



Published in final edited form as:

Lang Cogn Process. 2011 ; 26(3): 376–405. doi:10.1080/01690965.2010.492642.

Representational Complexity and Memory Retrieval in Language Comprehension

Philip Hofmeister

Center for Research in Language, University of California-San Diego, La Jolla, CA, Telephone: (650)-387-6641

Philip Hofmeister: phofmeister@ucsd.edu

Abstract

Mental representations formed from words or phrases may vary considerably in their feature-based complexity. Modern theories of retrieval in sentence comprehension do not indicate how this variation and the role of encoding processes should influence memory performance. Here, memory retrieval in language comprehension is shown to be influenced by a target's representational complexity in terms of syntactic and semantic features. Three self-paced reading experiments provide evidence that reading times at retrieval sites (but not earlier) decrease when more complex phrases occur as filler-phrases in filler-gap dependencies. The data also show that complexity-based effects are not dependent on string length, syntactic differences, or the amount of processing the stimuli elicit. Activation boosting and reduced similarity-based interference are implicated as likely sources of these complexity-based effects.

Keywords

complexity; memory retrieval; encoding; filler-gap dependencies; sentence processing

Introduction

When you hear a description or read it in a text, you form a mental representation that can be thought of as a bundle of features (Miller, 1956; Anderson, Bothell, Byrne, Douglass, & Lebiere, 2004). Some of these features capture syntactic information, while others express semantic attributes. How rich or elaborate this representation is depends upon how much you hear or read about the entity in question. Hence, linguistic representations formed during language comprehension (hereafter, the encoding phase) may vary in terms of how complex they are. A phrase may only provide you with very little information (e.g. “something”), or it may give you lots of specific details (e.g. “the red book sitting right in front you”). Sometimes, of course, language understanding requires you to remember linguistic representations you encountered earlier, as when a pronoun must be processed or when some syntactic dependent must be retrieved.

These basic points raise a number of questions about the encoding-retrieval relationship in language comprehension. Do differences in representational properties, such as semantic richness or syntactic complexity, impact retrieval processes in language understanding? If so, how? Do the cognitive processes engaged during the encoding phase, particularly syntactic and semantic processing, influence how the resulting representation is retrieved from memory?

The present paper considers these questions in the context of so-called filler-gap dependencies, where a syntactic element is displaced from its syntactic head and standard hierarchical position. This process leaves an “empty” syntactic position, as shown below:

- (1) There was a meteor shower last night that I hoped that you got a chance to see ___ before it disappeared.

To understand this sentence, “a meteor shower” must be interpreted as the object of “see”, despite the fact that this object noun phrase appears eleven words before its verbal head. The relationship between these two sentence constituents is hence typically called a long-distance dependency and, more particularly, a filler-gap dependency (Fodor, 1978). Filler-gap dependencies link two phrasal constituents, one of which (*the filler*) is displaced from its standard structural position, leaving an empty structural *gap*. Hence, the displaced noun phrase “a meteor shower” must be maintained in memory until it can be interpreted as the argument of the verb “see.” Evidence that the information in a filler-phrase is retrieved at the syntactic head comes from a variety of empirical methods, including probe recognition tasks, cross-modal priming, reading time studies, and electrophysiological techniques (Tanenhaus, Carlson, & Seidenberg, 1985; Nicol & Swinney, 1989; Kluender & Kutas, 1993; Osterhout & Swinney, 1993; Swinney, Ford, Bresnan, & Frauenfelder, 1988; McElree, 2000).

Psycholinguistic research on dependency processing, however, does not directly address whether properties of the target representation or the processes engaged during the encoding of the target affect retrieval performance (except when the target features overlap with other representations in memory). The variation in the complexity of linguistic forms raises the possibility that differences in representational properties, such as semantic richness or syntactic complexity, may significantly impact the retrieval process in long-distance dependencies and potentially other syntactic contexts. If these properties do not matter, it suggests that encoding and representational attributes have no independent bearing on memory retrieval in comprehension.

A wealth of recall and recognition research, in contrast, indicates that how something is retrieved from memory interacts with characteristics of the encoding process and the target representation itself. These include the distinctiveness of the representation (with respect to the surrounding encoding context or background knowledge) and the elaboration associated with the encoding process (Reder, 1980; McDaniel, 1981; Bradshaw & Anderson, 1982; Anderson & Reder, 1979; Anderson, 1983; Wiseman, MacLeod, & Lootsteen, 1985; O’Brien & Myers, 1985; Reder, Charney, & Morgan, 1986; McDaniel, Dunay, Lyman, & Kerwin, 1988; Gallo, Meadow, Johnson, & Foster, 2008).

Bradshaw and Anderson (1982), for instance, show that sentence recall improves when the proposition appears with other causally-related propositions. They explain these results by postulating that additional propositional information that relates to some core concept increases the number of possible retrieval paths to the target concept. That is, additional information expands or creates a network of information that links together certain traces in memory. Retrieval of the targeted information can consequently transition through the related information via an associative recall process. A similar account for word recall and recognition expresses the idea that the addition of semantic or conceptual features provides a set of unique distinguishing characteristics, “making these memories less susceptible to interference and/or providing more features that can be cued on a typical recall or recognition memory test” (Gallo et al., 2008, p. 1096). On these accounts, retrieval benefits from the presence of unique features that allow indirect access a representation, and potentially, related representations via an associative relationship.

These findings in the sphere of long-term memory for isolated words and propositions add further weight to the possibility that a systematic relationship between representational richness or “complexity” and retrieval exists in online comprehension. Of course, the dynamics of retrieval in overt recall and recognition tasks may well differ from what

happens during retrieval in language comprehension. Individuals may employ conscious strategies for recollection in memory tasks that are impractical in the rapid context of language comprehension. However, several lines of research in sentence processing also hint that properties of stored representations factor into the determination of retrieval difficulty.

The Informational Load Hypothesis (ILH) (Almor, 1999, 2004; Almor & Nair, 2007), for instance, defines how the semantic or conceptual distance between an anaphor and antecedent affects processing of the anaphor. Anaphor processing, like dependency processing, requires memory retrieval at some level of linguistic representation (McKoon & Ratcliff, 1980; Gernsbacher, 1989; Ariel, 1990, *inter alia*). Among other predictions, the ILH states that given an anaphor that is more general than its antecedent, a greater conceptual distance between the anaphor and antecedent results in a smaller informational load and thus easier anaphor processing. Thus, the anaphor-antecedent pair “bird-ostrich” has a greater conceptual difference than “bird-robin”, so the anaphor in the former case is expected to have a lower informational load. Almor (1999) explains that “specificity is one factor that affects semantic distance” such that when the anaphor is more general than its antecedent, increasing the specificity of the antecedent should lower conceptual distance. Under the critical assumption that a more specific description like “the crippled robin” is more semantically distant from “the bird” than a phrase like “the robin,” the ILH implies that antecedents of greater semantic complexity reduce the processing costs of anaphors. The computational model of Almor (2004), in fact, incorporates the idea that semantically richer representations are more activated: “overall activation was assumed to be affected by not only memory activation but also by the amount of semantic detail in the representation” (p. 91).

Relatedly, Cowles and Garnham (2005) report faster reading times for anaphoric expressions (e.g. “vehicle”) given a more specific/semantically richer antecedent (e.g. “hatchback” vs. “car”). Their description of this effect also invokes the notion of conceptual distance: “hatchback” is considered more conceptually distant from “vehicle” than “car” is. Without advocating any one proposal, Cowles & Garnham mention several possibilities for why conceptual distance could aid anaphoric processing, including the idea that similar NP forms (less conceptually distant) are only licensed when the antecedent is difficult to retrieve.

All of these findings implicate a relationship between properties of stored representations and the processing of linguistic material that calls for reaccessing those stored items. What they do not do, however, is define what properties of stored representations matter or what the cause of this relationship is. Nor do they suggest if such a relationship exists in the absence of an overt reference to the stored discourse entity.

One way in which references to the same entity may differ, as already mentioned, is in terms of semantic and syntactic complexity. Here, I employ the following working definition of complexity for the purposes of this paper:

- (2) For two descriptions, x_1 and x_2 , denoting a discourse entity e , if the semantic and syntactic feature-value pairs encoded by x_2 are a proper subset of the feature-value pairs encoded by x_1 , then x_1 is more complex than x_2 .

Accordingly, complexity is cashed out here in terms of features. This working definition does not supply the means for counting all the features in a linguistic representation. Instead, it appeals to the idea that we can make *relative* assessments when a linguistic unit clearly carries all the feature-value specifications that another does, as well as some additional features.

