Visual Arguments

Julie E. Boland

University of Michigan

Address correspondence to:

Julie Boland
Dept of Psychology
University of Michigan
525 E. University
Ann Arbor, MI 48109-1109

Phone: (734) 764-4488
Fax: (734) 764-4488
Email: jeboland@umich.edu

Abstract

Three experiments investigated the use of verb argument structure by tracking participants' eye movements across a set of related pictures as they listened to sentences. The assumption was that listeners would naturally look at relevant pictures as they were mentioned or implied. The primary hypothesis was that a verb would implicitly introduce relevant entities (linguistic arguments) that had not yet been mentioned, and thus a picture corresponding to such an entity would draw anticipatory looks. For example, upon hearing *...mother suggested...*, participants would look at a potential recipient of the suggestion. The only explicit task was responding to comprehension questions. Experiments 1 and 2 manipulated both the argument structure of the verb and the typicality/co-occurrence frequency of the target argument/adjunct, in order to distinguish between anticipatory looks to arguments specifically and anticipatory looks to pictures that were strongly associated with the verb, but did not have the linguistic status of argument. Experiment 3 manipulated argument status alone. In Experiments 1 and 3, there were more anticipatory looks to potential arguments than to potential adjuncts, beginning about 500 ms after the acoustic onset of the verb. Experiment 2 revealed a main effect of typicality. These findings indicate that both real world knowledge and argument structure guide visual attention within this paradigm, but that argument structure has a privileged status in focusing listener attention on relevant aspects of a visual scene.

Keywords: argument structure, parsing, thematic roles, sentence comprehension

Visual Arguments

A verb's argument structure specifies the syntactic and semantic constraints on its arguments via constructs such as subcategorization frames and thematic roles. A large body of literature has demonstrated that verb argument structure and other aspects of combinatory lexical knowledge play an important role in guiding sentence comprehension (e.g., Boland, Tanenhaus, & Garnsey, 1990; Boland, Tanenhaus, Garnsey, & Carlson, 1995; Ferreira & McClure, 1997; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Konieczny, Hemforth, Scheepers, & Strube, 1997; Stowe, Tanenhaus, & Carlson, 1991; Trueswell, Tanenhaus, & Garnsey, 1994; Trueswell, Tanenhaus, & Kello, 1994). For example, Mauner and Koenig (2000) found a contrast between passive and intransitive verb forms in sentences like those in (1). In a word-by-word stop-making-sense task, readers began to indicate that (1b) did not make sense at the word, *shut*. This is because the purpose clause (*To reduce…*) sets up an expectation for an agent that must be satisfied in the next clause. The semantic argument structure of the passive verb specifies an agent, so (1a) is fine; whomever shut the door is the person who wanted to reduce the noise. However, there is no such agent in the argument structure of the intransitive form, so (1b) is unacceptable. In earlier research, using *spray/load* verbs, Carlson and Tanenhaus (1988) also found that verb argument structure can implicitly introduce an entity into the discourse. This phenomenon is central to the experiments presented here.

(1). a. *To reduce the noise…the door <u>was shut</u> with great force.* (passive)

b. *To reduce the noise… the door <u>had shut</u> with great force.* (intransitive)

Both Mauner and Koenig (2000) and Carlson and Tanenhaus (1988) assumed that lexical access of the verb makes available the relevant set of thematic roles (characterizing the participants in the event described by the verb), and that these thematic roles are the mechanism by which an entity is introduced into the discourse. This type of view raises important questions about what counts as an argument and about the nature of thematic role information—both topics of current debate.

Most linguistic theories maintain a distinction between arguments, which are essential participants that must be lexically specified, and adjuncts, which are constrained by global (i.e., non-lexical) syntactic and semantic principles. For example, in the sentence *Chris recommended a movie to Kim in the hallway*, both *a movie* and *to Kim* are arguments of *recommend*, but *in the hallway* is an adjunct. Just hearing *Chris recommended...* implies that Chris recommended <u>something</u> <u>to someone</u>, but does not imply a <u>somewhere</u>. Most events do occur somewhere, but a location is not typically part of the verb's argument structure. Thus under traditional accounts, argument structure consists of syntactic and semantic specifications for a small set of participants in the situation described by the verb. Nonetheless, there is considerable debate in linguistics about what counts as an argument, with instruments (e.g. *sweep with a broom*) being a particularly contentious case (e.g., Carlson & Tanenhaus, 1988; Ferretti, McRae, & Hatherall, 2001; Jackendoff, 1987; Marantz, 1984; Van Valin & Lapolla, 1997).

