

# Experimental methods for studying the mental representation of language

Uri Hasson and Rachel Giora

## 1. Introduction

One of the central assumptions underlying research in cognitive linguistics is that language use reflects conceptual structure, and that therefore the study of language can inform us of the mental structures on which language is based. One of the goals of the field is therefore to properly determine what sorts of mental representations are constructed by various sorts of linguistic utterances. Initial research in the field (e.g., Fauconnier 1994, 1997; Lakoff & Johnson 1980; Langacker 1987) was conducted by way of theoretical discussions, which were based on the methods of introspection and rational reasoning. These methods were used to examine diverse topics such as the mental representation of presupposition, negation, counterfactuals and metaphor, to name a few (cf. Fauconnier 1994).

Unfortunately, the observation of one's mental structures via introspection may be limited in its accuracy (e.g., Nisbett & Wilson 1977). As a result, investigators have come to realize that it is important to examine theoretical claims by using experimental methods (cf., Gibbs 2000; Grady 2000). In this chapter, we will discuss such experimental methods with two goals in mind. Our first goal is to provide the reader who does not specialize in psychology a basic understanding of experimental methods that are used to study the mental representations constructed during language comprehension. Our second goal is to provide examples of how these methods have been used to tackle questions that are of interest to cognitive linguists. Because cognitive linguistic inquiry is interested in the mental representation prompted by language, we restrict ourselves to the mental representations that are later products of the comprehension process. Roughly, these correspond to what is referred to as the interpretation of a certain statement or text. We therefore do not address those initial stages of comprehension, which involve phonological access and access to the mental lexicon. Furthermore, for the sake of brevity, we will mostly restrict the discussion to textual units no larger than a single sentence.

The methods that we will discuss are ones that are often used in psycholinguistic research. These are:

- a. Lexical decision and naming measures.
- b. Memory measures.
- c. Item recognition measures.
- d. Reading times.
- e. Self report measures.
- f. The effects of language comprehension on a subsequent task.

Each of these methods is based on observing an experimental measure to draw conclusions about the mental representations constructed by a certain linguistic unit. For instance, reading times are used to examine the difficulty associated with understanding texts, whereas memory measures look at people's recall to understand how these texts were interpreted. We therefore distinguish between the different research methods based on the dependent variables they focus on (e.g., reading times, proportion of memory errors). In the context of each experimental method we highlight the following:

- a. The rationale behind using the dependent variable.
- b. Examples of studies using the method.
- c. Further considerations regarding the use of the measure, which address advantages, disadvantages and practical problems involved with the use of the method.

## 2. Lexical decision and naming latencies

### 2.1 Rationale

In a lexical decision task, participants are presented with a certain letter string on the screen, and their task is to decide (e.g., by pressing a key) whether that letter string makes up a word or not. In a naming task, participants have to read out loud a word presented on the screen. The materials in lexical decision and naming studies typically consist of 50% real words, and 50% non-words (e.g., *flurp*). The rationale behind both tasks is that people should be faster to identify or pronounce a word to the extent that the concept denoted by that word was made accessible by previous context. Response latencies are therefore thought to reflect the mental accessibility of a word. For example, in a seminal study, Meyer and Schvaneveldt (1971) demonstrated that people were faster to decide that *butter* is a word in English when it was preceded by the word *bread* than when it was preceded by the word *nurse*. This kind of facilitation is often referred to as *priming*, and it is attributed to automatic spread of activation between related meanings in the mental lexicon. In the context of lexical decision and naming studies, the first term is often referred to as the *prime*, and the second term is referred to as the *target*. Facilitation is said to occur when response times to a target term are faster after an experimental prime than after an unrelated control prime. Inhibition (or suppression) is said to occur when response times to a target term are slower after the experimental prime than after the control prime.

From the perspective of language comprehension, there are two main factors that are often manipulated in lexical decision and naming tasks. These are (a) the type of prior context, and (b) the interval between the termination of the priming stimulus and the

presentation of the target stimulus (sometimes referred to as *interstimulus interval* or ISI). Manipulations of ISI are particularly informative when the research goal is to study the timeline by which meaning is constructed. Note that lexical decision and naming tasks are also useful for investigating which sorts of concepts are made accessible by priming units that are longer than a single word, such as sentences or entire paragraphs.

## 2.2 Examples

Blasko and Connine (1993) used a lexical decision task to examine the time course of metaphor comprehension. Specifically, they were interested to know how quickly people interpret metaphors, and whether the initial stages of metaphor comprehension involve the construction of a literal-related meaning, as implied by the standard pragmatic model of metaphor comprehension (e.g., Searle 1979). For example, they presented participants with a statement that contained the phrase *hard work is a ladder*, and participants then made a lexical decision either to a literally-related target (*rungs*), a metaphorically-related target (*advance*) or a control target (*pastry*). They found that after participants heard familiar metaphors, response latencies to both metaphorically-related and literally-related targets were facilitated. This finding was interpreted as showing that the comprehension of familiar metaphors does not require the construction of the literal meaning of the utterance before the metaphoric meaning is derived. However, for unfamiliar metaphors the results were different; here only literally-related targets were facilitated in the initial stages of comprehension.

The usefulness of varying the ISI in a lexical decision task can be seen in a study by Till, Mross, and Kintsch (1988), who investigated the time course of sense creation in discourse context. They presented participants with paragraphs that contained statements such as *The servant lit the fire and then prepared the meal. He seasoned it and wrapped it in foil*. The presentation of the paragraph was then immediately interrupted, and participants were given a lexical decision to a critical target word. In one condition, the target was lexically associated with a word that appeared in the sentence (e.g., *tin*), and in another condition, the target word reflected an inference from the statement (e.g., *barbecue*). By manipulating the interval between the termination of the sentence and the presentation of the lexical decision target, the authors found that lexically-associated targets were facilitated fairly early after reading the statements (~ 200 msec), but inferentially-related targets were facilitated only later on; about 500 msec after reading the statements.

