A Memory Account of Children's Failure to Generalize Novel Names to Novel Instances and Novel Scenes

Thibaut, Jean-Pierre (jean-pierre.thibaut@univ-poitiers.fr)
University of Poitiers
LMDC, CNRS UMR 6215
99 avenue du recteur Pineau
86000 Poitiers France

Abstract

When they learn novel names, young children are thought to perfectly segregate an object from its environment and to associate it with its name no matter the scene in which the object is included and, in a post-test, to designate the correct object in novel scenes or contexts and to generalize the association to new instances that might differ from the original object according to various dimensions. We show, in two experiments, that children aged three do not always generalize new names across contexts and across instances even when instances are categorized in the same set as the learning stimuli. These results suggest that novel name generalization is, to some extent, independent of conceptual generalization. These results are interpreted in terms of general mechanisms of memory: it is argued that a failure to generalize novel words to novel stimuli or to new contexts might result from a lack of retrieval cues.

Young children are supposed to be proficient learners of novel names. Each day, young children learn new lexemes (Fenson, L., Dale, P., Reznick, S., Bates, E., Thal, D., Pethick, S., 1994). One important dimension of the task is to associate each novel word with the appropriate referent in the world and to be able to use this association for future namings. Creating the association is an instance of a general and fundamental learning mechanism (Siegler, 1997). Children can learn a novel category name from hearing its name only once (Smith, 1999). This ability is well captured by the notion of fast mapping. Children can grasp aspects of the meaning of a new word on the basis of a small number of incidental exposures, without any explicit training or feedback (Carey & Bartlett, 1978; Bloom, 2000 for an overview). Carey and Bartlett (1978) showed that many young children who had learned the novel word "chromium", still remembered part of its meaning six weeks later. Markson and Bloom (1997) showed that, in a pointing task, three- and four-year-old children performed well above chance after a month delay.

One important notion regarding lexical learning and generalization is that young children are thought to be perfectly able to segregate a target referent from the scene it is embedded in and to associate it with its name. For example, the concept of fast mapping implies that children grasp an object (for instance) and associate it with a word, no matter the scene, the spatial context, in which the object is included. Later, the child is supposed to be able to recognize, name or designate the correct referent in novel contexts. It is also generally recognized that children are able to generalize the new word to new instances that differ from the original object according to various dimensions (Clark, 1993; Mervis, 1987; Smith, 1999) (even though there is much debate regarding the dimensions according to which children will spontaneously generalize). This view of novel name learning and generalization predicts that the name of an object should be recovered later despite variations in the characteristics of the referents and of the scenes in which it was embedded in the learning phase.

The following experiments have several purposes. First, we assess the stability of performance across delays (immediate post-test, compared with a two-week delay). Following Markson and Bloom (1997), Bartlett and Carey (1978), we hypothesize that performance should be stable across delays. However, their results were obtained for a comprehension (pointing) task. In the same way, Childers and Tomasello (2002) assessed comprehension and naming at various intervals after training (immediate, one day and one week later) and found no effect of delay on both types of scores.

A second purpose of the present experiments, is to study generalization of novel names, as a function of superficial characteristics, such as the scene in which the stimulus was embedded during the learning phase and its congruence (see methods) with the scene in which the stimulus is embedded at test, or such as perceptual differences between the training referent and novel instances introduced later. More generally, our central question is whether the context in which the association between a novel name (e.g. the word "tapir") and a referent (the tapir) was "inserted" will influence future generalization tasks such as naming or pointing tasks. More specifically, we want to introduce memory factors in the question of novel word generalization. During lexical learning and generalization, children encounter objects belonging to the same category in various contexts. For example, a child plays with a new toy (e.g., a crane) and learns its name ("crane") in one room and, later, might play with a *new* crane in another room or another house. The same is true for pictures of objects: children can see a cat in a book and learn its name, while being confronted with a different cat later in another book.

Most lexical learning theories would assume that children spontaneously generalize a novel name to novel instances of the original referent, displayed in a new spatial setting if (i.e., a novel scene) (e.g., Mervis, 1987). In the same way, they do not consider the "delay" component of a lexical learning task as a central issue: in most experiments, one tests generalization immediately after a novel referent has been associated with a novel name and in situations in which the *learning object* remains *in view* (Mervis & Bertrand, 1994; Samuelson & Smith, 2000). Given the above evidence regarding delays (e.g., Childers & Tomasello, 2002) and the notion of fast mapping, memory factors do not seem to play any central role in lexical generalization. In other words, the implicit general hypothesis is that "context" parameters do not influence the encoding of the association between the word and the object and, thus, its later retrieval.