A phrase like “the white bidet”, for instance, would be considered both syntactically and semantically more complex than the phrase “the bidet.” The former contains additional semantic features (color specification) and additional syntactic features (an adjectival modifier). The above definition also differentiates phrases that are semantically equivalent but have different syntactic complexities (e.g. “who” vs. “which person,” “which” vs. “which one of them”). Similarly, two phrases can differ in semantic complexity via category hierarchy differences, despite having identical syntactic structures (e.g. “a person” vs. “a politician”). In this case, I assume that the memory representation for “a politician” is associated with features specifying occupation or type of individual, unlike “a person.” Due to the wording of this definition, however, we cannot compare the complexity of descriptions such as “the white bidet” and “the round bidet,” even though both may refer to the same discourse entity; distinct semantic feature-values may be more or less beneficial for retrieval, depending on the context.

A series of self-paced reading tasks address the question of whether the syntactic and semantic complexity of a linguistic description has significant consequences for the retrievability of the corresponding mental representation. Representational complexity could theoretically interact with memory retrieval processes in language comprehension in several ways. If a description contains lots of semantic detail and is syntactically complex (e.g. has numerous modifiers), this could make the phrase difficult to process when first heard or seen. Reaccessing such a complex representation in memory would recreate this difficulty, if more features need to be checked or evaluated to guarantee successful retrieval. On this scenario, complicated structures and meanings that are generated during the encoding phase of comprehension lead to both difficult encoding and retrieval.

Alternatively, building and storing a complex representation in memory may actually help you to reaccess it. This counter-intuitive idea can be implemented in sentence processing theories where only some representational features need to be reaccessed for successful retrieval (e.g. cue-based theories of retrieval such as McElree (2000), Anderson, Budiu, and Reder (2001), Van Dyke and Lewis (2003)). It may be unnecessary, for example, to remember everything about a description of an individual, such as whether the noun phrase was indefinite or definite, the prosody on the noun, etc. Some features, though, would be critical for memory retrieval, particularly those generated by the retrieval context (Lewis, Vasishth, & Van Dyke, 2006). On such a view, complex memory structures would not necessarily be any harder to reaccess than simpler ones, because the quantity of features to identify the target is determined by the retrieval context and not the memory target itself.

Such selective reaccess theories, however, do not predict by themselves that encoding complex representations facilitates retrieval processes. However, there are several principled reasons why more features in a representation might aid memory processes. First, the effort of putting information together and extended processing may raise the salience or activation level of the representational network (Gernsbacher, 1989; Lewis & Vasishth, 2005; Vasishth & Lewis, 2006). Secondly, additional features may cause the representation to be more unique in memory (i.e. less confusable with other representations in memory). To use a vague referring expression such as “somebody,” for instance, creates a representation with little to distinguish it from other animate entities in the discourse model. Hence, underspecification may cause expressions to be non-distinct in memory and subject to interference. Finally, remembering one feature may enable retrieval of other critical features for memory retrieval (Bradshaw & Anderson, 1982; Anderson et al., 2001). In this case, additional semantic and syntactic features that have been reaccessed may actually strengthen the activation of other related features.

Before considering the experimental evidence, I first overview the statistical methods used to analyze the data. In the subsequent section, I test whether the complexity of filler-phrases affects reading times at retrieval points in filler-gap dependencies. The remaining experiments seek to not only verify the existence of complexity-based effects in other environments, but to examine why these effects occur. Experiment II addresses whether complexity-based effects result only from length or syntactic complexity differences. The third experiment is designed to answer whether the complexity-based differences are simply the result of increased processing during the encoding phase. The concluding section of this paper ties together the present findings with other recent research in sentence processing and offers a rationale for the existence of complexity-based effects.

Statistical Analysis

Self-paced reading tasks are employed for all of the experiments discussed here. In these comprehension experiments, subjects read sentences at their own pace on a computer screen (Just, Carpenter, & Woolley, 1982). Initially, they are presented with a screen of dashes separated by spaces, representing the words for that experimental item. With each press of a predefined key, a new word appears on the screen and the previous word disappears. Blocking of items into lists and randomization within lists was automatically managed by the reading time software, LINGER v. 2.94, developed by Doug Rohde (available at <http://tedlab.mit.edu/~dr/Linger/>).

Reading times were analyzed with linear mixed-effects models, using the lme4 package in R (version 2.4.0). This method of statistical analysis not only avoids the loss of statistical power that comes with prior subject- and item-averaging (Pinheiro & Bates, 2000), but as Baayen (2004) points out, it allows for a principled way of incorporating longitudinal effects and covariates into the analysis, as well as being free from the assumptions of constant covariance and sphericity. Markov chain Monte Carlo (MCMC) sampling ($n = 25,000$) was used to estimate conservative p -values for the fixed effects.

Reading time results were handled and analyzed using the following method which borrows heavily from the analyses described in Jaeger, Fedorenko, Hofmeister, and Gibson (2008) and Jaeger, Fedorenko, and Gibson (submitted). Unrealistic reading times (> 2500 ms) were removed prior to further analysis, as well as data from subjects with performance accuracy below 60% or a reading time average more than 2.5 standard deviations from the participant mean. Subsequently, raw reading times were log-transformed to normalize the data. Next, residual reading times were computed for each subject by regressing the log-transformed reading times from all stimuli (experimental items and fillers) against a number of predictors: (a) construction type, (b) word length, (c) the restricted cubic spline of the word position in the sentence, and (d) the logarithm transformed position of the trial in the experiment (see also <http://hlplab.wordpress.com/2008/01/23/modeling-self-paced-reading-data-effects-of-word-length-word-position-spill-over-etc/> for the specific implementation in R).

The first of these factors—construction type—takes into account the possibility that experimental items and fillers may impose unique processing loads based on their overall sentence complexity. For all three experiments described here, average reading times for the critical experimental items differed significantly from the baseline items (Experiment I: $\beta = -.061$, $SE = .004$, $t = -14.53$, $p < .0001$; Experiment II: $\beta = .032$, $SE = .005$, $t = 6.71$, $p < .001$; Experiment III: $\beta = .050$, $SE = .003$, $t = 14.14$, $p < .0001$). Similarly, word length differences contribute to reading time differences and are commonly controlled for in reading time studies (Ferreira & Clifton, 1986). Unsurprisingly, reading times increase with longer word length in all three experiments (Experiment I: $\beta = .018$, $SE = .001$, $t = 30.51$, p

< .0001; Experiment II: $\beta = .016$, $SE = .001$, $t = 24.64$, $p < .0001$; Experiment III: $\beta = .018$, $SE = .001$, $t = 32.67$, $p < .0001$). Participants also speed up drastically across trials: in fact, trial position emerges as the most significant predictor of reading times in the current studies (Experiment I: $\beta = -.167$, $SE = .003$, $t = -63.97$, $p < .0001$; Experiment II: $\beta = -.123$, $SE = .003$, $t = -42.33$, $p < .0001$; Experiment III: $\beta = -.131$, $SE = .002$, $t = -55.12$, $p < .0001$). Lastly, word position has a clear non-linear effect on reading times in all experiments, as Figure 1 illustrates for Experiment I (raw reading times are shown in Figure 1 to better convey the magnitude of word position effects, although logged reading times were used in the actual analysis). To capture these non-linearities, the restricted cubic spline of this predictor was used to model the relationship between word position and reading times (Harrell, Lee, & Pollock, 1998; Harrell, 2001). These parameters are summarized in the multilevel model below used for each experiment:

$$(3) \quad \log RT_{ij} = \beta_0 + \beta_1 \text{Construction}_i + \beta_2 \text{Length}_i + \beta_3 \log(\text{ListPosition}_i) + \beta_4 \text{rcs}(\text{WordPosition}_i) + b_j + \epsilon_{ij}$$

Hence, log reading times for word i , read by subject j , are predicted by an intercept (β_0), four parameters ($\beta_1 \dots \beta_4$), a subject random effect (b_j) and residual error (ϵ_{ij}). All reading time data from the experiment were employed (with the exception of practice items) in computations of these residual reading times, including filler items. The residuals of these models (*residual log reading times*) constitute the data discussed throughout this paper. After computing these residual log reading times and excluding reading times from incorrectly answered stimuli, data points more than 2.5 standard deviations from the mean at each word region for each experimental condition were removed. This process affected 2.9% of the data points in Experiment I, 3.0% of the data in Experiment II, and 2.6% of the data in Experiment III. Using this residualization method, as opposed to raw reading times or including these reading time predictors directly in the final models, has several advantages. First, each factor regressed out strongly influences reading times, as shown above. Secondly, regressing these predictors against reading times for all stimuli, including fillers, reduces the likelihood of collinearity. For instance, since two of the experiments described here manipulate sentence length and the word position of critical regions, word position effects cannot be estimated on the basis of the experimental items alone without introducing strong collinearity between the predictors of word position and the primary experimental manipulation. Accordingly, better estimates of the effects of word position are achieved by using all experimental items, including fillers.