In recent years, there has been a trend within linguistics toward increased lexicalization of syntactic knowledge. Under some approaches, adjunct slots are specified lexically, blurring or eliminating the distinction between arguments and adjuncts (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Spivey-Knowlton & Sedivy, 1995). Under the MacDonald et al. account, arguments and adjuncts are both lexically specified, but tend to lie at opposite ends of a frequency continuum: Phrases that are traditionally classified as arguments occur more frequently with their heads than phrases that are traditionally classified as adjuncts. Note that verbs can only have a fixed number of arguments, but the number of adjuncts is unbounded, so if adjuncts are lexically encoded, the number of entities that are referenced in the lexical entry increases dramatically.

McRae, Ferretti, & Amyote (1997) maintained that every verb has a unique set of thematic roles specifying the prototypical participants for that verb, based upon situational schema that are formed through everyday experiences For example, a criminal is a poor agent, but a good patient, of the verb *arrest*. This Situation-Based view of thematic roles blurs the distinction between linguistic and real world knowledge. Nonetheless, the distinction between

arguments and adjuncts is clear in the conclusions of Ferretti et al. (2001). They demonstrated that isolated verbs prime typical agents (*arresting, cop*), patients (*arresting, criminal*), and instruments (*stirred, spoon*), but not locations (*swam, ocean*). Ferretti et al. concluded that agents, patients, and instruments are arguments while locations are not, based on the criterion that a phrase is an argument if a verb's lexical representation specifies semantic information about it.

The Situation-Based approach can be contrasted with a more traditional view of thematic roles, in which thematic roles specify the generic mode of participation (e.g., agent, theme) for each argument, without incorporating detailed situation-based knowledge (e.g., Jackendoff, 1972). Under the traditional approach, the line between linguistic and real world knowledge is fuzzy, but it is basically intact. For example, *The criminal arrested the ...* would not constitute a thematic role violation—it is merely implausible in our world. If we're discussing an altered world (e.g., *In my dream, a criminal arrested a clown.*) there's no semantic anomaly. Such examples provide some evidence that facts about the world can be distinguished from formal constraints within the language.

In this paper, thematic roles are assumed to be verb-specific, as in McRae et al. (1997). Thus, *dog* satisfies the agent requirements of *attack* but not *arrest*. Both verbs require animate entities in the agent role, but *arrest* also requires that the agent be a human or human-like entity. However, I depart from the McRae et al. approach by maintaining a distinction between thematic role constraints within the language and facts about the world that make thematic role assignments more or less plausible. Thus, as in traditional approaches, *criminal* satisfies the agent requirements of both *attack* and *arrest*, even though *criminal* is more plausible as the agent of *attack*.

Within the literature on sentence comprehension, the question of how and when real world knowledge is integrated with linguistic knowledge is hotly debated, and it is closely linked to the question of how and when thematic role knowledge is used. Some evidence suggests that the use of thematic knowledge is delayed, relative to syntactic knowledge (e.g., Ferreira &

Clifton, 1986; McElree & Griffith, 1995), while other evidence suggests that thematic knowledge guides syntactic ambiguity resolution (e.g., Trueswell et al., 1994) and that a verb's thematic roles are assigned to available discourse entities as soon as the verb is recognized, without the mediation of syntactic constituents (Altmann, 1999). If one adopts both the Situation-Based view of thematic roles and the view that thematic role information guides syntactic and semantic analysis, then one must conclude that real world knowledge is integrated with linguistic knowledge as soon as words are recognized, if not before.

Unfortunately, most of the research on the use of lexical information in sentence comprehension is based on reading. Reading paradigms have the advantage of providing a nearly continuous measure of local processing load, but they have the disadvantage of imposing a secondary skill (reading) between the cognitive process of interest and the dependent measure. As we will see, the choice of a reading or spoken language paradigm can force the adoption of somewhat different perspectives on the role of real world knowledge and co-occurrence frequency in language comprehension. The current study investigates the use of verb argument structure and real world knowledge in a spoken language paradigm.