By contrast, our contention is that memory factors might interact with generalization of novel names to novel instances especially when they are displayed in new scenes. Our reasoning is based on the encoding specificity hypothesis (Tulving & Thomson, 1973), according to which an event is stored in memory together with the spatial and temporal information associated with the target stimulus during encoding (e.g., the scene a stimulus is embedded in). Later recovery is facilitated when the cues encoded during learning are available at test in the sense that properties of effective retrieval cues are determined by the specific encoding operations performed by the system on the input (see Baddeley, 1997; Balsam & Tomie, 1985).

In a lexical learning task, this view can be translated in the following terms: a child always learns a novel word for a particular instance of a category in a particular scene, which results in the encoding of retrieval cues that are specific to this learning setting. Later presentations of novel instances of the same category in new scenes should elicit, by definition, poorer retrieval cues than the retrieval cues provided by the original instance displayed in its original scene. This should result in a poorer recall of the stimulus name association. By contrast, if, as mentioned above, a novel word is spontaneously generalized to new instances of the same category even when they are displayed on new scenes, one should not predict any influence of these differences between training and recall situations.

Note that our view has important consequences on the notion of undergeneralization which is quite common in early lexical learning. It is defined as a lexical use in which children extend a novel word to a subset of the referents to which the adults extend the same word (e.g., the word "dog" used for small dogs only). The common explanation of these undergeneralizations is conceptual: preschool children fail to extend novel words because they do not understand that the novel entities belong to the same category as the referents to which they already extend the word. Within our memory hypothesis, we want to suggest that some undergeneralizations might not be conceptual but might result from poorer retrieval cues available when novel referents and/or novel scene settings are introduced, compared with the retrieval cues available when old stimuli and scenes are displayed.

In order to show that undergeneralization is not conceptually-based, one has to show that children extend novel words less well to referents they have never seen before than to the learning referents; while, at the same time, they give independent evidence that they *understand* that these novel stimuli belong to the same category as the learning stimuli. In our first experiment we compared performance for old referents (i.e., the ones shown in the training phase) and new referents displayed on their novel training scenes or on new test scenes at two delays: immediate and two weeks after training.

Experiment 1

Methods

Subjects: Fifty-two three-year-old children (mean age= 42 months, range: 36 to 47 months, 25 boys and 27 girls) participated in this study. They were normally developing, had no hearing impairments, and were acquiring French as their native language. Informed consent was obtained from their parents.

Materials

In the learning phase, four pictures of unfamiliar basic level categories (two categories of mammals – tapir and ibex -, and two categories of tools – plane and trowel) were used. Each referent was displayed on a different background scene (see Figure 1 for the two categories of mammals). In the transfer phase, participants were shown two referents from two categories of the learning phase (e.g., the old tapir and the old plane) (same instance condition) and two new instances of the remaining two categories (e.g., a new ibex and a new trowel) (different instance condition). To this purpose, a new instance of the four training categories was also selected (in books or on the Web). To summarize, there were two instances for each of the four categories, one used as training stimulus (the old instance), and the other as the novel instance. Each stimulus was displayed on a scene, selected in a set of eight different scenes, each scene being associated with one stimulus only. The scenes for tools (e.g., an indoor scene such as a table) were not the same as the ones used for mammals (e.g., a wood). For each tool and each mammal, two versions were constructed: each referent was associated (pasted on) with two different scenes. The stimuli were constructed with the Photoshop© software. In the training phase, one instance of each of the four basic level categories was selected randomly, that is embedded in one of the two scenes each referent was associated with. The four test stimuli were selected according to the scene condition in which a child participated: in the same scene condition, all the scenes were the scenes shown in the training phase, two scenes associated with two training instances, and two scenes associated with two novel instances of the two remaining categories. In the novel scene condition, two novel scenes were associated with two old instances and the two remaining novel scenes were associated with a novel instance of each of the two remaining categories. We also constructed another version of the material, with referents pasted on a white background. This version was used to check children's knowledge of the real name of the referents, prior to the training phase itself. Four novel words (non-words) were used as labels of the referents throughout the experiment: moupa, duban, togon, kéni. For a given child, each name was randomly associated with one of the four referents.