The second major phase in analyzing the data for each experiment involved regressing the residual log reading times at a particular word region against the complexity variable and the random effects of subjects and items. Because reading time differences across conditions at a particular word region may reflect preexisting differences from a preceding word or region that have spilled over (Sanford & Garrod, 1989), I also treated reading times at the previous word as a fixed factor, in addition to the experimental manipulation of complexity. For analyses of multi-word regions, I employed the reading times from the word immediately preceding the region to estimate spillover effects. Note that all significant effects reported here remain significant (most effects even increase in size) with the use of multiple spillover variables (e.g. based on reading times at words $n-1$, $n-2$, $n-3$, etc.) rather than one. In short, each of the mixed-effects models described within the experimental sections contained two fixed factors (complexity and spillover) and two random factors (subjects and items) regressed against the residual log reading times.

For the comprehension question data, reaction time z-scores were computed for each subject. After removing z-scores more than 2.5 standard deviations from the mean, these z-scores were also analyzed with linear mixed-effects models that included list position as a fixed effect and subjects and items as random effects. Question-answer accuracies were

evaluated using generalized linear mixed-effects models; however, there were no effects of accuracy in any of the experiments discussed here, so these models are not described.

Experiment I: Syntactic & Semantic Complexity

This first experiment evaluates whether the addition of one or more adjectives, increasing both syntactic and semantic complexity, to a dislocated noun phrase significantly influences processing at or around the retrieval site.

Participants

Thirty-six native English-speaking University of California-San Diego undergraduates participated in this study to fulfill a course requirement.

Materials and Procedure

Twenty-four clefted indefinites varied in terms of how many adjectives preceded the clefted head noun: zero (SIMPLE), one (MID), or two (COMPLEX), as illustrated in (4). In all items, the clefted indefinite was followed by a relative pronoun and a five-word subject NP and then a transitive verb with an object gap, requiring the retrieval of the clefted indefinite phrase.

(4) SIMPLE: It was a communist who the members of the club banned from ever entering the premises.

MID: It was an alleged communist who the members of the club banned from ever entering the premises.

COMPLEX: It was an alleged Venezuelan communist who the members of the club banned from ever entering the premises.

Yes/no comprehension questions followed all stimuli. Sixty other items acted as fillers. Polynomial coding was used for this experiment to identify linear or quadratic effects of complexity with the conditions ordered according to the number of adjectives in the clefted noun phrase.

Results

The encoding phase for these indefinites shows strong effects of complexity: greater referential complexity increases reading times. As depicted in Figure 2, the averaged reading times of the head noun of the filler-phrase (“communist” in (4)) and the next two words are significantly higher in the COMPLEX and MID, compared to the SIMPLE condition. This pattern produces a significant linear effect of complexity ($\beta = .034$, $SE = .012$, $t = 2.91$, $p < .01$), but also a significant quadratic effect ($\beta = -.026$, $SE = .012$, $t = -2.24$, $p = .026$), since the MID and COMPLEX conditions group together. These effects occur along with a highly significant effect of spillover ($\beta = .271$, $SE = .024$, $t = 11.33$, $p < .001$). The significance and positive coefficient of this predictor means that higher reading times at the previous word predict higher reading times within this averaged region. In other words, trouble reading word $n-1$ is likely to spill over on to word n . In the COMPLEX and MID conditions, the previous word is an adjective, but in the SIMPLE condition, the previous word is the indefinite determiner. Together with the fact that reading times are, on average, slower on the preceding adjectives than the determiner (mean residual RTs: SIMPLE = -0.098 , $SE = .015$; MID = -0.042 , $SE = .018$; COMPLEX = 0.033 , $SE = .020$), the spillover variable here implies that some of the differences at the noun plus subsequent two words are attributable to the fact that reading times were already faster in the simplest condition. Since the critical complexity variable remains significant even with the spillover variable in the same model,

this means that the effect is reliable even after taking reading times at the previous region into consideration.

Reading time differences at the head noun alone reflect a linear effect of the complexity of the clefted phrases ($\beta = .048$, $SE = .020$, $t = 2.47$, $p < .05$), while the quadratic term is only marginal ($\beta = -.032$, $SE = .019$, $t = -1.64$, $p = .10$). The spillover variable again captures the fact that the model's most important predictor of reading times at the head noun is the reading time at the previous word ($\beta = .401$, $SE = .041$, $t = 9.80$, $p < .001$).

Immediately after the subcategorizing verb (e.g. "banned"), this pattern of results reverses. At the first word after the verb, reading times for the SIMPLE condition are *slower* than those for the MID condition, which are in turn slower than those of the COMPLEX condition, yielding a significant linear effect of complexity ($\beta = -.042$, $SE = .016$, $t = -2.53$, $p = .012$). The quadratic term is non-significant. As at the encoding site, reading times at the previous word have a significant effect on reading times ($\beta = .147$, $SE = .031$, $t = 4.68$, $p < .001$).

Considering the three-word region following the verb (regions 9 through 11 in Figure 2), an even stronger linear effect of complexity is observable ($\beta = -.032$, $SE = .011$, $t = -2.96$, $p < .01$); again, no quadratic effect is evident. This linear effect occurs along with a significant effect of spillover ($\beta = .144$, $SE = .021$, $t = 6.95$, $p < .001$). Notably, the processing facilitation for the more complex conditions does not emerge until the retrieval site. At intermediate regions between the retrieval site and the *wh*-relativizer, there are no significant effects of complexity.

Question-answer accuracy was unaffected by condition (COMPLEX: 89.6%, $SE = 1.80$; MID: 89.9%, $SE = 1.78$; SIMPLE: 91.7%, $SE = 1.63$). Similarly, question-answering times were unaffected by condition (all t s < 1).

Discussion

The reading time evidence indicates that greater semantic and syntactic complexity consumes more resources during the encoding phase of comprehension, but also leads to more efficient processing around the retrieval site. The syntactically and semantically more complex descriptions slowed reading when readers first encountered them, presumably due to the cost of constructing more syntactically and semantically elaborate representations. Faster response times for the more complex descriptions, however, are evident when the corresponding syntactic constituents have to be retrieved to complete the filler-gap dependency. In fact, a linear effect of complexity emerges at the retrieval site, as increasing representational complexity leads to faster processing. Since the faster reading times for the more complex conditions do not begin until the word regions where evidence of the missing constituent is available, it is reasonable to suppose that the reading time differences are tied to cognitive events connected to the retrieval process.

At a more general level, these findings stress that differences in representational complexity contribute to retrieval processes in language comprehension. Therefore, properties of the encoding phase appear to matter for retrieval in language comprehension, as they do in other memory-related tasks. Notably, retrieval in sentence comprehension happens covertly and rapidly (restricting the deployment of sophisticated recall strategies), suggesting that encoding-retrieval interactions can operate at an automatic, unconscious level of cognition.

Experiment II: Semantic Complexity

The critical targets for retrieval in the previous experiment differ in terms of both syntactic and semantic complexity. The disparity in syntactic complexity further creates length-based differences, which raises the possibility that the observed effects relate either to syntactic differences alone or to the increased time spent building a single representation.

Time-based facilitation in language comprehension has been proposed before to account for so-called anti-locality effects (Konieczny, 2000) and O'Brien and Myers (1985) report evidence that increased study time for sentences with unexpected words leads to improved recognition of the words (cf. Waddill and McDaniel 1998). The extra study time arguably opens the way to more extensive processing and elaboration, as suggested in Anderson and Reder (1979) & Reder (1979).

It is reasonable to speculate, therefore, that stretching out the encoding process of a particular discourse representation over a long period of time benefits retrieval. Extra words may provide additional time for encoding and perhaps even redundant encoding. Thus, Experiment II tests whether retrieval differences persist in the absence of differing syntactic complexities. By using extracted elements with the same number of words, but which differ in semantic complexity, effects of time can be distinguished from effects of complexity.

Participants

Twenty-eight English-speaking Stanford University students participated in this study for course credit. None participated in any of the other experiments described here.

Materials and Procedure

The sole manipulation for the sixteen items in this experiment was the semantic complexity of an embedded *which-N'* phrase (e.g. "which guard"). In the SIMPLE condition, the head noun was always "person." In the COMPLEX condition, the head noun describes the category or type of individual, typically with an occupational title. The *wh*-phrase was separated from its subcategorizing verb by six words, as in (5) below. The six words always consisted of a definite description with an attached relative clause.

