the prototype representation; or in other words it is “a constrained form of similarity in which the respects (and their relative importance) can be determined by the conceptual representation itself” (Hampton 1998:139).

The other important point concerns the nature of the prototype representation. In a more colloquial sense, the term ‘prototype’ can be used to refer to a prototypical member or ‘best example’ of a category. But ‘prototype’ can also mean the prototype representation of a category. It needs to be made clear that in this case the ‘prototype’ can be a particular exemplar³; still, usually it is none of its particular instances, but rather an abstract, generalised representation in the format of a set of criterial features that may also be weighted. According to the latter interpretation, a trout can be said to be the prototype of the category FISH only in so much as it is likely to be judged a prototypical member of this category, but the prototype representation of the category FISH is an abstract representation consisting of (weighted) features like ‘lives in water’, ‘has fins’, ‘is slimy’, and so on.

³ Either a particular instance, or a hyponym, such as ‘sparrow’ to ‘bird’.
Categorisation: prototypes versus exemplars

The exemplar view of categorisation is in many important respects akin to the prototype view. Both those approaches converge on the same theoretical assumptions listed above (A’, a’ to e’), and have often been grouped together as the ‘similarity’, or ‘probabilistic’ views. Consequently, they might not always get differentiated, with the exemplar view being sometimes subsumed under the ‘prototype’ label, especially in the crucial aspect of the opposition to the classical view; as we shall see, the distinction between prototype and exemplar models is not easy to make in every context.

Exemplar models appear to have been somewhat neglected in cognitive science, at least until recently. Although exemplar models date back to the 1970s (cf. Medin, Ross and Markman, 2001:379), the influence of the exemplar view on cognitive science has been relatively limited, especially in comparison to the tremendous impact that the prototype view has exerted in the areas of linguistics and philosophy. This might have been due to hindered interdisciplinary communication: exemplar models, developed within mathematical psychology, remained rather technical, with a strong bias towards the field of perceptual categorisation. A technical debate over the application of exemplar and prototype models continues between Robert M. Nosofsky and J.D. Smith (e.g. Nosofsky and Johansen, 2000; Nosofsky and Zaki, 2002). Storms (2004: 5–7) identifies two major stumbling blocks in the application of exemplar models to semantic-lexical categorisation. Firstly, instances of natural language concepts seem to have to be abstractions themselves (“what is an exemplar”), and secondly, the appropriate selection of features is problematic (“what are the relevant features”). One might add that the emphasis on the statistical and distributed nature of representation leads to heavy reliance on advanced computational procedures. While features and dimensions used in perceptual categorisation are much more amenable to purely quantitative treatment, for lexical-semantic categorisation this is much more problematic. In the case of lexical categories, some level of intentional description seems to be inevitable, which may be more difficult to reconcile with the inherently mathematical-computational nature of the exemplar approach.

The basic assumption of the exemplar models is to deny the abstraction of criterial features⁴. On this view, people store in memory all the encountered instances of a given category to form a distributed, collective representation of this category. Most models assume

---

⁴ That is, at least to some degree; this issue will be developed further in the text.
that every occurrence of an item leaves a separate memory trace; but in some we store only individuals, so that each occurrence of the same individual would not leave a separate trace, but update and strengthen the present one for this individual (Barsalou, Huttenlocher and Lamberts, 1998). Information preserved in each memory trace is assumed to be fairly extensive, including non-essential and contextual properties, which results in the system’s building up a rich base of category-related knowledge. According to exemplar-based models, a categorisation decision process consists in retrieving from memory a number of exemplars that are similar to the test instance, and classifying the test instance on the basis of its relative similarity to the exemplars from the various categories. This means that e.g. on coming across some animal, we invoke a number of similar exemplars, choose the few closest matches and extract the category label that they have in common, and assign the label (e.g. ‘cat’) to the encountered animal.

As stated above, the exemplar view of categorisation exhibits substantial similarities to the prototype view. Typicality effects, even including the prototype enhancement effect, will hold in exemplar-based models; however, they will be accounted for in a different way, without assuming them to be a direct consequence of the category’s representation having the prototype format. According to the exemplar view, an exemplar-based model will be most efficient with instances that are highly similar to a large number of stored exemplars, that is, will predict the fastest and most accurate categorisation for such instances. More prototypical instances of a category are, as a rule, more similar to (a greater number of) stored exemplars of the category than non-prototypical instances, and, as such, would be categorised faster and more accurately, thereby giving rise to typicality effects. ‘Prototypes’, therefore, would not be inherent, but resultant in the system; not so much the actual representation of a category but rather an emergent property, generated in the process of categorisation-related processing.