Procedure

Each child was tested individually at school, in a quiet room. The experiment was composed of two phases: a learning phase, and a testing phase, the latter being composed of two posttests, immediate, and after two-week delay.

Learning phase. Prior to the learning phase, the experimenter introduced the stimuli pasted on the white background and asked the child to give the name of the referents. There was no case in which the child knew the correct name. The learning phase itself was composed of four trials, so that all the children saw the referents and heard their names the same number of times. The task was introduced as a novel name learning task. The novel names were introduced as basic level names: "I'm going to show you animals and tools, and you will have to learn their name. Listen carefully because you will have to tell me their names, after". In each trial, the experimenter first showed one of the four stimuli and produced one of the four novel words: "this is a ….". Then, the child had to repeat the novel word, in order to assess his/her capacity to repeat it. When the child failed to repeat the name, the experimenter repeated it, until the child was able to repeat it correctly. The experimenter stopped these repetitions after four failures (which never happened since most children could repeat the word after the first presentation). The experimenter introduced the second stimulus in the same way, and so on with the third and fourth stimuli. Then, he introduced the four stimuli in a row and gave the name of the corresponding referent. Last, he showed the stimuli one at a time in a random order and asked the referent name. He gave a feedback for each production of the child (e.g., "yes, it is… ", "no, it's a … ". This sequence composed a trial. There were four such trials.

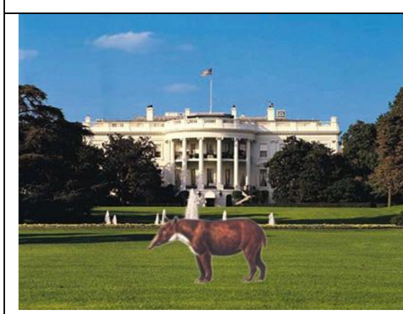
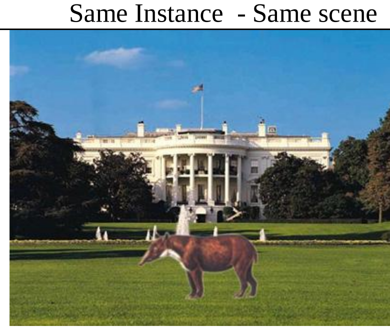
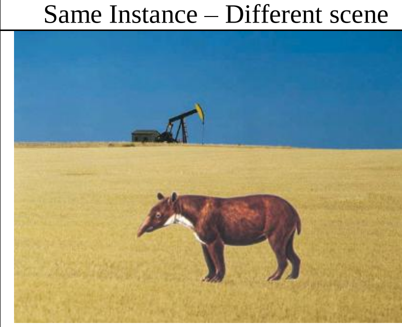
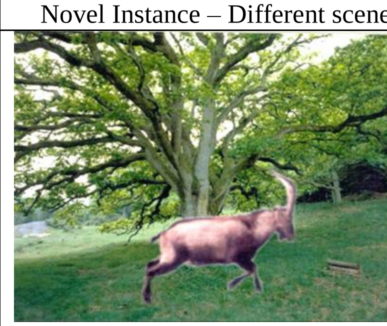
Learning Phase Stimuli	Post-test	Post-test
	Same Instance - Same scene	Same Instance – Different scene
		
	Novel Instance - Same scene	Novel Instance – Different scene
		

Figure 1: Examples of the stimuli and scenes used in Experiment 1: The left column displays training stimuli with the scene they were embedded in. The middle column and the rightmost columns display test stimuli.

Test phase. There were two post-tests: an immediate posttest which took place one minute after the last learning trial, and a second post-test two weeks after the learning phase. Each post-test was composed of a naming phase in which the four stimuli were introduced one by one, in a random order ("do you remember the name of this one?"). The naming task was followed by a pointing task in which the four stimuli were interspersed with three new stimuli that were never shown during the previous stages of the experiment and that acted as fillers. The child had to point to each stimulus corresponding to the name given by the experimenter ("could you show me the *duban*?"). The two conditions defining the variable instance (old vs. novel referents) differed by the referents introduced in the posttests: in the "old" condition, the experimenter showed the same instances as in the learning phase, whereas in the "novel" condition, novel instances were shown to the children (e.g., a tapir not introduced during the previous stages of the experiment).