Hasson and Glucksberg (2006) manipulated the interstimulus interval to see what sorts of meaning are prompted by negated utterances. In their study, participants were presented either with affirmative metaphors (e.g., *this lawyer is a shark*) or with their negation (*this lawyer is not a shark*). Participants pressed a spacebar after reading such affirmative or negative metaphors, and were then presented with a target that was related either to the affirmative meaning (e.g., *vicious*) or its negation (e.g., *gentle*). When the target words were presented 100 msec after reading the sentence, then targets related to the affirmative meaning were facilitated after both affirmative and negative metaphors. Targets related to the negative meaning were not facilitated in either case. In contrast, when the target words were presented 1000 msec after reading the sentence, then affirmative-related tar-

gets were facilitated after affirmative metaphors, but not after negative metaphors. Again, negative-related targets were not facilitated in either case. The findings show that the comprehension of negative metaphors involves initial activation of the affirmative meaning followed by a reduction of that activation. Giora et al. (2005) conducted a similar investigation of the comprehension of negated adjectival attributions. In their study, participants made lexical decisions to affirmative-related terms after reading either affirmative or negative adjectival attributions (e.g., *this instrument was / was not sharp* followed by the target, *pricking*), and the ISI was a short 100 msec. They found that response latencies to affirmative-related terms were equally fast after reading affirmative or negative adjectival attributions. This finding also suggests that in the initial stages of comprehension, negations are understood as affirmations.

### 2.3 Further considerations

When used properly, lexical decision and naming methods are useful for studying which sorts of representations are constructed by linguistic units. However, the validity of these measures (like any other measure) depends on the extent to which they reflect the theoretical constructs of interest and are unaffected by other factors. If performance on lexical decision and naming tasks reflected *just* the degree to which a linguistic context affects the accessibility of a certain concept, then these methods would display what is known as high construct validity. In this case, differences in lexical decision times could be unambiguously interpreted as indicating differences in the accessibility of concepts. However, if the performance on these tasks was shown to be susceptible to the influence of other factors, then the results may be of less theoretical interest, because in that case it would be unclear whether task performance actually reflects differences in concept accessibility.

Luckily, the literature on lexical decision and naming tasks has identified a number of exogenous factors that affect performance on these tasks, and so it is possible to circumvent many potential pitfalls (see Neely 1991, for a comprehensive summary of such factors). For example, lexical decision tasks are susceptible to 'expectancy' effects: if participants expect that a certain target will follow a certain prime, and that target fails to appear, then response latencies to the unexpected target will be slowed down. Participants may also wrongly answer 'No' to the lexical decision task when the target does not match the preceding context. Such expectation and backward-checking effects are a result of the decision component of the task and have been reviewed in the literature (see, e.g., Haberlandt 1994; Keenan, Potts, Golding, & Jennings 1990). Another problem is that at long ISI latencies, lexical decision tasks are sensitive to the proportion of pairs in which the prime and the target are related to each other: the greater the proportion of related prime-target pairs in a study, the larger the facilitation seen for the target (McKoon & Ratcliff 1995).

### 3. Memory measures

#### 3.1 Rationale

The logic underlying the study of language by looking at memory performance is that people's memory for a certain expression could indicate how that expression was mentally represented at the time it was committed to memory, or encoded. In particular, the errors that people make in a memory task may be indicative of the representations they constructed. The basic format of memory tasks consists of two stages: a *learning stage*, in which certain expressions are presented to participants, and a *test stage*, which evaluates participants' memory for those expressions. Between the learning and the test stage there is typically a break for a certain period of time, in which participants are engaged in an unrelated task. This break serves to clear working memory from the materials just encountered in the learning stage. The evaluation of memory during the test stage can be carried out by asking people to recall the items presented in the learning stage (free recall), or by presenting them with old and new test items and asking them to indicate for each item whether it had appeared in the learning stage (this is known as an old/new recognition task).

The recognition measure is particularly useful when the goal is to examine the degree to which different expressions are similar in meaning. The greater the similarity in meaning between two expressions, the more likely people are to mistake one for the other in a recognition test. For instance, a hypothesis might be that statement *a* is more similar in meaning to statement *b* than to statement *c*. To test this hypothesis, statement *a* might be presented in the learning stage of a study, and either statements *b* or *c* would be presented in the test stage. Ideally, participants should judge items *b* and *c* as 'new' items, i.e., as items that had not appeared in the learning stage. However, if the results show that item *b* is incorrectly judged as 'old' reliably more often than item *c*, then the results would support the hypothesis in question.

#### 3.2 Examples

The utility of the recognition measure is demonstrated in a classic study by Bransford, Barclay, and Franks (1972). In that study, participants studied sentences, and were later given a recognition task in which they were to be asked whether a certain sentence was presented in the learning stage. The critical materials in the study were ones in which a sentence presented in the learning stage and a sentence presented in the test stage either reflected the same situation in the world, or not. Take, for example, sentences (1) and (2), which describe the same situation in the world:

- (1) Three turtles rested on a floating log and a fish swam beneath them.
- (2) Three turtles rested on a floating log and a fish swam beneath it.

The authors found that when participants were presented with sentence (1) in the learning stage, they later tended to confuse it with sentence (2) in the test stage; i.e., they incorrectly judged that sentence (2) was presented in the learning stage. In contrast, when the

two sentences did not describe the same situation in the world, participants did not tend to confuse them. For example, participants who were presented with sentence (3) in the learning stage seldom mistook sentence (4) for it:

- (3) Three turtles rested beside a floating log and a fish swam beneath them.
- (4) Three turtles rested beside a floating log and a fish swam beneath it.

These results demonstrated that people are more likely to confuse two statements when the statements refer to the same situation in the world, and indicated that the mental representation of a statement is not *just* a representation of the propositional phrase-structure of the text.

A number of researchers (e.g., Fillenbaum 1966; Just & Carpenter 1976; Smith 1981) have used memory measures to tackle a long-standing question: does the mental representation of negation consist of a representation of what is denied? If comprehending negation involves the representation of the affirmative counterpart, then in the context of a memory task, people who study negated statements should wrongly mistake them for affirmations in the test stage, but the opposite mistake should occur less often. Smith (1981) presented participants with affirmative and negative statements (e.g., *the boy hit the girl* vs. *the boy did not hit the girl*). In the test stage, some of the affirmative statements were presented in negative form, and some of the negative statements were presented in affirmative form. Smith (1981) found that after people learned a negative statement they were more likely to report that they had learned its affirmative form than vice versa. Interesting, this happened for statements with high content of mental imagery (e.g., *the enormous elephant did not lift the fallen tree*), but not for abstract statements (e.g., *his greatest virtue was not his irrepressible confidence*). The fact that abstract negations were remembered better than concrete negations could suggest that the negation of a concrete sentence involved mental imagery of what was said not to be the case. This concrete imagery of the counterfactual state of affairs was later "read off" from memory and mistaken for the representation of an affirmative proposition.