It has been observed before that characteristic of the exemplar view is denying reduction and postulating the level of individual instances as the level on which a category is represented. Consequently, exemplar-based models presuppose the preservation of relatively rich information on each category, in comparison to only ‘skeletal’ category-related information preserved by classical or prototype models. The former position seems to be consistent with the way in which we put categories to use, e.g. in inductive inference. People possess knowledge not only of the attributes of a category, but also on its internal structure, including the correlations between the attributes; this supports within-category predictions and inferences. The canonical examples here involve people’s ability to make the inferences that if a bird is large, it is unable to sing (Medin, Ross and Markman, 2001:379), and if it has
large wings, it is more likely to inhabit regions near the sea and live on fish (Medin and Smith, 1984:125).

What is more, people apparently base their categorisation decisions on attributes that are non-critical to category membership. This would not be so surprising in situations where the defining attributes were obscure and had themselves to be established first. But Smith et. al. (1998) found evidence that subjects’ categorization decisions were influenced by irrelevant features even when they had been provided with an explicitly formulated rule to base their decisions on. All the above examples suggest that categorisation cannot be reduced to a fairly narrow set of criterial features, and that a broad foundation of encyclopaedic knowledge cannot be eliminated from the classification process.

**Similarity revisited and the representational format of features**

As noted above, an important advantage of the exemplar view is the introduction of background knowledge that, as empirically proven, bears on categorisation. This brings us back to the notion of similarity: exemplar-based models classify instances depending on how similar they are to other stored instances, but here the variable ‘similarity’ becomes unconstrained again. Since the process goes beyond any ‘key attributes’, ‘similarity’ cannot be redefined as typicality, as was the case in prototype models. Such a flexible, unconstrained notion of similarity, although extremely problematic (this was mentioned before, but will be discussed later in the text), is at least as much appealing. In particular, bringing forward overall, holistic similarity allows one to greatly reduce initial assumptions about the *category’s representational format*. Ideally, this would mean that the category representation would not have to be composed of a list of verbally encoded features.

All of the most common models of categorisation relevant to natural language categories propose that a systems’ representation of a category has the form of a set of attributes that can be characterised verbally; that is, the attributes are either simple ‘concepts’ like [+/- ANIMATE; +/- MALE], atomic word meanings, or short propositional descriptions. In all those cases an underlying assumption can be traced, an assumption which, even if necessary and unavoidable, needs to be pronounced explicitly: that adequate and relevant categorisation features are overt and can be given in a verbal representational format. The

---

5 Smith et. al. (1998) quote and then replicate (with a few alterations) Allen and Brooks’ experiment, in which subjects were considerably worse at classifying those instances that, though technically members of the category, were at the same time highly similar (in non-criterial respects) to non-members of the category.
generation of such features requires analytical scrutiny of the studied concept, involving selective attention and drawing on working memory resources, and the subsequent encoding of the features in a symbolic format (in the process of describing a feature verbally). The cognitive components used in such a task are thus characteristic only of controlled, rule-based processing.

It is never guaranteed that the features which we use in this analytical, post-factum, symbolic description of a particular category are exactly the features that we actually base our categorisation decision upon. It might be useful to revert to Wittgenstein’s famous example of family resemblance in a literal way. On encountering a (known) human face we are instantly able to categorise it as a particular familiar face; we complete this categorisation task using low-level cognitive processes that do not enter our consciousness. We can describe the face verbally, employing propositions such as ‘has bushy eyebrows’, ‘has a long nose’ or ‘has a pointy chin’, but there is little doubt that even a very detailed description of this kind will be highly unsatisfactory; the features actually involved in the categorisation process must be of a rather different nature. The true features are below the level of our verbal access.

One way of reaching beyond the symbolic-descriptive level is to include ‘raw’ perceptual similarity into categorisation, basing on more analogue than conceptual representations. At least two categorisation mechanisms seem to be relevant to this stance: categorisation by blurring and by rating overall similarity to an exemplar.

It has been argued that judging category membership by overall perceptual similarity is an often used categorisation procedure. It is accomplished by a rapid response to a stimulus by holistic processing, without – or prior to – breaking the stimulus down into separate features or aspects. Goldstone and Barsalou (1998:240-41) make a case for wider application of undifferentiated overall visual similarity especially in rapid categorisation. A related strategy is ‘categorisation by blurring’, whereby one classifies novel instances without forming the precise schema of a category, but purely by comparing the novel instance to a known member of the category and blurring over the irrelevant differences. “One does not need to know what makes something a dog in order to categorize the neighbor’s poodle as a dog, as long as one knows that a beagle is a dog, and is able to ignore (blur over) the differences between poodles and beagles” (Goldstone and Barsalou, 1998:250).