At the end of the second post-test, participants were shown all the referents, that is to say, the "old" items and their "novel" counterparts. The experimenter showed the training stimuli and asked children to put the test stimuli with the corresponding training stimuli. This was done in order to verify that children were able to classify the novel instances in the same category as the old instances. This was necessary since we wanted to avoid, in the novel condition, naming or pointing failures that would be due to children's (conceptual) misunderstandings that novel instances were belonging to the same category as their learning counterparts.

Experimental Design

The variables Instance (old vs. novel), Task (pointing vs. naming) and Post-test (immediate vs. 2 weeks later) were within variables, while the variable Scene (same vs. different) was a between factor.

Results

Eight children were removed from the sample because they failed to correctly classify *all* the novel referents with their "old" counterparts. Children's performance was calculated in the following way. In the naming task, a correct word pronounced correctly and correctly associated with its referent was given 1 point. However, sometimes children productions were only partly correct. In that case, we calculated the number of correct phonemes in a given word. Each correct phoneme in the correct location in the word was rated .25 (e.g., "keno" for "kéni" would give a score of .75). This was necessary because the mean number of entire words correctly pronounced was very low. In the pointing task, each correct association between a name and a referent was scored 1.

We conducted two separate mixed three-way ANOVAs on the data for the naming task and the pointing task, with the variable scene (same vs. different) as a between factor, and the variables post-tests (immediate vs. 2 weeks later) and instances (old vs. novel) as within factors. Since there were only two old instances and two novel instances, the maximum score was 2 for each cell of the experimental design. The 2 x 2 x 2 analysis on the naming scores revealed a significant effect of the post-test variable, F (1,42) = 6.22, $p < 0.05$, with better performance for the immediate post-test (M = .35) than for the delayed post-test (M = .19). More importantly, there was a significant main effect of instance, F (1,42) = 4.82, $p < 0.05$, with better performance for the old instances (M = .35) than for novel instances (M = .17). No other effect or interaction reached significance.

The parallel ANOVA on the pointing scores also revealed a significant effect of instance, F (1,42) = 20.23 et $p < 0.001$, with better performance in the Old condition (M = 1.19) than in the Novel condition (M = 0.81). There was also a significant post-test x scene interaction, F (1,42) = 8.83 et $p < 0.005$, showing that the difference between the same scene and the different scene condition, in favour of the same scene condition, was important in the second post-test, not in the immediate post-test (see Figure 2).

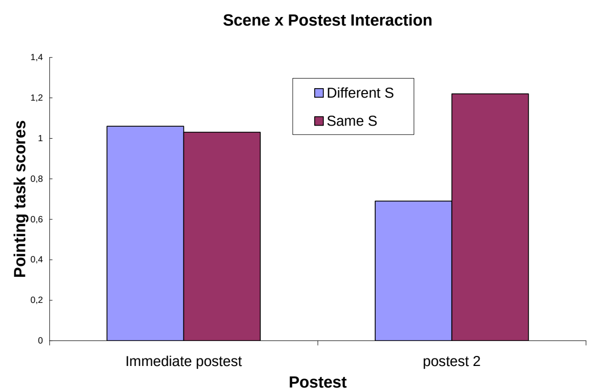


Figure 2: Scene x Post-test interaction in the pointing task

Discussion

In both the naming and the pointing tasks, performance was better in the old condition than in the novel instance condition. This result (which can be described as undergeneralization, that is a failure to apply a novel name to instances to which it could be applied) cannot be explained by conceptual factors since all the children kept in the analysis understood that the novel instances were to be classified with the corresponding old instances. Another interesting result was obtained for the variable scene, in the post-test x scene interaction, showing that participants' performance was weaker for the different scene condition and was not influenced by the scene congruence in the immediate test. Both results (on the variables scene and instances) are consistent with our retrieval cue hypothesis.

