Title
The Structure of Semantic Memory: Category-based vs. Modality-based

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

ISSN
1069-7977

Author
Gottlieb, Jeremy F.

Publication Date
2005

Peer reviewed

The Structure of Semantic Memory: Category-based vs. Modality-based

Jeremy F. Gottlieb (jgottlieb@carthage.edu)
Carthage College, 2001 Alford Park Drive
Kenosha, WI 53140 USA

Abstract

A source of significant debate in psychology is the issue of how information is stored in semantic memory. The two primary frameworks are the Unitary Content Hypothesis, which holds that information is stored based on categories, and the Multiple Semantics Hypothesis, which holds that information is stored based on sensory modality. In a series of three experiments, I attempt to shed some light on which of these two frameworks is the most probable explanation of a number of phenomena associated with semantic memory. The results indicate that the Multiple Semantics Hypothesis is the most likely explanation.
Keywords: semantic memory

Introduction

A significant debate in cognitive science centers around the question of the underlying organizational structure of the information stored in semantic memory. Specifically, the question that has been argued is whether information is organized categorically (e.g., Caramazza, Hillis, Rapp & Romani, 1990; Caramazza & Shelton, 1998) or based on the sensory modality of the information (e.g., Shallice, 1987; McCarthy & Warrington, 1988; Beauvois, 1982).

The prototypical theory that posits a category-based organization is the Unitary Content Hypothesis (UCH) of Caramazza et al. (1990). The UCH posits that semantic memory is an abstract, amodal storage system where information is organized categorically. Every input modality can access all of the information about a given category in semantic memory, and semantic memory provides information to every output modality. Thus, damage to semantic memory should affect semantic processing in a category-based fashion.

There are several reports of patients who exhibit category-specific deficits (see Forde & Humphreys, 1999, and Caramazza & Shelton, 1998, for review). Patients suffering from these deficits are generally impaired at processing semantic information about living things *vis a vis* man-made objects (e.g., Warrington & Shallice, 1984; Gainotti and Silveri, 1996) or vice versa (e.g., Warrington & McCarthy 1983, 1987).

The representative theory of a modality-based semantic organization is Shallice's (1987) Multiple Semantics Hypothesis (MSH). The MSH posits that semantic memory is actually divided into different semantic subsystems based on the various input modalities. There is a visual-semantic system, a tactile-semantic system, a verbal-semantic system, etc. These systems are linked together, allowing for information to pass between them. Categories are a result of how these systems are linked together. Pieces of information that are related to each other have strong links with each other, regardless of the subsystem they are stored in. Thus, while the associations between pieces of information are still category-based, the physical layout of information in semantic memory is based on modality. Furthermore, a given input modality can only directly access semantic memory through its corresponding semantic subsystem, and only accesses the other subsystems indirectly through the links between them. Damage to the semantic system should result in selective impairments in processing semantic information associated with particular input modalities, instead of particular categories.

There are a number of reports of patients demonstrating modality-specific semantic deficits, such as optic aphasia (Beauvois, 1982; Riddoch & Humphreys, 1987). In this deficit, patients are unable to name objects that are presented visually, but are not impaired at naming them from any other sensory domain, such as touch. They are also unable to point to the proper object when given its spoken name. However, these patients still exhibit semantic knowledge of the objects they were unable to name, such as being able to correctly mime the use of the objects and being able to draw objects and complex scenes upon a verbal request. Beauvois (1982) also reports similar cases of tactile and auditory aphasia. According to proponents of the MSH, the most parsimonious explanation of optic aphasia is to theorize that the connection between visual semantics and verbal semantics has been severed.

Much of the rest of the evidence brought to bear in this particular debate is also relatively ambiguous. One frequently reported finding is that words are named faster than pictures, but pictures are categorized faster than words (e.g., Potter & Faulconer, 1975; Guenther, Klatzky, & Putnam, 1980; Seifert, 1997; Theios & Amrhein, 1989). However, word "naming" is really just word reading, which arguably involves simply mapping graphemes to phonemes and can bypass semantic memory entirely, as in many dual-route models of word reading (Coltheart, Curtis, Atkins, & Haller, 1993).

PET scanning has showed that areas of the brain related to visual processing are more active when living things are being named, while those related to action are more active when tools are being named (Martin, Wiggs, Ungerleider, & Haxby, 1996), indicating that information in semantic memory is organized by sensory modality. However, there is no true indication that this is the result of a modality-based organization instead of a category-based one. It could simply be the case that living things are processed in one part of the brain and man-made objects in another.