- (5) a. COMPLEX: The lieutenant could not remember **which soldier** the commander that was deeply respected ordered to scout the area ahead.
- b. SIMPLE: The lieutenant could not remember **which person** the commander that was deeply respected ordered to scout the area ahead.

A yes/no comprehension question followed each trial (e.g. *Did the captain recall who was doing reconnaissance?*). Half had 'yes' answers and half had 'no' answers. The materials for this experiment accompanied sixty distractor items. The levels of complexity were treatment coded, such that the SIMPLE condition acted as the reference level. Negative coefficients thus indicate faster reading in the COMPLEX condition.

Results

Immediately after the *which-N'* phrases, there is a marginal slowdown in the COMPLEX condition ($\beta = .046$, $SE = .024$, $t = 1.93$, $p = .053$), compatible with the idea that the extra semantic features require additional resources to identify and integrate. As in the previous experiment, spillover effects are also evident after processing the filler-phrases ($\beta = .349$, $SE = .061$, $t = 6.88$, $p < .0001$). Conditional differences, however, dissipate between this initial post-NP region and the retrieval site, as depicted in Figure 3.

Reading times at the verb and subsequent word, however, show the reversal seen in the previous experiment: the semantically more complex COMPLEX condition is now read faster than the SIMPLE condition ($\beta = -.049$, $SE = .023$, $t = -2.14$, $p < .05$; Spillover: $\beta = 0.214$, $SE = .048$, $t = 4.46$, $p < .0001$).

Although there is a temporary slowdown of the COMPLEX condition at region 9 (two words after the subcategorizing verb), the subsequent two regions show the same pattern evidenced at the verb ($\beta = -.047$, $SE = .017$, $t = -2.74$, $p < .01$; Spillover: $\beta = .287$, $SE = .029$, $t = 7.49$, $p < .0001$). Consequently, no special significance is attributed to the pattern of results at region 9.

Question-answer accuracies do not reveal any effect of semantic complexity (COMPLEX: 77.2%, $SE = 2.7$; SIMPLE: 79.9%, $SE = 2.8$). Additionally, while question-answering times were numerically faster in the TYPE condition, this contrast is non-significant ($t = -1.53$, $p > .1$).

Discussion

According to the reading time evidence, facilitation effects for more complex descriptions do not depend exclusively on syntactic complexity differences. As in the previous experiment, the complexity-based differences accompany several salient features: (1) longer reading times immediately after the semantically more complex NPs, (2) faster reading times at the retrieval site for more complex NPs, and (3) the absence of reliably faster reading times for the complex NPs prior to the retrieval site.

Importantly, the findings from this experiment suggest that the reduction in processing effort at the retrieval site is not strictly dependent upon length-based differences. This does not rule out the possibility, however, that length plays a role. The data only confirm that complexity-based effects occur independently of differences in string length, as well as syntactic complexity.

Lastly, the fact that the facilitation effect for complex NPs kicks in precisely at the suspected retrieval site adds further weight to the interpretation that these effects are retrieval-based. Both experiments reviewed so far share this characteristic: equivalent or slower reading times for the more complex conditions prior to the retrieval site. Hence, these two sets of results argue for a processing difference that emerges during the process of retrieving and integrating the dislocated filler-phrase.

Experiment III: Processing Effort

In the preceding experiments, faster retrieval times accompany slower processing around the encoding of the critical noun phrases. One possible explanation, therefore, for the results is that more complex descriptions initiate additional processing that supports memory retention. Linguistic complexity, therefore, may constitute only one of innumerable ways for manipulating processing effort during the encoding phase.

Experiment III thus tests the hypothesis that the relationship between complexity and memory retrieval hinges upon how much processing a representation elicits. Additional processing is sometimes portrayed as the mechanism that supports enhanced memory performance for orthographically or semantically distinct words in recall and recognition tasks, since these stimuli generally stimulate a greater amount of processing than contextually nondistinct items (Hirshman, Whelley, & Palij, 1989; Watkins, LeCompte, & Kim, 2000). On the hypothesis that faster retrieval ultimately depends upon the amount of processing, additional predictable features should help retrieval less than unexpected or

atypical features, because the latter should elicit more attention and processing. Alternatively, if predictable features lead to faster or equivalent retrieval times, compared to unexpected features, then processing time during the encoding phase is not the key component to retrieval facilitation.

Participants

Thirty-seven students from Stanford University and Foothill Community College participated in this experiment—all of whom identified as native speakers of English. Subjects either received course credit or were paid \$12 for their participation.

Materials and Procedure

The materials for this experiment consisted of twenty-four items with three conditions each. In the SIMPLE condition (6a), participants saw a definite NP with no additional modifiers, but in the other two conditions, the definite object NP contained two additional words. The second word was held identical in both conditions, but the first word was manipulated for distinctiveness or typicality, as shown in (6). In the COMPLEX-TYP(ICAL) condition (6b), the first word expressed a common or highly predictable feature of the head noun (e.g. ruthlessness is a predictable characteristic of dictators). By contrast, the first word in the COMPLEX-ATYP(ICAL) condition, (6c), encoded an unexpected or uncommon feature (e.g. dictators are not typically associated with being lovable):

- (6)
- a. The diplomat contacted **the dictator** who the activist looking for more contributions encouraged to preserve natural habitats and resources.
 - b. The diplomat contacted **the ruthless military dictator** who the activist looking for more contributions encouraged to preserve natural habitats and resources.
 - c. The diplomat contacted **the lovable military dictator** who the activist looking for more contributions encouraged to preserve natural habitats and resources.

Stimuli were followed by yes/no comprehension questions, half of which had ‘yes’ answers and half ‘no’ answers. An equal number of comprehension questions (4) targeted information in the matrix subject, matrix verb, matrix object, embedded subject, embedded verb, and post-verbal spillover region. These items appeared along with forty-eight filler items, twelve of which contained similarly complex sentences but with syntactically and semantically complex subject phrases, instead of object phrases. As in the previous experiment, the experimental conditions were treatment coded with the SIMPLE condition as the reference level.

The typicality of modifying features was determined by two methods. First of all, Gigaword, a 1.2 billion word corpus of written English, was used to evaluate how often the critical adjective appeared before the head noun. The results of these corpus searches are shown in Table 1. While the overall conditional probabilities for the adjective-noun combinations prove to be quite low for both conditions, the ratios of the probabilities show that the adjective-noun combinations in COMPLEX-TYP condition are much more likely. Even taking into account the difference in mean frequencies of the adjectives, the conditional probabilities for the combinations in the COMPLEX-TYP condition are still much higher.

Secondly, a norming study was conducted with eleven English-speaking participants to verify the categorizations (also compensated with course credit). Subjects rated on a scale of 0–10 how likely an individual of the sort described by the object head noun (e.g. *dictator*) is to have the characteristic described by the adjective (e.g. *ruthless*). Each subject saw each of

the critical twenty-four head nouns used in the reading experiment with either the typical or atypical adjective, along with twenty-six filler items. The scores were computed for each data point by subtracting the original score from the subject's mean score across all items.

Items categorized as typical (COMPLEX-TYP: 2.00, $SE = .127$) were judged to be more likely characteristics without exception than items categorized as atypical (COMPLEX-ATYP: -2.00, $SE = .180$). A linear mixed-effects model with subjects and items as random factors and likelihood as the sole fixed factor shows that these differences are highly reliable ($\beta = 4.05$, $t = 18.94$, $p < .0001$).

Results

As in the first experiment, the more complex definite descriptions elicit slower reading times at the head noun and immediately subsequent word regions. At the head noun (region 1 in Figure 4), both the COMPLEX-TYP and COMPLEX-ATYP conditions are read slower than the SIMPLE condition (COMPLEX-ATYP: $\beta = .128$, $SE = .026$, $t = 4.96$, $p < .0001$; COMPLEX-TYP: $\beta = .077$, $SE = .025$, $t = 3.02$, $p < .01$). These effects are independent of spillover effects from the previous word ($\beta = .351$, $SE = .038$, $t = 9.27$, $p < .0001$).

Averaging the reading times at the head noun and the next two words, a stronger three-way split is evident: the atypical NPs lead to the slowest reading times, followed by the typical NPs, with the bare definites leading to the fastest reading times (COMPLEX-ATYP: $\beta = .113$, $SE = .019$, $t = 6.06$, $p < .0001$; COMPLEX-TYP: $\beta = .059$, $SE = .019$, $t = 3.16$, $p < .01$). Again, reading times at the word prior to the averaged region emerge as a significant predictor of reading times ($\beta = .054$, $SE = .026$, $t = 2.05$, $p < .05$). A separate model fitted with only variables for the COMPLEX-TYP and COMPLEX-ATYP conditions verifies that the difference between them is significant (COMPLEX-TYP: $\beta = -.054$, $SE = .018$, $t = -2.94$, $p < .01$).