Note that we did not obtain a significant effect of scene in the naming task. This might be due to the very low scores or to the important variance in the results. A scene effect is particularly interesting because it reflects a lack of generalization of a novel which is certainly not due to conceptual factors, especially if the learning stimuli are the same in the training and the transfer phase. In the following experiment, we compared three conditions of scene congruence or incongruence between the training and the transfer phases, with the same instances used in both phases.

Experiment 2
Methods

Participants

A total of 113 french-speaking children participated in the study, (mean age = 52 months, range: 42 to 59 months). Children were tested in their school. (Note that the children were younger in this experiment than in the first one. In fact, Experiment 2 was conducted before Experiment 1, for which we decided to test younger children.) French was their native language. Informed consent was obtained from their parents.

Materials

Training phase stimuli. There were four experimental pictures representing real entities. Stimuli were coloured images (15 x 21 cm) of four unfamiliar animals or four unfamiliar musical instruments. Each stimulus was included in a scene. Four scenes highly contrasted were used for each category of stimuli. Animals were associated with a desert, snow, a park, or a house. Musical instruments were associated with a desk, in a room, on a seat, or in the street. A stimulus was pasted on a scene with the Photoshop© software. The four target stimuli of a category gave rise to several sets of stimuli, each set resulting from a different association between each of the four novel stimuli and a particular scene with the additional constrain that, in a given set, two stimuli could not be associated with the same scene.

Test phase stimuli. Three conditions were devised: same scene, mixed scene, new scene. In the *same* scene condition, a stimulus was associated with the same scene in the learning and the test phase. In the *mixed* scene condition, each stimulus swapped the scene it was associated with in the learning phase with the scene of one of the three remaining learning stimuli. In the *new* scene condition, eight new scenes were created for the test phase, four for the animals and four for musical instruments. In this condition, each learning stimulus was associated with one of the four new scenes.

Procedure

The training phase was similar to the training phase in Experiment 1 except that there were 5 trials instead of 4. Note that a child had to learn the name of four stimuli from the same category, either the four animals or the musical instruments. At the time of the experiment, we preferred to use two types of categories for the sake of generality. We also decided to separate the two types of categories in order to keep the conceptual heterogeneity of the stimuli for each child as low as possible. We agree that mixing the two types of stimuli would have been a possible strategy.

The test phase was composed of three post-tests. The first one took place one minute after the end of the learning phase, the second, two days later, and the third 2 weeks after the end of the learning phase. There were 3 post-tests here instead of 2 in the first experiment because, as mentioned above, this experiment was performed before Experiment 1 and at that time we had no idea of the effect of the length of the delay. In each post-test, the child had to perform the same naming and pointing tasks as in Experiment 1.

Results

We decided to split the group of participants into two subgroups of children, younger and older ones. The idea was to check whether potential scene effects would be equivalent in both age groups. Indeed, it is interesting to study whether younger children would be more influenced by irrelevant cues than older ones. Two separate analyses were performed on the naming and the pointing tasks. The analyses were mixed three-way-analyses of variance (ANOVA) 2 (Age: under 50 months vs. over 50 months) x 3 (Type of scene: same vs. mixed vs. new) x 3 (Time of post-test: immediate vs. 2 days vs. 2 weeks) with repeated measure on the factor Post-test variable.

For the naming task, the analysis of variance showed a significant effect of the variable Age, with older children obtaining better results than young children, $F(1, 107) = 6.36$, $p < 0.05$. Interestingly, there was a significant effect of the scene, $F(2, 107) = 3.479$, $p < 0.05$. A posteriori test (Tukey HSD) revealed that the “same scene” condition was only marginally significantly higher than the other two scenes (same scene, mean = 1.24, mixed scene, mean = .88, new scene, mean = .88). The post-test variable was significant, $F(2, 214) = 6.670$; $p < 0.01$. A posteriori test (Tukey HSD) showed that the first post-test was significantly lower than the other post-tests ($p < .05$; mean, first post-test = .81, second post-test = 1.06, third post-test = 1.13).

For the pointing task, there was also a significant effect of scene, $F(2, 107) = 8.351$; $p < 0.001$. The condition “same scene” was significantly higher than the two other conditions (Tukey HSD, $p < .05$) (same scene, mean = 2.96, mixed scene = 2.07, new scene = 2.40). No other effect reached significance.