\[6\] Jackendoff (1990 [1983]: 42–43) introduces the example of face recognition to reach a similar conclusion, although in a slightly different context.
Exemplars and the context of natural language

Quite obviously, perceptual similarity cannot be treated as the sole basis of conceptual categorisation. Abstract concepts or terms from psychology and linguistics constitute the most evident examples of categories in which perceptual cores are clearly absent. But neither is overall similarity a sufficient ground for categorisation in the case of natural objects; even young children are not misled by external appearances and suspect deeper structural regularities to underlie categories. On the other hand, the notion of ‘overall similarity’ unrestricted to perceptual similarity, but holding over an unspecified number of psychological dimensions, is rather inconceivable. Furthermore, there is a problem of directionality in the relation between similarity and categorisation, as one can give rise to the other and vice versa: not only can an item be categorised on a basis of similarity, but two items can be viewed as (more) similar because they share a category label (cf. e.g. Schyns, 1997; Medin et. al., 2001:389).

For all those reasons, the application of nonsymbolic representations is fairly limited in conceptual categorisation. In creating a transparent theory of linguistic categorisation, one simply has to fall back upon conventional featural representations. The assumption that prototypes or exemplars are accurately represented intensionally in the form of a list of verbally formulated features (like *is sweet* or *has seeds*) is practically inevitable. This results in a partial limitation of the potential merits of the exemplar view, bringing this conception closer to the prototype view. The following passage from Storms et. al. (2000:53) illustrates comprehensively the distinction between the two views in the context of lexical categorisation, a distinction that becomes rather a matter of degree than a sharp qualitative change:

To summarize, in the context of natural language categories like *fruits*, *vegetables*, *vehicles*, etc., three different theoretical views may be distinguished depending on the levels at which abstraction does or does not take place. The first view assumes that no abstraction whatsoever takes place and that only memory traces of particular encountered instances are stored. Any category-related judgment is based on these memory traces, as no abstract information is stored with verbal concepts. The second view assumes that abstraction may take place, but only at a level lower than the concepts studied, that is, at the level of *tomatoes* in case *vegetables* are studied. The representation of the studied natural language concepts, like *vegetables* and *vehicles*, is comprised of lower level concepts like *tomatoes* and *bikes*, respectively. Finally, the third view states that abstraction (also) takes place at the level of the studied natural language concepts and that (characteristic) features of their exemplars are directly stored at this level. The latter view can
be labeled the prototype view, and the first two views are exemplar views. (Storms, De Boeck, and Ruts, 2000: 53)

In the above formulation, the difference between exemplar and prototype model is very much reduced to the type of computational process (calculating feature overlap with many exemplars rather than a single prototype). But even here some of the strengths of the exemplar view can be retained, namely, it preserves substantial body of category-related knowledge. Because a category representation comprises \( x \) stored exemplars, and each of the exemplars can be characterised by a list of \( y \) features, the distributed representation effectively consists of the total of \( x^*y \) features; many of them will of course be common to most of the exemplars, but the total number of nonrepeating features will still be greater than for prototype models. Idiosyncratic features of seemingly marginal relevance, that would be lost in classical and prototype models, still reside in the representation and, though pushed to deep background, can nevertheless be retrieved to influence categorisation. Moreover, information about feature correlations is preserved, too (such as \( \text{is large} \sim \text{is unable to sing} \) for BIRD). Those points are particularly significant if one remembers the dynamic character of categorisation: categorisation decisions tend to be influenced by context (e.g. the same tool held by a cook and a soldier might be judged as a kitchen utensil or a weapon, respectively), and the inclusion of a broader knowledge-base helps to account for this context-dependence.

**Conclusion**

One important statement that should be made at this point is that different theories of (linguistic) categorisation need not necessarily be considered as conflicting. Especially classical versus ‘similarity’ models can be viewed as complementary rather than mutually exclusionary both in simulating the actual process by which we categorise and for the formal description of categories. As to the descriptive side of the problem, the matter amounts to setting a distinction between encyclopaedic and linguistic knowledge. The distinction is not supported on empirical grounds, but if we are willing to accept some idealisation and set an arbitrary division, the classical view offers a transparent and formally elegant account of the categories from the point of view of linguistic knowledge. Incorporation of an increasing proportion of encyclopaedic knowledge into the models of categorisation results in greater psychological reality at the cost of becoming more and more elusive to a precise formal description. Also, in categorisation as a process, we are inclined not only to rely on different
strategies of categorisation (rule-based, described well by the classical models, and instance-based, described by the exemplar models) depending on the type of the category, but also apply the different procedures simultaneously to the same task (cf. Smith et. al, 1998). It is important to remember that so far, no single theory of categorisation can claim to be the only right theory.

BIBLIOGRAPHY:


[Online]: http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad02.symlang.htm