Although the COMPLEX-ATYP condition produces the slowest reading times when the critical NP is first read, reading times at the retrieval site (region 9 in Figure 4) for this condition are comparable to that of the SIMPLE condition (COMPLEX-ATYP: $\beta = -.021$, $SE = .024$, $t = -0.88$, $p = .38$). Additional modeling shows that the absence of a difference between the SIMPLE and COMPLEX-ATYP conditions persists even when excluding the data from the COMPLEX-TYP condition (i.e. removing a possible source of collinearity; COMPLEX-ATYP: $\beta = -.021$, $SE = .025$, $t = -0.81$, $p = .41$). At the subsequent word (region 10), the SIMPLE condition is also marginally faster than the COMPLEX-ATYP condition ($\beta = .044$, $SE = .023$, $t = 1.91$, $p = .06$). The COMPLEX-TYP condition does, however, evidence facilitated processing at the retrieval site, when contrasted with reading times in the SIMPLE condition (COMPLEX-TYP: $\beta = -.069$, $SE = .024$, $t = -2.94$, $p < .01$; Spillover: $\beta = .300$, $SE = .040$, $t = 7.49$, $p < .0001$).

To further understand the relationship between representational likelihood and retrieval, the reading time data from region 9 (the verb) were regressed against the mean likelihood scores from the norming study in a separate linear mixed-effects model with subjects and items as random factors. Like the previously described models, this model also included a spillover predictor. The mean likelihood score proves to be a good predictor of reading times in this region ($\beta = -.011$, $SE = .005$, $t = -2.12$, $p < .05$). As is reflected by the negativity of the coefficient, reading times go down as mean likelihood goes up. In other words, as the likelihood of the representation increases, processing times at the retrieval site go down.

As in the previous experiment, complexity had no effect on comprehension question accuracy (COMPLEX-ATYP: 75.7%, $SE = 2.5$; SIMPLE: 75.0%, $SE = 2.5$; COMPLEX-

TYP: 77.0%, $SE = 2.4$). Similarly, there were no differences across conditions with respect to question-answering times (all $t_s < 1$).

Discussion

The main finding of this experiment is that the strength of the retrieval facilitation is not reducible to the amount of processing that the encoded representations elicit. Atypical complex definites elicited more initial processing than either of the other conditions, yet were not processed fastest at the retrieval site. This argues against an interpretation of the retrieval facilitation effects as strictly dependent upon the amount of processing that occurs during the encoding phase.

Richer representations containing typical or highly predictable feature combinations, in contrast, yielded the expected retrieval facilitation at the retrieval site. Compared to a bare definite description, the extra features in these complex phrases led to slower reading times around the encoding of the critical noun phrase. In this case, though, the additional syntactic and semantic features accompanied faster processing at the verb that triggers retrieval. In fact, the model which used likelihood evaluations as a predictor of the reading times points to a linear trend for faster retrieval as typicality increases. Hence, the division of complex NPs into atypical and typical actually masks a continuum of likelihood with increasing likelihood associated with enhanced memory performance.

Overall, these results suggests several things: (1) the retrieval facilitation does not appear to be strictly based on how much processing the to-be-retrieved target elicits during encoding; (2) the quality of information also factors into the determination of retrieval ease. This echoes the conclusions of Bradshaw & Anderson (1982) that whereas elaborations that bear a causal relationship to a target proposition improve recall, unrelated elaborations do not. That is, extra propositional information in and of itself does not aid recall. The spirit of these findings is summed up in Stein et al. (1978): “effective elaboration seems to depend on the *quality* rather than the quantity of the information expressed.” In the present case, infrequent and less plausible features combinations were unhelpful in the retrieval process; however, relatively frequent and plausible combinations did facilitate retrieval processing.

There are several (non-mutually exclusive) candidate explanations for this difference, which can be divided into their relevance for the encoding versus the retrieval process. One explanation is that unexpected feature combinations may interfere with the success of the encoding process. Accurate encoding may occur less often when key components are improbable together, reducing the chances for successful retrieval. Given a successful encoding process, though, retrieval performance could also vary with the associative strength between the representational components. Retrieving a mental representation (or a set of representational features) from memory may first involve accessing *part* of that representation. If two features encoded as part of a larger representation are highly associated (e.g. “ruthless” & “dictator”), then successful retrieval of one may facilitate retrieval of the other (e.g. Anderson et al., 2001). In contrast, two features with low associative strength (e.g. “lovable” & “dictator”) should be less capable of facilitating each other’s retrieval and may even interfere with retrieval if the individual features are implausible enough together.

Alternatively, the activation of elements with high associative strength may increase via *activation spreading* (Budiu, 2001; Budiu & Anderson, 2004). Elements in the focus of attention (e.g. that are being read or heard), spread activation to other memory items, according to the proposal in Budiu and Anderson (2004). The more semantically related two elements are, the greater the boost in activation. Semantic features that are highly associated with other properties translate to higher activation of the related memory items and thus

more efficient retrieval. If two elements are unrelated enough, the activation spread may actually be negative, leading to a more difficult retrieval process.

On all of these scenarios, improbable feature combinations can be unhelpful or even harmful for retrieval. It appears then that semantically rich representations benefit retrieval most when the representational components are strongly associated. Most importantly, these findings indicate that the retrieval benefits linked to complexity cannot be reduced to the extent of processing during the encoding phase.

General Discussion

The data show that differences in representational complexity play an important role in retrieval in language comprehension. In particular, the reading time evidence from Experiment I shows that greater syntactic and semantic complexity is associated with faster processing at retrieval sites. These effects persist even after controlling for length (Experiment II) and are not strictly contingent upon the amount of processing in the encoding phase (Experiment III). Furthermore, these effects remain after the effects of spillover, construction type, word position, word length, and list position are taken into account.

Recent sentence processing research and modeling offer some potential motivations for these findings. A considerable body of research suggests that repeated mentions strengthen a representation's activation, arguably because the triggered retrieval process restores the retrieved memory chunk to attentional focus (Gernsbacher, 1989; Anderson & Lebiere, 1998; Anderson et al., 2001, 2004; Anderson, 2005). Hence, the more often some particular representation is retrieved, the higher its activation level will be, making future memory retrievals easier. Building on this research, Lewis and Vasishth (2005) argue that processing a word or phrase whose interpretation depends upon or modifies some representation already in memory requires reactivating that memory representation. This reactivation boosts activation, comparable to the effect of a repeated mention. In short, modifying or embellishing some representation requires its retrieval to integrate new information.

The relationship between representational complexity and memory retrieval can be similarly justified (at least in part). On the view espoused by Lewis and Vasishth, each of the modifiers in Experiment I preactivates the predicted head noun category and processing the head noun itself naturally boosts its own activation. The retrieval differences thus can be linked to the process of repeatedly accessing or predicting some syntactic head. The processes of access and integration are costly from a short-term perspective, but may facilitate processing when the relevant representation must be re-accessed (see Kluender (1998) for a related point). The complexity-based retrieval effects can thus be thought of as being related to the necessary processes of building a complex representation.

Retrieval differences persist, however, when holding syntactic complexity constant (Experiment II) and no differences emerge when the additional syntactic complexity corresponds with improbable semantic features (Experiment III). Therefore, activation boosting due to syntactically triggered retrievals cannot be the full story. Similar activation increases due to meaning computation processes could theoretically explain the results of Experiment II—encoding semantically complex phrases could involve multiple retrievals. The plausibility of such an account depends partly on whether initiating multiple retrievals when processing semantically rich constituents accords with the time constraints implicated by the reading time data. Even if such an account is not ultimately plausible, semantically rich phrases may have more feature associations that allow indirect access to the target representation.

Along these lines, the reading time evidence emphasizes that it is not merely the quantity, but also the quality of the features that matters. Targeted representations with features that have high associative strengths lead to faster processing at retrieval sites, according to the results from Experiment III. This aligns with the notion that words spread activation to other semantically-related words (Budi & Anderson, 2004). Alternatively, retrieval—rather than encoding—of one set of features activates other features (Bradshaw & Anderson, 1982; Waddill & McDaniel, 1998; Anderson et al., 2001). Here, too, the resulting degree of activation would be a function of associative strength. Complex representations thus pose more opportunities for successful retrieval, if recovering one feature facilitates the recovery of other features in that representation. In fact, the effects in Experiment II are also subject to an explanation in terms of associativity. The verbs that trigger retrieval (e.g. “order”) may activate features associated with the semantically richer NPs (e.g. “soldier”) but which are not linked to the semantically simpler NPs. On such a view, complexity benefits retrieval because of an increasing probability that a retrieval cue will be associated with a feature of the target representation.