Discussion

This experiment confirmed the role of the scene already demonstrated in the first experiment. The effect was present in both tasks and revealed better results for the “same scene” condition compared with the two other conditions, which is totally consistent with our retrieval cues hypothesis and is not compatible with a conceptual view of undergeneralization, since the stimuli were the same in the learning and the test phases. Note also the effect of age in the naming task, not in the pointing task. This interesting result suggests that the association between the name and the referent was equivalent in the two age groups. The difference in the naming task might reflect differences at the level of the production system and/or in the ability to retrieve the word when memory cues are poorer or a combination of these two variables.

General Discussion

Our purpose was to show that children might fail to generalize a novel word to new situations for non-conceptual reasons. The differences obtained between novel instances and old instances, in Experiment 1 the one hand, and between *same* scene and *new* scene in both experiments are important because they suggest that children did not use novel word for memory reasons. More precisely, when children did not use a novel word-referent association, this was not due to a failure to learn the association, or because they did not understand that the stimuli in the test phase belonged to the same category as the training stimuli. Our interpretation is that in a number of cases they encoded the word-referent association in terms of specific cues that were no longer or less available in the test phase.

Interestingly, our results suggest that a number of undergeneralizations described in previous contributions, especially those observed in natural communication situations, and that have been described as conceptually-based failures to generalize were probably due to the memory factors underlined here. Obviously, our claim is not that conceptual undergeneralizations do not exist: there are many cases in which children did not grasp the right dimensions leading to a correct generalization of the corresponding term in the adult language. Our claim is that, before assuming a conceptual undergeneralization, one should verify that children did not mention or understand that the novel stimuli did not belong to the same category as the corresponding learning stimulus. Moreover, it is quite clear that the scene effects observed in Experiment 2 cannot be described in conceptual terms.

More generally, in the lexical development literature, different views have been proposed. To summarize, on the one hand, authors have claimed that lexical learning depends on mechanisms that are specific to this task, either because these mechanisms are innately dedicated to lexical learning (Markman, 1989) or because children have learned productive regularities regarding lexemes in the course of their development (Smith, 1999). On the other hand, other authors view lexical learning as the product of general learning mechanisms, not specific to language, such as memory mechanisms or more conceptual mechanisms associated with the emergence of a theory of mind (Bloom, 2000). Our results, we think, are more consistent with this general mechanisms view of lexical learning. Indeed, our memory hypothesis is directly connected with the encoding specificity hypothesis (Tulving & Thomson, 1973) that has been applied to a wide number of encoding situations outside the language domain. Generally speaking, this view is consistent with the negative influence of non-congruent scenes observed in the present experiment.

Acknowledgments

This research also benefited from a grant from the “Fonds spéciaux de la recherche” of the University of Liège. We would like to thank Johanne Gregoire, Agnes Sadzot for assistance for data collection and analysis.

References

Baddeley, A.D. (1997). *Human Memory: Theory and Practice*. Psychology Press.

Balsam, P.D. & Tomie, A. (1985). Context and learning. Hillsdale, NJ. Lawrence Erlbaum Associates.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge: MA, MIT Press.

Carey, S. & Bartlett, E. (1978). Acquiring a single new word. *Papers and reports on child langage development, 15*, 17-29.

Childers, J. & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, *38*, 967-978.

Clark, E.V. (1993). *The lexicon in acquisition.* Cambridge : Cambridge University Press.

Fenson, L., Dale, P., Reznick, S., Bates, E., Thal, D., Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*, (no. 242).

Markman, E.M. (1989). Categorization and naming in children. Problems of induction. Cambridge, MA. Cambridge University Press.

Markson, L. and Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature 385*, 813-815.

Mervis, CB (1987). Child-basic object categories and early lexical development. IN U. Neisser (Ed.), Concepts and conceptual development. Cambridge: Cambridge University Press.

Mervis, C.B., & Bertrand, J. (1994). Acquisition of the Novel Name-Nameless Category (N3C) Principle *Child Development*, 65, 1646-1662.

Siegler, R.S. (1998). *Children's thinking*. Englewood Cliffs, NJ: Prentice Hall.

Smith (1999). Children's noun learning: How general learning processes make specialized learning mechanisms. In B. MacWhinney (Ed.). *The emergence of language* (pp. 277-303).

Tulving, E. & Thomson, D.W. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review*, *80*, 352-73,