Experiments

The experiments described below were designed to better examine how information is organized in semantic memory. All three experiments shared a common methodology. Participants were simultaneously presented with two stimuli. The stimuli were either both pictures, both words, or there was one of each type of stimulus. The two things that were varied across the experiments were the exact nature of the stimuli (what they depicted or described) and the question that participants were supposed to answer upon being presented with the stimuli.

Experiment 1 – Natural categorization

The primary purpose of Experiment 1 was to give a better overview of the time-course for making judgments about category membership for both words and pictures.

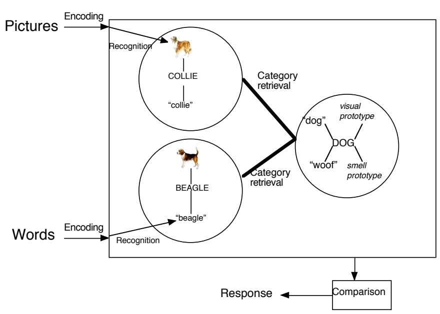

Figure 1: UCH model of Experiment 1

Predictions The task in Experiment 1 requires four basic processing stages:

1. Perceptual encoding and recognition – converting the input stimulus into a usable representation and finding the semantic representation of the input.
2. Category retrieval – retrieving the superordinate category of the input
3. Comparison – comparing the two retrieved categories to each other.
4. Response preparation and generation

There is no difference between the models for steps 1, 3, and 4. Step 1 happens independently for the two stimuli, and steps 3 and 4 occur independently of how semantic memory is structured. Thus, the crux of the differences between the two models occurs at the point of category retrieval. To see how, we will examine what each model has to say about the cognitive processing involved in this task.

If the UCH is correct, then both the picture and word representations of objects should access the same semantic representation of the concept, as in Figure 1. Thus, both pictures and words ought to be equivalent with regards to retrieving the superordinate category of that representation. Thus, we can postulate that the reaction times[1] for participants across the three stimulus conditions should take the form of:

Picture-picture: $(2*ER_p) + (2*CR) + C + R$
Picture-word: $ER_p + ER_w + (2*CR) + C + R$
Word-word: $(2*ER_w) + (2*CR) + C + R$

In other words, participants' reaction times across the three conditions should be essentially linear with a slope of $|ER_p - ER_w|$ if the UCH is correct.

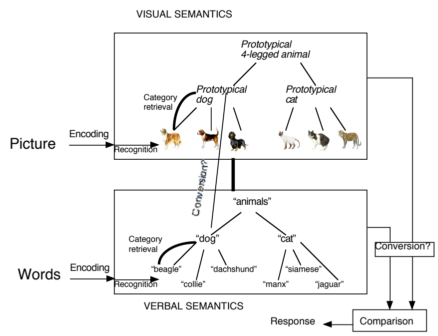

Figure 2: MSH model of Experiment 1

Under the MSH the category retrieval stage becomes more complicated. In particular, the issue arises of whether or not the category information being retrieved is the same for both pictures and words. If it is not, then some sort of conversion step becomes necessary in the picture-word condition (see Figure 2) to allow for a comparison of the category retrieved by the picture to the category retrieved by the word. If this is the case, then we would postulate that the reaction times across the three stimulus conditions would take the form of:

Picture-picture: $(2*ER_p) + (2*CR) + C + R$
Picture-word: $ER_p + ER_w + (2*CR) + Conversion + C + R$
Word-word: $(2*ER_w) + (2*CR) + C + R$

Thus, in this case, we would expect to see a non-linear pattern of reaction times, with the picture-word condition showing slower reaction times than it would if the distribution were linear.[2]

Stimuli The picture stimuli were 138 colored pictures of living things that fell into nine different categories – dogs, cats, horses, fish, trees, birds, fruits, vegetables, and flowers. The word stimuli were the names of the objects used as the pictures. Stimuli were presented side by side on a Macintosh

---

[1] ER_p is encouding and recognition for pictures; ER_w is the same for words; CR is category retrieval; C is comparison; R is response preparation and generation.