Lastly, representational complexity can alleviate some of the difficulty that accompanies the simultaneous presence of representations with overlapping features in memory (i.e. similarity-based interference). Increasing representational complexity increases the probability that some features will be unique and therefore helps distinguish a representation from other competitors in memory. On cue-based retrieval theories, such uniqueness may create a better match with the set of retrieval cues (Van Dyke & McElree, 2006), and parsing highly similar descriptions may generally reduce the activation among these items due to *encoding interference* (Suckow, Vasissth, Lewis, & Smith, 2006). Representations with unique features, as the result of complexity, thus avoid potential drops in activation due to similarity. As pointed out by Waddill and McDaniel (1998, p. 118) and others, though, uniqueness is relative: “Discriminative features are useful only when some, but not all, of the items have the features.” The uniqueness of complex representations therefore depends not only on features of the to-be-remembered item, but also on whether or not other representations in the discourse context share these same features. From this perspective, the encoding-retrieval relationship in language comprehension does not depend purely upon the number of features in the memory target: memory retrieval is simply best served by targeting maximally unique representations that have highly associated (possibly even redundant) features.

Memory research going back to Morris, Bransford, and Franks (1977) shows that whether some encoding difference improves or impairs retrieval depends on the nature and context of the retrieval task. In language comprehension, memory retrieval necessitates accessing some set of semantic and syntactic features that are potentially shared amongst multiple linguistic representations, creating one of the main problems for retrieval in sentence comprehension (Lewis, 1996; Van Dyke & McElree, 2006; Lewis et al., 2006; Van Dyke, 2007). For other tasks—even other language-related tasks—other types of features may be more pertinent for memory retrieval. For instance, syntactic and semantic features may be largely irrelevant in memory tasks targeting phonetic properties (e.g. remembering whether you saw a word that rhymes with a recall prompt—syntactic complexity may even complicate retrieval since it correlates with increased phonetic content). When viewed this way, the present data do not point to a general mnemonic advantage for syntactic and semantic processing across all memory tasks. For sentence processing, however, syntactic and semantic complexity appear to interact with retrieval because increasing complexity along these dimensions increases the probability that features that uniquely predict a target will come into play. That is, what matters for retrieval in sentence processing are syntactic and semantic features and whether the target stands out contextually in these respects. Complexity thus constitutes one out of

many theoretically possible means for modulating representational uniqueness, but the ultimate effect of any such modulation depends upon the memory retrieval setting.

Considering these different explanations, representational complexity may influence retrieval processes in language comprehension for distinct reasons. Lewis and Vasishth's work supports the view that the process of building the complex representation creates a highly activated memory network. At the same time, as the resulting representational complexity increases, the probability increases that the ensuing cognitive structure carries unique distinguishing features. Put slightly differently, both the *process* and the *result* of building a complex representation potentially contribute to the dynamics of the retrieval process in language comprehension. It is an open question as to whether complexity-based effects can be fully captured by either aspect of complexity alone, or whether both are necessary.

In sum, these results demonstrate that the memory processes in language comprehension are sensitive to the properties of stored representations, just as they are in other settings. Whereas a rich history of memory research has considered the encoding-relationship in detail, this relationship has been left largely unexplored in the context of sentence processing. The current work represents a first step in that direction.

Acknowledgments

For valuable comments and suggestions on this manuscript, I thank Marta Kutas, Lyn Frazier, Florian Jaeger, Ivan Sag, and the anonymous reviewers. All errors and inaccuracies are my own. This research was supported by NIH Training Grant T32-DC000041 via the Center for Research in Language at UC-San Diego.

References

- Almor A. Noun-phrase anaphora and focus: the informational load hypothesis. *Psychological Review*. 1999; 106:748–765. [PubMed: 10560327]
- Almor, A. A computational investigation of reference in production and comprehension. In: Trueswell, J.; Tanenhaus, M., editors. *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press; 2004. p. 285-301.
- Almor A, Nair V. The form of referential expressions in discourse. *Language and Linguistics Compass*. 2007; 1:84–99.
- Anderson, JR. *The architecture of cognition*. Harvard, MA: Harvard University Press; 1983.
- Anderson JR. Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science*. 2005; 29:313–341. [PubMed: 21702777]
- Anderson JR, Bothell D, Byrne M, Douglass S, Lebiere C. An integrated theory of mind. *Psychological Review*. 2004; 111:1036–1060. [PubMed: 15482072]
- Anderson JR, Budson R, Reder L. A theory of sentence memory as part of a general theory of memory. *Journal of Memory and Language*. 2001; 45:337–367.
- Anderson, JR.; Lebiere, C. *The atomic components of thought*. Mahwah, NJ: Erlbaum; 1998.
- Anderson, JR.; Reder, L. An elaborative processing explanation of depth of processing. In: Cermak, L.; Craik, F., editors. *Levels of processing in human memory*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1979. p. 385-404.
- Ariel, M. *Accessing noun-phrase antecedents*. London: Routledge; 1990.
- Baayen RH. Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers*. 2004; 1:1–45.
- Bradshaw G, Anderson JR. Elaborative encoding as an explanation of levels of processing. *Journal of Verbal Learning and Verbal Behavior*. 1982; 21:165–174.
- Budson, R. Unpublished doctoral dissertation. Carnegie Mellon University; 2001. The role of background knowledge in sentence processing.

- Budiu R, Anderson JR. Interpretation-based processing: A unified theory of semantic sentence comprehension. *Cognitive Science*. 2004; 28:1–44.
- Cowles H, Garnham A. Antecedent focus and conceptual distance effects in category noun-phrase anaphora. *Language and Cognitive Processes*. 2005; 20:725–750.
- Ferreira F, Clifton C. The independence of syntactic processing. *Journal of Memory and Language*. 1986; 25:348–368.
- Fodor J. Parsing strategies and constraints on transformations. *Linguistic Inquiry*. 1978; 9:427–473.
- Gallo D, Meadow N, Johnson E, Foster K. Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language*. 2008; 58:1095–1111.
- Gernsbacher M. Mechanisms that improve referential access. *Cognition*. 1989; 32:99–156. [PubMed: 2752708]
- Harrell, F. Regression modeling strategies. Oxford: Springer; 2001.
- Harrell F, Lee K, Pollock B. Regression models in clinical studies: Determining relationships between predictors and response. *Journal of the National Cancer Institute*. 1998; 80:1198–1202. [PubMed: 3047407]
- Hirshman E, Whelley M, Palij M. An investigation of paradoxical memory effects. *Journal of Memory and Language*. 1989; 28:594–609.
- Jaeger TF, Fedorenko E, Gibson E. Anti-locality in English: Consequences for theories of sentence comprehension. submitted.
- Jaeger, TF.; Fedorenko, E.; Hofmeister, P.; Gibson, E. Proceedings of the 21st Annual CUNY Sentence Processing Conference. University of North Carolina; Chapel Hill: 2008. Expectation-based syntactic processing: Anti-locality effects outside of head-final languages.
- Just M, Carpenter P, Woolley J. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*. 1982; 111:228–238. [PubMed: 6213735]
- Kluender, R. On the distinction between strong and weak islands: A processing perspective. In: Culicover, P.; McNally, L., editors. *Syntax and semantics 29: The limits of syntax*. San Diego, CA: Academic Press; 1998. p. 241-279.
- Kluender R, Kutas M. Bridging the gap: Evidence from erps on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*. 1993; 5:196–214.
- Konieczny L. Locality and parsing complexity. *Journal of Psycholinguistic Research*. 2000; 29:627–645. [PubMed: 11196066]
- Lewis R. Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*. 1996; 25:93–115. [PubMed: 8789368]
- Lewis R, Vasishth S. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*. 2005; 29:1–45. [PubMed: 21702765]
- Lewis R, Vasishth S, Van Dyke J. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*. 2006; 10:447–454.
- McDaniel M. Syntactic complexity and elaborative processing. *Memory and Cognition*. 1981; 31:35–43.
- McDaniel M, Dunay P, Lyman B, Kerwin ML. Effects of elaboration and relational distinctiveness on sentence memory. *The American Journal of Psychology*. 1988; 101:357–369. [PubMed: 3177699]
- McElree B. Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*. 2000; 29:111–123. [PubMed: 10709178]
- McKoon G, Ratcliff R. The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior*. 1980; 19:668–682.
- Miller G. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*. 1956; 63:81–97. [PubMed: 13310704]
- Morris CD, Bransford JD, Franks J. Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*. 1977; 16:519–533.
- Nicol J, Swinney D. The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*. 1989; 18:5–19. [PubMed: 2647962]