Fillenbaum (1966) examined a related question: is negation always represented as affirmation, or does this depend on the type of concept negated? In Fillenbaum's (1966) study, participants were presented with statements in which adjectives were negated (e.g., *the man was not alive*, *the surface was not rough*). Some negated adjectives offered a direct implication (not alive  $\rightarrow$  dead), whereas some did not (not rough  $\rightarrow$  smooth?). In the test stage, participants were given a forced choice recognition test. For example, after being presented with the sentence *the postman is not alive* in the learning stage, they were asked to recognize which sentence they had seen earlier from the following options: *the postman is alive*, *not alive*, *dead*, *not dead*. The data of interest were the type of recognition errors that participants made. These fell into two categories:

- a. Negation drop: e.g., misrecognizing "not alive" as alive.
- b. Gist substitution: e.g., misrecognizing "not alive" as dead, or "not rough" as smooth.

Fillenbaum found that for dichotomous adjectives such as *alive*, participants made more substitution errors than negation-drop errors, but the opposite tendency was found for scalar adjectives such as *rough*. These results suggested that people's encoding of negated

propositions may depend on the inference that is afforded by the negation, and that negation is not always encoded as the falsity of the affirmative proposition (cf. Clark & Chase 1972). Fillenbaum's results have been recently extended in two studies using different materials and methods (Mayo, Schul, & Burnstein 2004; Hasson, Simmons, & Todorov 2005).

Sloutsky and Goldvarg (2004) employed memory measures to examine the representation of statements containing logical connectives. Specifically, they were interested to know whether people represent conditional statements of the general form *if p then q* as (more simple) conjunctions (i.e., *p and q*). In the learning stage of their study, they presented participants with statements based on conjunctions, disjunctions, or conditional connectives. These statements were based on abstract contents, e.g., *if he takes medicine, then he likes the zoo*. In the following test stage, they presented participants with old and new statements. The crucial statements in the recognition stage were new statements that differed from an original sentence only in the logical connective. They found that after studying a conditional statement, participants were more likely to confuse that conditional with a conjunction form (e.g., *he takes medicine and he likes the zoo*) than to confuse the conditional with a disjunction form (*he takes medicine or he likes the zoo*). The findings suggested that people might build a 'minimal' representation of conditional statements: although a conditional is logically consistent with three sorts of states of affairs, people seem to represent that possibility in which both the antecedent and consequent are true, and therefore later confuse between conjunctions and disjunctions.

### 3.3 Further considerations

Though memory measures often provide useful information for evaluating certain hypotheses, there are certain methodological weaknesses that are inherent in such methods. As a result, it is best to use such measures in combination with other methods targeting the questions of interest (this is generally true).

One weakness is that memory performance on a recall or recognition task does not depend solely on the way in which a certain linguistic element was encoded during the learning stage, but also on various processes that occur during the memory test stage itself. This sort of weakness is inherent in various measures that are conducted "offline", that is, after participants had finished comprehending the stimuli. Consider for example Smith's (1981) finding that people are more likely to mis-recall negation as affirmation than vice versa. This finding was interpreted as showing that negation is sometimes represented and encoded as affirmation. However, the finding might also be a result of a certain response bias that is manifested in the test stage: when participants are unsure whether a sentence was presented in affirmative or negative form, they might answer, for whatever reason, that the statement was originally presented in affirmative form.

Also, the instructions given to participants in the learning stage may affect how participants understand the statements, and consequently, their performance in the test stage. In particular, participants may use specific comprehension and encoding strategies when they are told to memorize the materials presented in the learning stage, and these strate-

gies may be very different from those used in routine comprehension (cf. Clark & Clark 1977: Ch. 2).

Finally, one practical consideration is that materials presented at the beginning and end of the learning stage are often recalled better than items in the middle of the list (Brown 1958; Peterson & Peterson 1959). For this reason, it is recommended that the first items and last items presented in the learning stage be “filler items” – i.e., materials that will not be analyzed. Also, it is advisable to randomize the order of presentation of the items of interest in both the learning and test stages.

## 4. Item recognition measures

### 4.1 Rationale

This method, like lexical decision and naming tasks, seeks to establish the relative accessibility of a certain term following text comprehension. In practice, participants read a text, and immediately after reading that text (typically, within 1 or 2 seconds), they are asked to decide whether a term had appeared in the text they had just read. The latency to make the decision is taken to reflect the relative accessibility of the term. More generally, it is assumed that different levels of accessibility reflect levels of activation for the concept referred to by the term.

### 4.2 Examples

MacDonald and Just (1989) used the recognition method to examine whether the representation of negation differs from affirmation. Participants were presented with statements that referred to two entities, one of which was negated (e.g., Elizabeth baked *some bread*, but *no cookies*). Participants read such statements at their own pace, and after each statement they were presented with a term on the screen, and had to verify whether it had appeared in the statement or not. In the crucial trials, the terms presented for verification corresponded either to the negated term (e.g., *cookies*) or to the non-negated term (e.g., *bread*) in the sentence just read. MacDonald and Just (1989) found that verification times were slower for negated terms than for non-negated terms suggesting that, “negation decreases accessibility of a negated noun” (p. 641).

The same procedure was used by Kaup (2001) to examine a different hypothesis: that the mental representation of negation depends on whether the negated term refers to an entity that is present or absent from the situation described in the sentence. Kaup (2001) presented participants with statements such as (5) and (6):

- (5) Almost every weekend, Mary bakes *some bread* but *no cookies* for the children.
- (6) Elizabeth tidied up her drawers. She burned the old *letters* but *not the photographs*.