[2] The predictions presented here are predicate on the idea that the processing stages are linear, as per the traditional models of memory. However, even in a system where one or more of these stages are performed in parallel, the MSH would still presume separate semantic representations for visual and verbal information, and thus there would still be a cost associated with having to transfer information from one semantic subsystem to the other.

computer using the PsyScope experiment software (Cohen, MacWhinney, Flatt & Provost, 1993). The pictures were normalized to all be roughly 1.5 inches square and the words were presented so that they were, on average, roughly as wide as the pictures. There was, on average, a one inch horizontal separation between the two stimuli.

Procedures As noted above, on any given trial, participants were presented with two stimuli simultaneously – two pictures, two words, or one of each. Within a block of trials, no picture or word was repeated, and only one member of the picture/name pair would be presented. Thus, if a participant saw the word "beagle" at some point in a given block of trials, they would not see the picture of the beagle within that same block. Each participant was given three blocks of 69 trials each.

For individual trials, there was an equal probability for which type of stimulus pair participants would see. In the picture-word trials, there was an even chance that the stimuli would appear with the picture on the left and the word on the right or vice versa.

Participants were asked to judge whether the two stimuli "depict or describe" objects that belong to the same category. They were not told what the possible categories were, nor were they told any criteria to use to categorize the objects. The stimuli were left on the screen until the participant responded.

Results Trials that participants answered incorrectly or that took longer than 3000ms were excluded from this analysis. These accounted for 11.6% of the total trials (9.3% and 2.3% respectively).

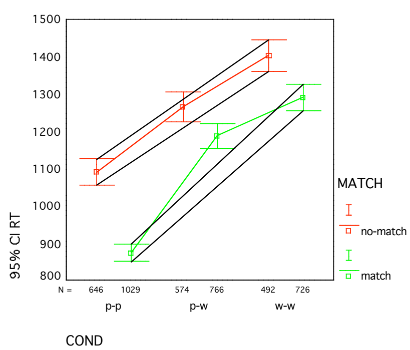

Figure 3: Mean reaction time by condition by response for Experiment 1

As can be seen from Figure 3, there was a deviation from linearity by the picture-word condition when the two stimuli belonged to the same category (the match condition), but not when they belonged to different categories (the no-match condition). To confirm the statistical significance of this deviation, a contrast-coded regression analysis was performed. In the no-match condition, the analysis was not significant, $F(1) = 1.05$, $p > .05$. However, for the match condition, this test was highly significant, $F(1) = 30.78$, $p < .001$.

Discussion The results from Experiment 1 seem to provide contradictory evidence, based on the predictions made above. The results from the no-match condition would seem to confirm the UCH, while the results from the match condition would seem to confirm the MSH.

One likely explanation of this discrepancy involves a possibility that I failed to consider when generating my predictions. When a participant is presented with two pictures from the same category – such as a beagle and a collie – the first picture will be processed, which will activate the DOG category node in visual semantic memory. When the second picture is processed, it will need to access the same category node in order for the task to be performed properly. In this case, one could argue that a priming effect would cause the retrieval of the category for the second picture to be faster than it was for the first picture. Similarly, we would anticipate seeing a priming effect when both stimuli are words belonging to the same category.

However, when the stimuli are of different modalities (a word and a picture), they would access different category nodes – one in visual semantics and one in verbal semantics. Thus, there would be no priming effect in the picture-word condition, and so it should be slower than it would be if the distribution across the three conditions were linear. Likewise, for the no-match condition, none of the stimuli benefit from any sort of priming effect, and thus we would expect those reaction times to be linear.

Put another way, my predictions presumed that the match and no-match conditions would operate in roughly the same fashion, and thus be subject to essentially the same rules. However, this does not appear to be the case. The no-match conditions would be governed by the equations given above for the UCH. The match conditions would be governed by these same equations, except that the picture-picture and word-word conditions should subtract a priming effect. To wit:

Picture-picture: $(2*ER_p) + (2*CR - P) + C + R$

Picture-word: $ER_p + ER_w + (2*CR) + C + R$

Word-word: $(2*ER_w) + (2*CR - P) + C + R$

Thus, it would appear that what Experiment 1 has found is not a conversion parameter, but a priming effect.

The UCH has a difficult time explaining this discrepancy between the match and no-match conditions. Since the underlying feature of the UCH framework is that inputs of any modality activate the same category information in semantic memory, then presumably any semantic priming effect observed in the same-modality trials in the match condition should also be present in the cross-modal trials, since the picture and the word are accessing the same category node in semantic memory. As a result, the pattern of reaction times should still be linear across the three types of stimulus presentation in the match condition. Clearly, this is not what is happening.