Research on spoken language comprehension has shown that eye fixations on relevant objects are time-locked to lexical access of isolated words (e.g. Allopenna, Magnuson, & Tanenhaus, 1998), identification of referents for syntactically ambiguous phrases (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Novick & Trueswell, 2001), and pronoun resolution in discourse context (Arnold, Eisnband, Brown-Schmidt, & Trueswell, 2000). That is, we tend to look at objects and people as they are mentioned. In that sense, measuring eye movements in a listening paradigm can provide an index of referential success or allow us to infer the resolution of a syntactic ambiguity once the linguistic input has been comprehended. However, it is not clear that eye fixations are time-locked to developing syntactic and semantic representations in a manner that can be distinguished from the listener's ultimate conceptual representation of the linguistic input. For example, suppose there were an intermediate set of representations built using a only a restricted body of linguistic knowledge, and real world

knowledge was used to select the best candidate (as in Boland, 1997a). Could this research paradigm be used to investigate those intermediate representations or only the final output of language comprehension?

In a promising study, Altmann and Kamide (1999) recorded eye movements as people listened to a sentence like *The boy will move/eat the cake* and looked at a semi-realistic scene with a boy, a cake, and other moveable but inedible objects. Altmann and Kamide found that people looked at the cake more quickly upon hearing *eat* compared to *move*. In fact, participants often looked at *cake* in the *eat* condition prior to the onset of the noun. Altmann and Kamide concluded that the verb's thematic roles were used to pro-actively restrict the domain of subsequent reference. In other words, those anticipatory looks reflected the rapid activation of combinatory linguistic knowledge, coincident with verb recognition. However, there are other accounts of this finding that do not invoke the verb's argument structure or thematic roles. For example, if other objects associated with eating (silverware, a table, etc.) had been depicted, they might have drawn anticipatory looks as well, even if they were not possible arguments of the verb. If so, the anticipatory looks would be more akin to associative priming than an argument structure effect. As such, they would not be directly related to the developing structural representation of the linguistic input. Thus, the implications of the anticipatory looks are not clear in terms of the debate over the integration of linguistic and real world knowledge. An essential question is whether the anticipatory looks are limited to arguments.

Altmann and Kamide (1999) themselves discussed the possibility that real world plausibility formed the basis for their findings. In more recent work, Kamide, Altmann, and Haywood (2003) found that the anticipatory looks were guided by combinatory semantics and real world plausibility, rather than simple lexical relationships. For example, when viewing a carnival scene, listeners were more likely to look at a motorcycle upon hearing *The man will ride…* and more likely to look at a merry-go-round upon hearing *The girl will ride….*

To understand how linguistic and real world knowledge are integrated during sentence comprehension, plausibility and linguistic knowledge must be unconfounded. One way to do this

is to manipulate typicality for both arguments and non-arguments. If verb argument structure alone is driving anticipatory eye movements, then listeners should not look at a bed upon hearing *The girl slept. Bed* cannot be an argument of *slept,* though it is a perfectly good adjunct. In fact, beds are part of a prototypical sleeping event and are thus conceptually and episodically related to *sleep*. Linguistic co-occurrence frequency is relatively high and the co-occurrence of sleeping and a bed in participants' actual experience is likely to be extremely high. Therefore, if linguistic or real world co-occurrence are driving the anticipatory looks, then listeners should be just as likely to look at typical locations for intransitive verbs as they are to look at potential arguments, such as the direct objects in the experiments by Altmann and colleagues.

The experiments in the current paper were designed to investigate how linguistic and real world knowledge are used during spoken language comprehension. To accomplish that goal, and to test alternative accounts of the findings from Altmann and colleagues, both argument structure and real world typicality were manipulated in Experiments 1 and 2. In Experiment 3, only argument structure is manipulated.

Experiment 1

The basic paradigm used by Altmann and Kamide (1999, Experiment 2) was adopted here, with a few modifications inspired by related research. Experimenters have found that a listener's gaze tends to track a spoken sentence when pictures/objects are available that correspond to entities mentioned in the sentence. This appears to be true whether the pictures/objects are arranged in a semi-realistic scene (as in Arnold et al., 2000, or Altmann & Kamide) or if they are presented in a grid (as in Cooper, 1974). In most previous experiments, the stimuli have been simple line drawings (Altmann & Kamide; Cooper, 1974), familiar cartoon figures (Arnold et al.), or actual objects (Eberhard et al., 1995; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Tanenhaus et al., 1995). In the current study, photographs of actual people and objects were arranged them in a four-quadrant grid. The pictures and their placement are described below.