Note that in (5), the negation implies absence of cookies from the scene, whereas in (6), the negated term is implied as present while the affirmative one is absent. The results showed that, on the whole, verification latencies were slower for negated terms than for

affirmative terms. However, this difference between verification latencies for the negated and non-negated terms depended on whether the negated terms referred to items absent or present in the scene. The difference between response latencies to negated and non-negated nouns was largest when the negated term was absent from the scene and the non-negated term was present in the scene (example 5). When the negated term was present in the scene and the non-negated noun absent (example 6), then the difference between verification latencies for affirmative and negated terms was smaller. Such findings indicate that negation reduces the accessibility of terms in its scope, but they also highlight the importance of a second factor, viz., the presence or absence of a given entity in the situation described.

The item verification method was also used in a study conducted by Glenberg, Meyer, and Lindem (1987), which examined whether elements that are foregrounded in a text are more cognitively accessible. In their study, participants read a story in which a certain element was either foregrounded or not. For example, one story described a person preparing for a marathon, who *takes off* his shirt before jogging around the lake. A different version of this story was constructed, in which a person is described as *putting on* the sweatshirt before jogging around the lake. After reading the story, participants were asked whether a certain word, e.g., *sweatshirt*, had appeared in the story. The authors found that (in certain conditions), people were faster to verify the appearance of a term in the text when that term corresponded to a foregrounded entity than when the term corresponded to a non-foregrounded entity. They argued that these findings support the notion that people build mental-model representations of the discourse they are reading – a representation of what the text is about (for alternative explanations, see McKoon & Ratcliff 1992).

A study by Horton and Rapp (2003) makes a similar point. They presented participants with stories in which a certain object was implied to be occluded from the protagonist, or not. For example, the protagonist might be described as observing a vase, and later on, a certain action was performed that may or may not have resulted in the occlusion of the vase from the protagonist. After reading these stories, participants were asked whether a vase was mentioned in the scene just described. Participants were slower to verify that the object had appeared in the scene when the object was described as being occluded from the viewpoint of the protagonist than when the object was not implied to be occluded.

Finally, item verification may also be used to study inferences prompted by the text. In such studies, researchers may be interested in the speed by which people accurately determine that a certain word did not appear in the preceding text. For example, people might take longer to decide that the word 'hammer' did not appear in sentence (7) than to decide that it did not appear in (8), because the term is implied more strongly by the former:

- (7) He pounded the nails into the wood.
- (8) He bought some nails and wood.

### 4.3 Further considerations

As reviewed, item verification measures are useful for examining hypotheses concerning the degree of accessibility of different terms appearing in a text. Their relative drawback is that it is sometimes unclear whether differences in verification latencies necessarily indicate differences in the mental activation of concepts. Take, for example, the findings by MacDonald and Just (1989) and Kaup (2001), which showed that verification latencies for a term were longer when it was in the scope of negation. These findings were interpreted as showing that negation reduces the activation level of concepts. However, the increased verification latencies are also consistent with the possibility that negation prompts the construction of two mental spaces: a factual and a counterfactual one, and that this more complex representation of negated sentences leads to longer verification latencies for the negated items, because those items are represented in two spaces. It is therefore useful to corroborate the results of such studies by other measures of accessibility, such as lexical decision tasks.

Item recognition, like lexical decisions, involves a decision stage which is non-automatic. Because a decision is involved, the responses on such tasks could be biased by backward-checking heuristics of the sorts that also affect lexical decision tasks. For example, participants may check for the compatibility of the target item with the previous context and this could affect verification times. It is therefore suggested that there be a short latency between the termination of the sentence and the presentation of the target term (McKoon & Ratcliff 1986) to minimize strategic processes.

## 5. Reading times

### 5.1 Rationale

In some experiments, the time needed to read a text is taken to be indicative of the processing difficulty that is associated with the comprehension of that text. Reading times are particularly useful for examining whether preceding contexts have differential effects on text comprehension. In such designs, a given text appears in different experimental conditions that vary in the type of context that precedes the text. Reading times are also used to compare the processing difficulty of different expressions, but in such cases it may be more difficult to interpret any differences between experimental conditions.

### 5.2 Examples

Gentner, Bowdle, Wolff, and Boronat (2001) report a study that examined whether the comprehension of metaphorical statements activates the source domain of the metaphor. They constructed two sorts of experimental conditions. Both conditions ended with the exact same statement, e.g., a statement drawn from the mapping A DEBATE IS A RACE. In one condition (the consistent condition), the preceding statements in the paragraph were instances of the same cross-domain mapping. In the other condition (the inconsistent

condition), previous statements were instances of a different mapping, e.g., A DEBATE IS A WAR. The authors found that the last statement in the paragraph was read faster in the consistent condition than in the inconsistent condition (see also, Allbritton, McKoon, & Gerrig 1995). These findings suggest that when people read metaphorical statements, they activate rich conceptual structures that correspond to the mappings between the source and target domains, and the accessibility of this knowledge assists the comprehension of later statements based on the same mapping.

Unfortunately, differences in reading times are sometimes more difficult to interpret. Johnson (1996) examined comprehension times for metaphorical statements and similes. Johnson argued that if metaphors are simply disguised similes, then metaphors should be understood just as fast as similes, or somewhat slower than similes. Participants read similes and metaphors at their own pace, and it was found that similes were read more slowly than metaphors (a control condition demonstrated that this was not due to similes containing an additional word). Johnson (1996) interpreted the findings as suggesting that similes are mentally transformed into class inclusion statements and therefore take longer to comprehend. However, the data are also consistent with the interpretation that metaphors and similes are understood by different mental processes which lead to distinct mental representations. Comprehension latencies alone cannot decide between these interpretations.

In a more elaborate study of this topic, Gentner and Bowdle (2001) report an experiment in which they manipulated the sort of source domain that appeared in the metaphors and similes so that in some cases the source was novel (e.g., a novel is / is like a *glacier*), whereas in other cases the source had a conventionalized sense (e.g., a gene is / is like a *blueprint*). They found that when sources were novel, then similes were comprehended faster than metaphors, but when sources were conventionalized, then metaphors were comprehended faster than similes. They interpreted the findings as showing that when sources are novel, both tropes are understood as comparisons. Therefore similes have an advantage, as they indicate the comparison directly. But why were opposite results found for conventionalized sources? Gentner and Bowdle suggested that (a) when sources have a conventional meaning, then metaphors are understood as categorization statements, whereas similes are understood as comparisons, and (b) that categorization statements are inherently easier to comprehend than comparisons. As can be seen from this discussion, reading times alone are weak constraints on theoretical accounts. Therefore, such data may not be sufficiently informative if the purpose is to make specific claims about the on-line construction of mental representation. In such cases they are useful when considered alongside other sorts of evidence.