There is, however, one possible explanation for these results that would allow the UCH to explain this deviation from linearity in the match condition. Specifically, there is the possibility that the deviation is due to a task-switching effect that results from having to mentally "switch gears," so to speak, from processing pictures to processing words (or vice versa). This effect would only be present in the picture-word case, causing the non-linearity I found.

Experiment 2 – Object/lexical decision task

The purpose of Experiment 2 is to determine whether the effect found in Experiment 1 was due to a simple task-switching effect.

Predictions This experiment uses a lexical/object decision task, where participants are asked to judge whether the pictures and letter strings they see are real or not. This task eliminates the category retrieval and comparison stages from the task in Experiment 1. In doing so, I am hoping to isolate any task switching effect in the perceptual encoding and recognition stage (since it clearly cannot be in the response stage). Thus, if we see a deviation from linearity in the pattern of reaction times similar to the one we saw in the match condition of Experiment 1, arguably that effect is due to a task switching effect, and not to a semantic effect.

Stimuli The stimuli for this experiment were the same pictures and words used in Experiment 1 (the *real* stimuli), as well as modified versions that were divided into two groups:

1. Plausibly unreal stimuli – pictures of objects with analogous parts interchanged (e.g. a beagle with a collie's head); pronouncable non-words (e.g. "mave").
2. Unreal stimuli – pictures of objects with non-analogous parts interchanged (e.g., the bottom of a beagle with the top of a tree); random, unpronounceable letter strings.

Procedures The procedures were exactly the same as those in Experiment 1 except for how participants were directed to respond. Half of the participants were instructed to decide whether both items they saw were real or not (2-yes) while the other half were instructed to decide whether at least one item was real (1-yes). This was to control for the fact that in the 2-yes condition, participants could also just respond "no" as soon as they saw an unreal item, meaning that rather than always having to look at both stimuli and make a decision, they could just be making their decisions as soon as they saw a stimulus that determined the answer (an unreal one). Thus, the 1-yes condition was used to balance this by having a set of trials where the deciding stimulus was the real one.

Results As before, incorrect trials and those trials over 3000ms were excluded from the analysis. This accounted for 11.6% of the total trials (8.8% and 2.8% respectively). The results from the 1-yes and 2-yes conditions were collapsed as there was no statistical significance between the groups.

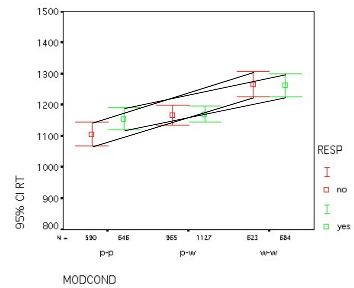

Figure 4: Reaction time by condition by response

As can be seen from Figure 4, there is no apparent deviation from linearity in either yes or no response conditions. A contrast-coded regression confirms the probable linearity of the results in both the yes condition ($\underline{F}(1) = 0.57$, $\underline{p} > .10$) and the no condition ($\underline{F}(1) = 3.06$, $\underline{p} > .05$).

Discussion The results from Experiment 2 demonstrate that the non-linearity observed in the match condition of Experiment 1 is not, in fact, due to a task-switching effect in the perceptual encoding and recognition stage. This was expected since a task-switching effect presumably should have also been present in the non-match condition of Experiment 1, but the results from Experiment 2 provide a more concrete demonstration of the absence of a task switching effect in this experimental design.

As a result, it is reasonable to conclude that a MSH-based model of Experiment 1 in conjunction with a semantic priming effect in the like-modal, match condition trials is the most likely explanation for the data gathered in Experiment 1.

Experiment 3 – Functional categorization

Experiment 3 was designed to provide a direct test of one of the key components of the MSH model. Certain neurological patients have category-specific deficits, where they show severe impairments when trying to semantically process living things *vis a vis* man-made objects, or vice versa. This syndrome is difficult to explain under the MSH framework – how can there be category-specific deficits resulting from a modality-based organizational structure?