- O'Brien E, Myers J. When comprehension difficulty improves memory for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1985; 11:12–21.
- Osterhout L, Swinney D. On the temporal course of gap-filling during comprehension of verbal passives. *Journal of Psycholinguistic Research*. 1993; 22:273–286. [PubMed: 8366478]
- Pinheiro, J.; Bates, D. *Mixed-effects models in S and S-PLUS statistics and computing*. New York: Springer; 2000.
- Reder L. The role of elaborations in memory for prose. *Cognitive Psychology*. 1979; 11:221–234.
- Reder L. The role of elaboration in the comprehension and retention of prose: A critical review. *Review of Educational Research*. 1980; 50:5–53.
- Reder L, Charney D, Morgan K. The role of elaborations in learning a skill from an instructional text. *Memory and Cognition*. 1986; 14:64–78.
- Sanford A, Garrod S. What, when, and how?: Questions of immediacy in anaphoric reference resolution. *Language and Cognitive Processes*. 1989; 4:235–262.
- Suckow, K.; Vasishth, S.; Lewis, R.; Smith, M. Interference and memory overload during parsing of grammatical and ungrammatical embeddings. *Proceedings of the 19th Annual CUNY Sentence Processing Conference*; 2006.
- Swinney, D.; Ford, M.; Bresnan, J.; Frauenfelder, U. Coreference assignment during sentence processing. In: Macken, M., editor. *Language structure and processing*. Stanford, CA: CSLI; 1988.
- Tanenhaus, M.; Carlson, G.; Seidenberg, M. Do listeners compute linguistic representations?. In: Zwicky, A.; Karttunen, L.; Dowty, D., editors. *Natural language parsing: Psycholinguistic, theoretical, and computational perspectives*. London and New York: Cambridge University Press; 1985. p. 359-408.
- Van Dyke J. Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2007; 33(2): 407–430.
- Van Dyke J, Lewis R. Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*. 2003; 49:285–316. [PubMed: 18079972]
- Van Dyke J, McElree B. Retrieval interference in sentence comprehension. *Journal of Memory and Language*. 2006; 55:157–166. [PubMed: 18209744]
- Vasishth S, Lewis RL. Argument-head distance and processing complexity: Explaining both locality and anti-locality effects. *Language*. 2006; 82:767–794.
- Waddill P, McDaniel M. Distinctiveness effects in recall: Differential processing or privileged retrieval? *Memory & Cognition*. 1998; 26:108.
- Watkins M, LeCompte D, Kim K. Role of study strategy in recall of mixed lists of common and rare words. *Journal of Experimental Psychology: Learning Memory and Cognition*. 2000; 26:239–245.
- Wiseman S, MacLeod CM, Lootsteen P. Picture recognition improves with subsequent verbal information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1985; 11:588–95.

Appendix A

Experiment I Items

Experimental stimuli appeared with 0, 1, or 2 adjectives in the clefted noun phrase.

1. It was a ((reclusive) English) writer that the dignitaries at the ceremony awarded with a medal in Stockholm.
2. It was a ((famous) deaf) sculptor that the aristocrats at the gallery ridiculed during the exclusive art show.
3. It was an ((unsuspecting) young) sophomore that the pranksters at the fair frightened by setting off loud firecrackers.

4. It was a ((dangerous) Russian) mobster that the jurors in the trial imprisoned for thirty years without parole.
5. It was a ((clueless) hospice) nurse that the surgeons in the hospital accused of poisoning the elderly man.
6. It was a ((heartless) professional) mercenary that the commanders at the base hired for the mission in Guatemala.
7. It was an ((incompetent) prison) guard that the warden of the prison blamed for the escape attempt yesterday.
8. It was a ((notorious,) rich) Frenchman that the authorities in San Francisco identified as the smuggler of diamonds.
9. It was a ((poor) local) fisherman that the villagers on the beach saw on the stormy ocean seas.
10. It was a ((Texas) cattle) rancher that the officials for the state subsidized throughout the worst drought periods.
11. It was a ((crooked) government) bureaucrat that the citizens of the county despised for always accepting illegal bribes.
12. It was an ((alleged) Venezuelan) communist that the members of the club banned from ever entering the premises.
13. It was a ((successful) marketing) entrepreneur that the investors in the company invited to the banquet on Thursday.
14. It was a ((fearless) German) environmentalist that the activists for energy conservation congratulated for all the crucial accomplishments.
15. It was a ((generous) oil) billionaire that the organizers of the campaign thanked for the largest ever contribution.
16. It was a ((helpless,) crying) child that the neighbors from next door pulled from the burning apartment building.
17. It was a ((ruthless) military) dictator that the diplomats from neighboring countries advised to avoid another election process.
18. It was a ((wounded) American) soldier that the townspeople in the square rescued from the tank that was on fire.
19. It was a ((hilarious) stand-up) comedian that the audience in the bar heckled all the way through the routine.
20. It was a ((peaceful) Buddhist) monk that the protestors at the rally supported in the quest for Tibetan freedom.
21. It was a ((victorious) four-star) general that the committee on foreign relations questioned for over two hours yesterday.
22. It was a ((struggling) rock) musician that the fans at the club booed for forgetting the lyrics to the songs.
23. It was a ((daring) ocean) explorer that the crew of the ship resented for taking them closer to the ice cap.
24. It was a ((nerdy) computer) programmer that the executives at the firm chose to lead the design team for their website.

Experiment II Items

Experimental stimuli appeared either with the *wh*-phrase listed below or “which person”.

1. The lieutenant could not remember which soldier the commander that was deeply respected ordered to scout the area ahead.
2. Stephen could not identify which volunteer the terrorist that was widely feared captured during a humanitarian aid mission.
3. Harold figured out which student the author that was frequently quoted encouraged to try writing a novel.
4. The foreman sensed which novice the architect that had recently died inspired to attempt a daring design.
5. The manager forgot which customer the salesman that hardly ever laughed conned into accepting the extended warranty.
6. Naomi indicated which applicant the committee that was annoyingly rushed rejected without reading the personal statement.
7. She recorded which trustee the chairman that was just replaced phoned to vehemently express his anger.
8. He learned which toddler the pediatrician that was usually insightful diagnosed with a slight stomach flu.
9. The activist determined which representative the tycoon that was without morals bribed to log the ancient rainforest.
10. Heather admitted which scientist the dean that was recently instated reprimanded for plagiarizing some rare texts.
11. He theorized which friend the host that was getting angry ignored for reasons nobody really knows.
12. She kept secret which model the photographer that had been drinking molested to avoid a serious scandal.
13. A jury decided which accomplice the kidnapper that was already sentenced corrupted too much to let go.
14. A journalist investigated which athlete the trainer that was quickly indicted injected with some illegal steroid substance.
15. A health inspector asked which janitor the chef that had two restaurants hired to clean on the weekends.
16. A bystander noticed which steelworker the girder that was dangerously dangling struck without doing any serious harm.

Experiment III Items

Experimental stimuli appeared either with 0 or 2 modifiers. The adjective for the COMPLEX-TYP condition appears first in each item in the list below, followed by the adjective in the COMPLEX-ATYP condition. The second modifier occurs in both complex conditions.

1. The reviewer criticized the (famous/blind young) actor who the director making the art film ignored during the opening night festivities.