Reading times are more useful when there is an a priori hypothesis that one sort of statement is more difficult to process than another. McElree and colleagues (McElree, Traxler, Pickering, Seely, & Jackendoff 2001; Traxler, Pickering, & McElree 2002) have examined whether sentential contexts that are thought to involve type-shifting (9) take longer to understand than those that do not demand shifting (10).

(9) The author was starting the book in his house on the island.

(10) The author was writing the book in his house on the island.

Verbs like “starting” typically require an activity as a complement, but in certain cases they can be followed by a noun. In such cases, the noun phrase has to be interpreted as referring to an action (e.g., writing the book). In this study, the sentences were presented one word at a time, in a self-paced manner, and each statement was followed by a comprehension question to ensure that participants read for comprehension. The authors found that the noun (e.g., *book*) was read more slowly in they type-shifting context (9) than in the non-shift context (10), and the same held for the word following that noun. These results are consistent with the idea that type-shifting contexts involve further psychological elaboration of the noun phrase.

Note that there are alternative explanations for these data. For instance, the phrase “writing the” could result in an expectation for the word “book”, whereas the phrase “starting the” would not. If so, the findings would not reflect type shifting, but differential predictability of the critical noun in the two sentential contexts. For this reason, the authors took great care to eliminate such alternative explanations for the data. For instance, they verified that the sentences that were used in the different conditions were equally *plausible*. They also verified that the verbs used in these sentences were equally *frequent* in use. Finally, they verified that the noun phrase (e.g., *the book*) was not better predicted by one verb than the other.

Black, Turner, and Bower (1979) observed reading times to see whether readers are aware of points of view established in narratives. They constructed statements that were based on deictic verbs, so that in some cases a consistent perspective was maintained throughout the sentence (e.g., 11), and in other cases there was a change in perspective in mid sentence (e.g., 12).

- (11) Bill was sitting in the living room reading the paper when John *came* into the living room.
- (12) Bill was sitting in the living room reading the paper when John *went* into the living room.

Participants took longer to read those sentences that involved a change in perspective. These results were interpreted as showing that readers are sensitive to narrative perspective and prefer consistent viewpoints.

### 5.3 Further considerations

Reading-time measures are a relatively precise measure, and one that is easily implemented in a lab. Depending on the goal of the study, an experimenter might want to present the materials one sentence at a time or one word at a time on the computer screen. The rate of presentation may be predetermined by the experimenter, or self-paced by the participants in the study. Presenting the sentences one word at a time can help identify points of difficulty in the reading of the sentence, but has the drawback of making the task less natural than normal reading. In particular, when the words are presented one at a time, readers cannot go back to parts of the sentence they had already read, and cannot look ahead at upcoming words – both of which are possible in normal reading. These might be some of the reasons for why reading times for individual words are typically much longer than when those words appear in the context of a sentence (see Haberlandt 1994, for a review).

## 6. Self report measures: Listing features, choosing features

### 6.1 Rationale

In a feature-listing task, participants are typically asked to write down properties that best capture the meaning of a certain expression. For instance, they might be asked to write down (or choose from a list) which properties are implied by a metaphorical statement such as *this lawyer is a shark*. In another variant of the task, they might be asked to choose from a list which property (or paraphrase) best captures the meaning of the expression. The rationale behind the method is that if people can understand the meaning of an expression, then they can also explain what that expression means and which features are associated with it. This method is often used to examine how linguistic contexts affect sense generation.

### 6.2 Examples

Interesting examples of the use of features-listing are found in studies of noun-noun combinations. Typically, such studies aim to examine the relation between the meaning of the constituents of the combination and the meaning of the combination itself. For instance, Wilkenfeld and Ward (2001) examined to what extent the properties of a noun-noun combination (e.g., *motorcycle-carpet*) reflect the meaning of the constituents of that combination. In their study, one group of participants wrote down salient features of the individual constituent nouns (e.g., *motorcycle*, *carpet*), whereas a different group of participants defined the combination, and then wrote which features were associated with the definition. For instance, one definition of *motorcycle-carpet* was “a field of thousands of motorcycles”, and a feature listed for this definition was *crowded*. This design allowed the authors to examine the extent to which features of the individual constituents overlap with those properties associated with the combination. They were particularly interested to see whether combinations can imply features that are not directly reflected in their constituents – i.e., emergent features. The authors operationalized an emergent feature as one that was listed for a combination, but not for either of the individual constituents alone. They found that when the terms in the combination were dissimilar (e.g., as in the combination *couch-skate*) then 28% of the features listed were emergent. When the terms were similar (as in *zebra-horse*), 21% of the features were emergent. The analysis in this study was based simply on whether or not a feature of the combination was mentioned for its constituents.

A more detailed analysis of this issue was undertaken by Hampton (1987), who examined whether important features of the constituents necessarily end up being important features of the combination. In one experiment, participants either listed features for individual terms (e.g., *sports*, *games*) or for conjunctive combinations of those terms (e.g., *sports that are also games*). This study revealed that some of the combination features were emergent; i.e., were not listed for either of the constituents. In a second study, Hampton asked participants to rank order the importance of the features for the definitions of concepts. The ratings were made for individual terms and for combinations of the terms.

Features that were rated as important for constituents were usually rated as important for the conjunctions of those concepts. However, there were a few cases (less than 10%) in which features that were unimportant for the constituents tended to be rated as important for their conjunction. For example, the properties *small* and *lives in a cage* were rated as important for the conjunction *pets that are also birds*, but as unimportant for *pets* or *birds* separately.