The solution was the sensory/functional hypothesis (SFH). The SFH is a theory about what exactly separates living things from man-made objects within semantic memory. The main idea is that there are significant differences in the composition of the representations for the different classes of objects. We interact with natural objects primarily in a visual fashion – we observe trees and flowers and animals and what-not and classify them based on what they look like. Thus, the bulk of the features used to represent these objects in semantic memory are going to be visual in nature, and thus stored in visual semantic memory.

Man-made objects, on the other hand, are interacted with in a primarily functional fashion. We use man-made objects, we build them to serve particular functions, and we classify them based on what they are designed to be used for. Therefore, the representations of man-made objects are going to contain proportionally more functional features as opposed to visual features. Arguably, these functional features would be stored in a semantic subsystem other than visual semantics.

Category-specific deficits, in this hypothesis, are simply the result of damage to either the visual semantic system or to whatever system contains functional information. Experiment 3 was thus designed to test whether or not functional semantic information is stored differentially from at least visual semantic information, and possibly from verbal semantic information.

Predictions Experiment 3 was designed to be as analogous as possible to Experiment 1, with the exception that the category retrieval stage in Experiment 1 needs to be replaced with two stages – a function retrieval stage and a functional category retrieval stage.

Under both the UCH and the MSH we would expect a linear distribution of reaction times across the three modality conditions. Under the UCH, both pictures and words would be accessing the same semantic representation of each object, including functional information[3]. Not only should the distribution be linear, but the reaction times across the three modality conditions should be commensurate with the corresponding reaction times from Experiment 1. If anything, they should be a little bit slower due to the added step of needing to retrieve the functional category after retrieving the function.

In the case of the MSH, however, the argument is that functional information is stored in a separate semantic subsystem from visual information. Thus, how the reaction times change will depend on precisely where this functional information is stored. If it is stored in close association with verbal semantics, then we would expect the picture-picture condition to be significantly slower than the same condition in Experiment 1 and the word-word condition to be significantly faster. On the other hand, if functional information is stored in a completely separate subsystem, we would expect reaction times to be significantly slower across all three conditions, to reflect the extra time needed to access the functional semantic system.

Stimuli The picture stimuli were 63 colored pictures of man-made objects. The words used were the names of the objects used as the pictures.

[3] Please note that when I refer to "functional information" what I am referring to is the semantic representations of functional information that we have stored, not to usage information that may be directly derived from visual input.

Procedures The procedures used were exactly the same as those used in Experiment 1, except that participants were asked to judge whether or not the two stimuli "depict or describe" objects that belong to the same *functional* category.

Results Once again, trials with errors or that took more than 3000ms were excluded. This accounted for 10.4% of total trials (8.8% and 1.6% respectively).

As can be seen from Figure 5, the reaction times seem to be relatively linear in both the match and no-match conditions. Coded-contrast regressions show that, in fact, the no-match condition does not deviate from linearity, $F(1) = 2.64$, $p > .10$. However, the match condition does show a slight deviation from linearity, $F(1) = 4.71$, $p < .05$.

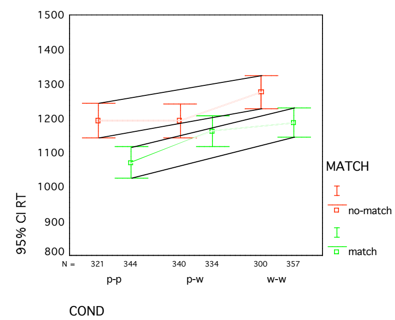

Figure 5: Reaction time by condition by response

Table 1 shows the reaction times from this experiment and Experiment 1 side-by-side. Of particular interest is the fact that the distribution of reaction times in Experiment 3 is much flatter than the distribution in Experiment 1. This difference is entirely the result of the picture-picture conditions being slower in Experiment 3 while the word-word conditions are faster. The picture-word conditions are virtually identical in Experiments 1 and 3.

Table 1: RT across conditions in Experiments 1 and 3

		Experiment 1	Experiment 3
match	Picture-picture	874.09	1071.04
	Picture-word	1187.26	1161.91
	Word-word	1290.69	1186.97
no-match	Picture-picture	1090.12	1192.53
	Picture-word	1265.32	1192.93
	Word-word	1403.02	1275.72

Discussion The results from Experiment 3, when compared with those from Experiment 1, match the predictions based on functional information being stored either in the verbal system, or in a system more closely tied to the verbal than to the visual semantic system (see Figure 6). In this scenario, it is clearly the case that pictures would have slower access to functional information than to categorical information as they would need to take an extra step to retrieve that functional information from the verbal/abstract subsystem.