2. The official congratulated the (sensitive/childish English) poet who the professor trying to get tenure admired since reading the book reviews.
3. The diplomat contacted the (ruthless/lovable military) dictator who the activist looking for more contributions encouraged to preserve natural habitats and resources.
4. The reporter interviewed the (wealthy/bankrupt teen) celebrity who the photographer struggling to find work embarrassed last week at a charity dinner.
5. The student defended the (brilliant/secretive schizophrenic) mathematician who the dean attempting to avoid controversy removed from the university ethics committee.
6. The defendant accused the (corrupt/scholarly narcotics) cop who the judge presiding in the case silenced after a disturbing courtroom outburst.
7. The guard aided the (dangerous/responsible exiled) criminal who the agent overseeing the dramatic pursuit apprehended following a long and tiring chase.
8. The employee obeyed the (professional/immature plant) supervisor who the inspector reviewing the safety measures cautioned about the poor work conditions.
9. The captain evaluated the (brave/lonely volunteer) fireman who the veteran planning to retire soon trained over the course of six months.
10. The investigator summoned the (injured/sickly rookie) athlete who the coach running out of ideas invited to try out for a spot on the team.
11. The pundit ridiculed the (Republican/demented senate) candidate who the leader making the final decision supported despite the misgivings of other members.
12. The interrogator questioned the (heartless/courteous former) mercenary who the commander organizing the armed rebellion abandoned without any explanation or warning.
13. The fugitive robbed the (lost/dead American) tourist who the guide showing the group around warned about straying too from the group.
14. The industrialist threatened the (poor/gifted Russian) peasant who the investor searching for new opportunities protected without the least bit of hesitation.
15. The executive infuriated the (liberal/morbid socialist) politician who the lobbyist representing some oil companies bribed to vote for the upcoming bill.
16. The ranger followed the (fearless/girlish crocodile) hunter who the landowner building a new house despised for trespassing on his land.
17. The advisor lectured the (handsome/insane blonde) prince who the dignitary learning to speak English thanked at the end of the ceremony.
18. The waitress married the (reclusive/vindictive musical) genius who the psychiatrist known for being insightful treated for a mild case of depression.
19. The pedestrian dodged the angry/sullen taxi driver who the bystander waiting for a bus identified later on at the police station.
20. The customer offended the (helpful/senile new) assistant who the manager hoping for a raise called into his office after the incident.
21. The nurse consoled the (dying/hostile elderly) patient who the doctor working a double shift forgot due to a lack of sleep.

22. The representative consulted the (conservative/ adventurous legal) strategist who the prosecutor looking into corporate fraud knew since they went to school together.
23. The scientist avoided the (impartial/enraged academic) observer who the technician assisting the busy researchers escorted around the recently built facilities.
24. The secretary aggravated the (young/old female) intern who the partner negotiating a huge settlement hired less than three weeks ago.

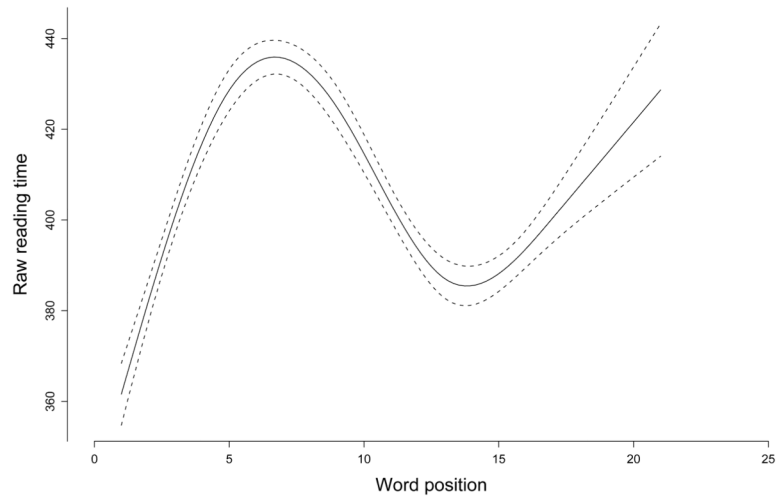


Figure 1. Effect of word position on raw reading times for all stimuli (critical items and fillers) in Experiment I. Dotted lines indicate confidence intervals.

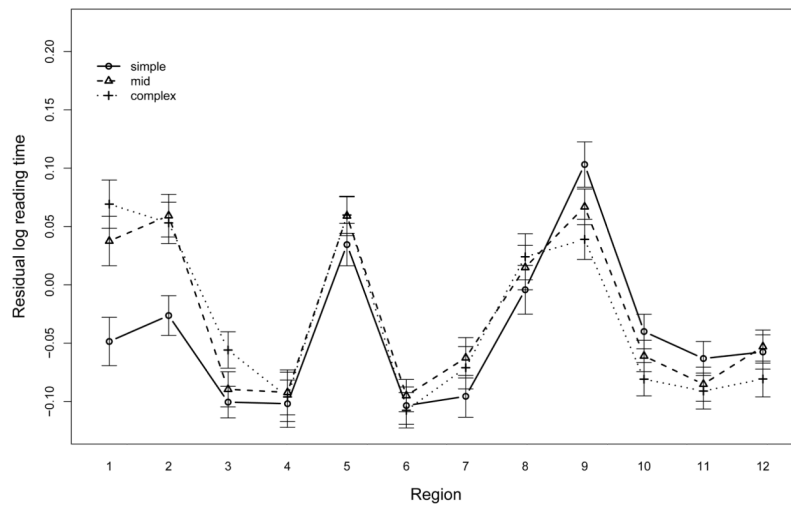


Figure 2.

Word-by-word mean residual log reading times in experiment I from clefted head noun to four words after the subcategorizing verb (communist₁ who₂ the₃ members₄ of₅ the₆ club₇ banned₈ from₉ ever₁₀ entering₁₁ the₁₂). Error bars show (+/-) one standard error.

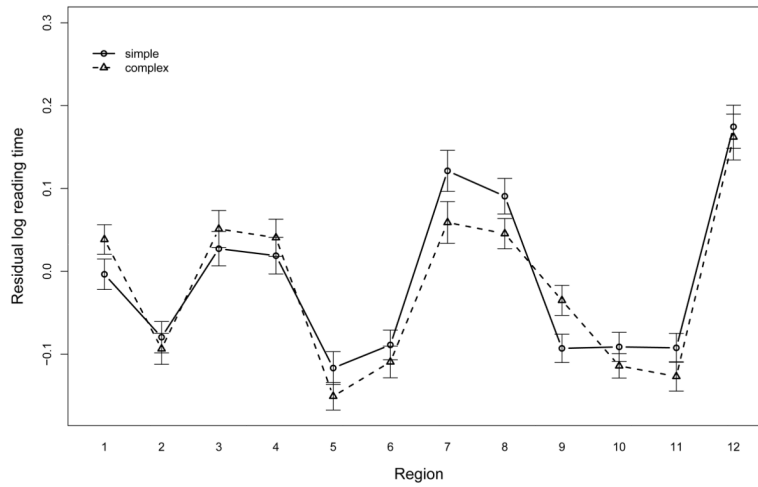


Figure 3.

Word-by-word mean residual log reading times in experiment II from the word immediately following the *wh*-phrase to five words after the subcategorizing verb (the₁ commander₂ that₃ was₄ deeply₅ respected₆ ordered₇ to₈ scout₉ the₁₀ area₁₁ ahead₁₂). Error bars show (+/-) one standard error.

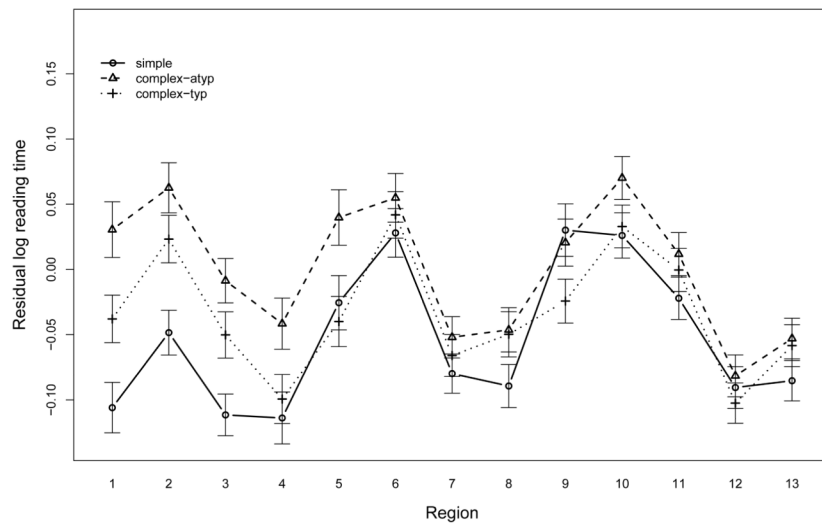


Figure 4. Word-by-word mean residual log reading times in experiment III from object head noun to four words after the subcategorizing verb (dictator₁ who₂ the₃ activist₄ looking₅ for₆ more₇ contributions₈ encouraged₉ to₁₀ preserve₁₁ natural₁₂ resources₁₃). Error bars show (+/-) one standard error.

Table 1

Corpus results from Gigaword for COMPLEX-TYP & COMPLEX-ATYP conditions: mean frequencies of adjectives, the mean conditional probability of the adjective preceding the head noun ($P(\text{adj}|\text{noun})$) and the mean conditional probability of the head noun following the adjective ($P(\text{noun}|\text{adj})$).

	MeanFreq.ofAdj.	P(adj noun)	P(noun adj)
COMPLEX-TYP	127613.17	.00377	.00359
COMPLEX-ATYP	86544.87	.00012	.00063
<i>Ratio</i>	1.475:1	31.176:1	5.697:1