Johnson and Keil (2000) also employed a feature-listing task to study the interpretation of noun-noun combinations. They used combinations that were interpretable but not completely familiar (e.g., *hospital-rat*, or *mountain-knife*). In their study, they asked participants to write thirty-six properties characteristic of each noun, and five properties characteristic of each combination. From this corpus, they constructed two master lists: one that contained properties listed for the nouns, and one that contained properties listed for the combinations. Note that comparing these two lists is sufficient for evaluating the prominence of the noun-features in the combination. However, before they analyzed the overlap between the lists, Johnson and Keil (2000) first derived two reduced lists from these lists. These reduced lists included only those properties that were mentioned by at least half of the participants for each of the nouns or for each of the combinations. These reduced lists therefore captured features that may be considered as more typical. They found that only 32% of the typical features of the combinations were also listed as typical features of the head noun. Thirty percent appeared as non typical features of the head noun, and 38% were not mentioned as features of the head noun (e.g., the property *stationary* was listed for the combination *hospital-bicycle* but not for *bicycle*).

Feature listing can also be used to examine how different sentential contexts mediate the accessibility of noun-features. Coulson and Matlock (2001) examined the meanings associated with words that appeared in different sentential contexts. In the null context, a term (e.g., *anchor*) was presented alone, and participants listed properties associated with the term. In addition, there were three experimental conditions, in which the terms were embedded in different sentential contexts. These contexts made up either a literal context (e.g., *he almost forgot about the anchor*), a metaphorical context (*his wife was his anchor*), or a literal-mapping context (*we were able to use a barbell for an anchor*). When appearing in these sentences, the terms of interest appeared as the terminal words of the sentence. Participants read each sentence and then quickly listed two or three features for the term of interest, which was underlined. Coulson and Matlock (2001) found that, on the whole, a significant proportion of the features listed in the sentential contexts (~40%) was not listed for these terms in the null context condition. The authors then went on to analyze which set of sentential features was most *similar in meaning* to those given in the null context condition. To this end they used a *latent semantic analysis* method (Landauer, Foltz, & Laham 1998). They found that the features given in the metaphorical context were the least similar in meaning to those given in the null context condition, and that the features elicited in the literal context were most similar in meaning to that of the null context. A feature-listing task of this sort was also used to study if interpretation of metaphorical statements generates emergent features (Becker 1997).

In another variant of this task, participants might be asked to choose which properties or paraphrase best capture the meaning of an expression from a list of pre-constructed

options. For example, Costello and Keane (2001) examined noun-noun combinations to determine which properties of the modifier are attributed to the head noun (e.g., *moth*) in combinations such as *bumblebee-moth*. They examined four sorts of possible properties, which varied on two dimensions: whether the property was related to a common dimension of the categories, and whether the property was diagnostic of the modifier. For example, for the combination *bumblebee-moth* these four options were:

- a. A moth that is black and yellow (this property is related to a common dimension; color, and color is diagnostic of bees).
- b. A moth that is the size of a bumblebee (this property is related to a common dimension; size, and size is non-diagnostic of bees).
- c. A moth that stings (this property is unrelated to a common dimension, but having a sting is diagnostic of bees).
- d. A moth that fertilizes plants (this property is unrelated to a common dimension, and fertilizing plants is non-diagnostic of bees).

They found that participants preferred interpretations that were based on diagnostic properties, as in example *c* above (68% of the interpretations). Whether or not the property was related to a common dimension had no effect on participants' choices. The same results were found when participants wrote down their definitions, and these definitions were then categorized into the four sorts of interpretations outlined above.

### 6.3 Further considerations

Feature-listing studies are easy to conduct, and often provide intriguing data. However, their downside is that they are based on probing people's conscious impressions of the meaning of an expression, which might not accurately reflect mental representation. Furthermore, by the time that people list the features of an expression, its initial interpretation may no longer be accessible. In fact, there is no way of knowing whether the features that people list capture the direct meaning of the expression, or people's implicit theories about the meaning of such expressions. Take for example the study by Costello and Keane (2001). The authors assumed that people's choices in that task reflect direct and unmediated access to the interpretation of the expression. However, people might hold lay theories about what meanings are associated with such expressions (theories that might be quite similar to those suggested by the authors!), therefore obtaining the results found in the study.

## 7. The effects of comprehension on subsequent tasks

### 7.1 Rationale

In certain cases, it is possible to examine how people interpret language by observing aspects of their behavior in a subsequent context. The rationale behind the method is that to the extent that a linguistic expression invokes a mental representation, that mental representation can be studied by observing the effect it has on subsequent behavior.

These sorts of paradigms are often called “priming” paradigms. Often such designs are employed to examine how the comprehension of one linguistic expression affects the processing of another expression (for example, whether the reading of the word *doctor* speeds up the subsequent reading of the word *nurse*, Meyer & Schvaneveldt 1971). We have discussed such cases in our discussion of lexical decision and naming tasks. In other cases, a researcher might be interested to find out how comprehension affects non-linguistic behavior. The advantage of studying the effects of comprehension on non-linguistic behavior lies in that participants are unaware that their input may reflect a product language comprehension, and so there is less risk that the results are due to strategic thinking on behalf of participants. Furthermore, such methods are sometimes the only way to answer specific research questions. This section focuses on such cases.

## 7.2 Examples

Estes (2003) used an indirect measure of comprehension to study how people understand two sorts of noun-noun combinations. Some theories (e.g., Gagne 2001) argue that the comprehension of noun-noun combinations is based on finding the relation that links the modifier and the head. It is assumed that there exist many sorts of such relations; for example, the relation “is like” could mediate the comprehension of the *attributive* combination *cactus-carpet*, where a property of cactus is attributed to carpet. A different relation would mediate the comprehension of the *relational* combination *pancake-spatula*. Arguing against this view, Estes (2003) suggested that when people understand attributive combinations, they engage in comprehension procedures that are qualitatively distinct from those used to comprehend relational combinations. In his study, he first asked participants to define (i.e., write the meanings of) attributive and relational combinations. Participants were then asked to judge the similarity of the terms in the combination; e.g., “how similar are cacti and carpets?”. These similarity ratings were compared to the ratings given by a control group that had not defined the combinations beforehand. The results showed that after participants defined attributive combinations, the terms in the combination seemed less similar than they seemed in the control condition. However, after defining relational combinations, the similarity of those terms was rated as greater than it had been in the control condition. The results suggested that the meaning of attributive combination might be arrived at via a comparison process, which highlights differences, whereas relational combinations are understood in a different manner.