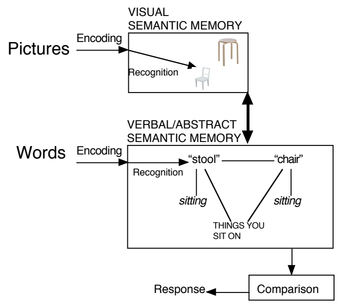

Figure 6: MSH explanation of Experiment 3

The fact that verbal input seems to have faster access to functional information than to categorical information can be explained by the fact that the functional information is possibly more strongly and/or directly connected to a particular concept node in verbal/abstract semantics than superordinate category information. This could potentially be the case for the man-made objects used in Experiment 3, as it would fit with the sensory/functional hypothesis discussed above.

Conclusion

The experiments presented in this paper evidence in favor of the multiple-semantics hypothesis explanation of how information is organized in semantic memory. The results of Experiment 1 are most consistent with the MSH view if one simply presumes, not without merit, that the like-modal match condition trials benefit from a semantic priming effect. Experiment 2 rules out the possibility of a task-switching effect being responsible for the results of Experiment 1, and the results from Experiment 3, when viewed in conjunction with the results from Experiment 1, are best explained by the MSH. Functional information would be more strongly related than category information to verbal semantics, but category information would be more strongly associated with visual semantics than functional information. Future work will explore categorization of man-made objects and functional categorization of natural objects to confirm these findings.

References

Beauvois, M. (1982). Optic aphasia: a process of interaction between vision and language. Philosophical Transactions of the Royal Society of London. 298, 35-47.

Caramazza, A., Hillis, A. E., Rapp, B. C., & Romani, C. (1990). The multiple semantics hypothesis: Multiple confusions? Cognitive Neuropsychology. 7, 161-189.

Caramazza, A. & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. Neurocase Special Issue: Category-specific deficits, 4(4-5), 339-351.

Cohen, J. D., MacWhinney, B., Flatt, M. & Provost, J. (1993).PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. Behavior Research Methods, Instruments, & Computers, 25(2), 257-271.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. Psychological Review, 100(4), 589-608.

Forde, E. M. E., & Humphreys, G. W. (1999). Category-specific recognition impairments: A review of important case studies and influential theories. Aphasiology. 13(3), 169-193.

Gainotti, G. & Silveri, M. C. (1996). Cognitive and anatomical locus of lesion in a patient with a category-specific semantic impairment for living beings. Cognitive Neuropsychology, 13(3), 357-389.

Guenther, R. K., Klatzky, R. L., and Putnam, W. (1980). Commonalities and Differences in Semantic Decisions about Pictures and Words. Journal of Verbal Learning and Verbal Behavior. 19, 54-74.

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. Science, 270(5233), 102-105.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. Nature, 379, 649-652

McCarthy, R. A. & Warrington, E. K. (1988). Evidence for modality-specific meaning systems in the brain. Nature. 334, 428-430.

Potter, M. C., and Faulconer, B. A. (1975). Time to understand pictures and words. Nature. 253, 437-438.

Riddoch, M. J., & Humphreys, G. W. (1987). Visual object processing in optic aphasia: A case of semantic access agnosia. Cognitive Neuropsychology, 4(2), 131-185.

Seifert, L. (1997). Activating Representations in Permanent Memory: Different Benefits for Pictures and Words. Journal of Experimental Psychology: Learning, Memory, and Cognition. 23, 1106-1121

Shallice, T. (1987). Impairments of semantic processing: Multiple dissociations. In M. Coltheart, R. Job, & G. Sartori (Eds.) The cognitive neuropsychology of language. London: Lawrence Earlbaum Associates Ltd.

Theios, J., and Amrhein, P.C. (1989). Theoretical Analysis of the Cognitive Processing of Lexical and Pictorial Stimuli: Reading, Naming, and Visual and Conceptual Comparisons. Psychological Review. 96, 5-24.

Warrington, E. K., & McCarthy, R. A. (1983). Category specific access dysphasia. Brain, 106, 859-878.

Warrington, E. K., & McCarthy, R. A. (1987). Categories of knowledge: further fractionation and an attempted integration. Brain, 110, 829-854.

Warrington, E.K., & Shallice, T. (1984). Category-specific semantic impairments. Brain, 107, 829-853.