The word fragment completion task is another task that can reveal which sorts of meanings are accessible. In this task, participants are typically presented with certain materials in the first stage of a study, and are later asked to complete a fragmented word (b-tt-r) with the first word they can think of (*bitter*, *butter*, *better*, *bettor*). A number of studies using this task have shown that the semantic processing employed in the first stage of the study, e.g., reading behavioral descriptions or a short story, results in indirect priming that is evident in people’s word completions (e.g., Richards & French 1991; Whitney, Waring, & Zingmark 1992).

For instance, Giora and Fein (1999) presented participants with short stories that biased a literal or ironic interpretation of the last sentence in the story. After reading these

vignettes, participants were asked to complete word-fragments related to either the literal or the ironic meaning of the targets. For example, the sentence *Moshe, I think you should eat something*, was embedded as the terminal sentence in either an ironic context (13a) or literal context (13b):

- (13) a. After he had finished eating pizza, falafel, ice-cream, wafers and half of the cream cake his mother had baked for his brother Benjamin's birthday party, Moshe started eating coated peanuts. His mother said to him: "Moshe, I think you should eat something".
- b. At two o'clock in the afternoon, Moshe started doing his homework and getting prepared for his Bible test. When his mother came home from work at eight p.m., Moshe was still seated at his desk, looking pale. His mother said to him: "Moshe, I think you should eat something".

Findings showed that ironically biased contexts (13a) facilitated both contextually compatible ironic responses as well as contextually incompatible-but-salient responses (related to the literal meaning of the irony). In contrast, literally biased targets (13b) facilitated only the salient, literally related concepts. Such findings showed that salient meanings were retained even when incompatible.

Another variant of this method is one where researchers examine how being exposed to a certain stimuli later affects people's behavior on a task that is ostensibly completely unrelated. This variant capitalizes on the fact that people's recent experiences affect how they comprehend ambiguous or neutral stimuli. In a classic study, Higgins, Rholes, and Jones (1977) presented participants with a description of a person, whose behavior was judged by participants in a control group to be mid-way between adventurous and reckless on a bipolar scale. Two other groups of participants also judged the behavior of the character on the basis of the same description, but only after being earlier exposed to certain words in the guise of an ostensibly unrelated 'perception study'. One of these groups was exposed to words related to recklessness, whereas the other had been exposed to words related to adventurousness. Higgins, Rholes, and Jones (1977) found that those participants that had been exposed to words related to recklessness later judged the character to be more reckless than adventurous, whereas participants that were exposed to words related to adventurousness later judged the character as more adventurous than reckless. These findings indicated that when people try to interpret an ambiguous meaning, they may be affected by cognitive constructs that have recently been made accessible.

McGlone and Harding (1998) used the disambiguation rationale to study whether people represent ego-moving and time-moving perspectives when comprehending temporal expressions. They presented participants with texts that contained ego-moving expressions (e.g., *we will arrive at the exam date in two days*) or time-moving expressions (e.g., *the exam date will arrive in two days*). Following, they asked the participants to disambiguate an ambiguous temporal expression; e.g., *if the meeting scheduled for next Wednesday has been moved forward two days, to which day has it moved?* This question is ambiguous, because the answer can be either Monday or Friday. Participants who had previously read ego-moving expressions tended to answer Friday (60% of responses), whereas those who had read time-moving expressions tended to answer Monday (69% of responses).

In a related study, Boroditsky and Ramscar (2002) examined whether thinking about movement through space influences how people interpret temporal expressions. They asked one group of participants to think of moving towards a certain point (which was in front of them), and asked another group of participants to think of moving a certain object towards them. Later, they presented these participants with the ambiguous question regarding Wednesday's meeting. Those participants who were earlier asked to think about moving forward tended to answer Friday (57% of responses), whereas participants who were asked to think of an object moving towards them tended to answer Monday (67% of responses).

### 7.3 Further considerations

The benefit of such methods is that they can demonstrate what constructs are cognitively accessible. However, the interpretation of such findings should be done with care. Specifically, an experimental finding demonstrating that thinking about one domain affects how people think about or behave towards another does not necessarily indicate that the two domains are based on common representational systems. For instance, Bargh and colleagues (Bargh & Chartrand 1999; Bargh, Chen, & Burrows 1996) found that participants that were shown films of elderly people later left the screening room more slowly than participants who had not seen such films. That is, thinking about the elderly caused people to walk more slowly. Clearly this does not mean that the dimensions of motion and age share a common representation. Instead, it shows that thinking about age may make certain constructs accessible (such as slowness), which can, in turn, affect behavior.

## 8. Summary

We have discussed some of the most common experimental methods used to study the interpretation of linguistic expressions in experimental settings. Some methods, such as lexical decisions, naming, and reading times rely on measuring behavior that occurs when participants are actively engaged with the experimental material. They are often referred to as "online measures" and are thought to index processes that are relatively low level and automatic rather than strategic ones. Other methods, such as those that rely on memory, fragment completion, or feature listing are "offline measures" as they rely on information (i.e., the dependent measure) that is collected well after the comprehension of the materials has been completed. These measures are therefore more susceptible to the working of strategic considerations employed by participants in the study.

Because our purpose was to introduce the reader to the different procedures, we did not elaborate on the more advanced technical considerations involved in each method. For those interested, Neely (1991) offers a comprehensive review of the different factors that affect lexical decision and naming tasks, Haberlandt (1994) reviews in detail different factors that affect performance on reading time and item-verification measures, and Keenan, Potts, Golding and Jennings (1990) provide a detailed review of different methodologies used to study inferences during reading.

## References

- Allbritton, D. W., McKoon, G., & Gerrig, R. J. (1995). Metaphor-based schemas and text representations: Making connections through conceptual metaphors. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21(3), 612–625.
- Bargh, J. A. & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, 54(7), 462–479.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality & Social Psychology*, 71(2), 230–244.
- Becker, A. H. (1997). Emergent and common features influence metaphor interpretation. *Metaphor & Symbol*, 12(4), 243–259.
- Black, J. B., Turner, T. J., & Bower, G. H. (1979). Point of view in narrative comprehension, memory, and production. *Journal of Verbal Learning & Verbal Behavior*, 18(2), 187–198.
- Blasko, D. G. & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 19(2), 295–308.
- Boroditsky, L. & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, 13(2), 185–189.
- Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 3(2), 193–209.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 10, 12–21.
- Clark, H. H. & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, 3, 472–517.
- Clark, H. H. & Clark, E. V. (1977). *Psychology and language. An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.
- Costello, F. J. & Keane, M. T. (2001). Testing two theories of conceptual combination: Alignment versus diagnosticity in the comprehension and production of combined concepts. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27(1), 255–271.
- Coulson, S. & Matlock, T. (2001). Metaphor and the space structuring model. *Metaphor & Symbol*, 16(3–4), 295–316.
- Estes, Z. (2003). A tale of two similarities: Comparison and integration in conceptual combination. *Cognitive Science*, 27(6), 911–921.
- Fauconnier, G. (1994). *Mental spaces: Aspects of meaning, construction in natural language*. New York, NY: Cambridge University Press.
- Fauconnier, G. (1997). *Mappings in thought and language*. Cambridge, U.K. New York, NY: Cambridge University Press.
- Fillenbaum, S. (1966). Memory for gist: Some relevant variables. *Language and Speech*, 9, 217–227.
- Gagne, C. L. (2001). Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27(1), 236–254.
- Gentner, D. & Bowdle, B. F. (2001). Convention, form, and figurative language processing. *Metaphor & Symbol*, 16(3), 223–247.
- Gentner, D., Bowdle, B. F., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. In D. Gentner & K. J. Holyoak (Eds.), *Analogical mind: Perspectives from cognitive science* (pp. 199–253). The MIT Press.
- Gibbs, R. W. (2000). Making good psychology out of blending theory. *Cognitive Linguistics*, 11(3/4), 347–358.
- Giora, R., Balaban, N., Fein, O., & Alkabetz, I. (2005). Negation as positivity in disguise. In H. L. Colston & A. Katz (Eds.), *Figurative language comprehension: Social and cultural influences* (pp. 233–258). Hillsdale, NJ: Erlbaum.

- Giora, R & Fein, O. (1999). Irony comprehension: The graded salience hypothesis. *Humor*, 12(4), 425–436.
- Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory & Language*, 26(1), 69–83.
- Grady, J. (2000). Cognitive mechanisms of conceptual integration. *Cognitive Linguistics*, 11(3/4), 335–345.
- Haberlandt, K. (1994). Methods in reading research. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1–31). San Diego, CA: Academic Press, Inc.
- Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition*, 15(1), 55–71.
- Hasson, U. & Glucksberg, S. (2006). Does negation entail affirmation? The case of negated metaphors. *Journal of Pragmatics*, 38, 1015–1032.
- Hasson, U., Simmons, J. P., & Todorov, A. (2005). Believe it or not: On the possibility of suspending belief. *Psychological Science*, 16, 566–571.
- Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, 13(2), 141–154.
- Horton, W. S. & Rapp, D. N. (2003). Out of sight, out of mind: Occlusion and the accessibility of information in narrative comprehension. *Psychonomic Bulletin & Review*, 10(1), 104–109.
- Johnson, A. T. (1996). Comprehension of metaphors and similes: A reaction time study. *Metaphor & Symbolic Activity*, 11(2), 145–159.
- Johnson, C. & Keil, F. (2000). Explanatory knowledge and conceptual combination. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 327–359). Cambridge, MA: The MIT Press.
- Just, M. A. & Carpenter, P. A. (1976). The relation between comprehending and remembering some complex sentences. *Memory & Cognition*, 4(3), 318–322.
- Kaup, B. (2001). Negation and its impact on the accessibility of text information. *Memory & Cognition*, 29(7), 960–967.
- Keenan, J. M., Potts, G. R., Golding, J. M., & Jennings, T. M. (1990). Which elaborative inferences are drawn during reading?: A question of methodologies. In D. Balota, G. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading*. Hillsdale, NJ: Erlbaum.
- Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2), 259–284.
- Langacker, R. W. (1987). *Foundations of cognitive grammar*. Stanford, CA: Stanford University Press.
- MacDonald, M. C. & Just, M. A. (1989). Changes in activation levels with negation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 15(4), 633–642.
- Mayo, R., Schul, Y., & Burnstein, E. (2004). "I am guilty" vs. "I am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, 40(4), 443–449.
- McElree, B., Traxler, M. J., Pickering, M. J., Seely, R. E., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition*, 78(1), B17–B25.
- McGlone, M. S. & Harding, J. L. (1998). Back (or forward?) to the future: The role of perspective in temporal language comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 24(5), 1211–1223.
- McKoon, G. & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 440–466.
- McKoon, G. & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99(3), 440–466.
- McKoon, G. & Ratcliff, R. (1995). Conceptual combinations and relational contexts in free association and in priming in lexical decision and naming. *Psychonomic Bulletin & Review*, 2(4), 527–533.
- Meyer, D. E. & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227–234.

- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259.
- Peterson, L. & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193–198.
- Richards, A. & French, C. C. (1991). Effects of encoding and anxiety on implicit and explicit memory performance. *Personality & Individual Differences*, 12(2), 131–139.
- Searle, J. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor & Thought* (pp. 92–123). New York, NY: Cambridge University Press.
- Sloutsky, V. & Goldvarg, Y. (2004). Mental representation of logical connectives. *The Quarterly Journal of Experimental Psychology A*, 57(4), 636–665.
- Smith, C. D. (1981). Recognition memory for sentences as a function of concreteness/abstractness and affirmation/negation. *British Journal of Psychology*, 72(1), 125–129.
- Till, R. E., Mross, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory & Cognition*, 16(4), 283–298.
- Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory & Language*, 47(4), 530–547.
- Whitney, P., Waring, D. A., & Zingmark, B. (1992). Task effects on the spontaneous activation of trait concepts. *Social Cognition*, 10(4), 377–396.
- Wilkenfeld, M. J. & Ward, T. B. (2001). Similarity and emergence in conceptual combination. *Journal of Memory & Language*, 45(1), 21–38.