As shown in Examples (2) – (4), both verb type and typicality of the target NP/picture were manipulated. The words referring to the target pictures are underlined. In the intransitive condition (2), the verb introduces no arguments. Therefore, the argument structure hypothesis predicts that attention will remain on the sentential subject until the location is mentioned. The target object is a typical or atypical location for the event.

(2)   Intransitive/Location. *The girl slept for a while on the bed/bus this afternoon.* (pictures: girl, bed/bus, pillow, toy car)

In the action verb condition (3), the verb requires a theme, but a plausible theme (a donkey) was already introduced, and in fact linguistically focused, at the beginning of the sentence. The target objects are a typical and atypical instrument for the action. Instruments represent a borderline case in the argument/adjunct distinction, patterning like arguments on some linguistic tests and like adjuncts on others. Some recent research suggests that some action verbs take instruments while other, quite similar action verbs, do not (Koenig, Mauner, & Bienvenue, 2002). The instruments in the current study are mainly adjuncts, using the Koenig et al. criteria. If the instruments in the current study are appropriately classified as adjuncts, the verb's argument structure does not introduce a new entity into the discourse. In this case, the listeners' attention may be divided between the direct object and subject until an instrument is mentioned.

(3)   Action/instrument. *The donkey would not move, so the farmer beat it vigorously with a stick/hat every day.* (pictures: donkey, farmer, stick/hat, grass)

In the dative condition (4), the verb requires both a theme and a recipient. A plausible theme was focused in the first clause, but an abstract recipient is introduced by the verb: Because suggestions are offered to someone, listeners should look for potential recipients as soon as they recognize the verb. Thus, the target objects are a typical and atypical recipient. This is the only verb condition in which the target objects clearly represent an argument of the verb.

(4)   Dative/Recipient. *The newspaper was difficult to read, but the mother suggested it anyway to her <u>teenager/toddler</u> last week.* (pictures: newspaper, mother, teen/toddler, dictionary)

The argument structure hypothesis predicts a main effect of verb type on looks to the target pictures, with more looks to dative recipients than intransitive locations. A verb type effect exhibiting this pattern will be called an "argument status effect." If the instruments are adjuncts, then they should pattern with the locations. The argument structure hypothesis also predicts effects of verb type on looks to the sentential subject, with the most looking time in the intransitive condition where there are no other pictured arguments.

Alternatively, participants might use real world knowledge to look at objects related to the verb, regardless of linguistic status. This hypothesis predicts a main effect of typicality for target pictures, with more looks to typical arguments and adjuncts alike.

Methods

Participants. Twenty undergraduates at University of Louisiana at Lafayette participated in the experiment, in exchange for extra credit in an introductory Psychology class.

Sentence Stimuli and Associated Norms. Sixteen items were constructed for each of three verb conditions: intransitive/location, action/instrument, and dative/recipient. Thus, there were 48 critical trials in the experiment. There were no filler trials. The sentence materials are provided in Appendix A. In each critical sentence, the target location, instrument or recipient was embedded within a prepositional phrase (PP). The critical PP always occurred after the main verb. In the action and dative conditions, the PP occurred after the direct object, which was pronominalized to prevent attachment ambiguities.

The intransitive verbs were taken from Levin's (1993) classifications of verbs involving the body (n = 7), appearance, disappearance, and existence (n = 5), light emision (n = 2), animal sounds (n = 1), and manner of motion (n = 1). These verbs do not take a noun phrase (NP) in

object position, except in certain highly restricted circumstances (cognate or reaction object, postverbal subject with either *there* insertion or locative inversion). These verbs occur frequently with locative PP's, but the PP's behave like adjuncts on linguistic tests. Thus, the intransitive condition represents a clear case in which the (locative) target is an adjunct. The locative PP's in the current stimuli are of two basic types. For nine of the items, the location specified the context for the event described by the verb (e.g., *dozed near the fire*). For the other seven items, the location specified the target of the event (e.g., *coughed into his hand, laughed at the clown*).

The dative items represent a clear case in which the (recipient) target is an argument. All but one verb (*suggest*) allowed the double object/prepositional object alternation. The verbs were taken from Levin's (1993) classifications of verbs of communication (n = 8), sending and carrying (n = 4), and change of possession (n = 4).

The action/instrument verbs represent a borderline case in the argument/adjunct distinction. The verbs were taken from Levin's (1993) classifications of verbs of creation and transfomation (n = 4), contact by impact (n = 6), removing (n = 4), killing (n = 1), and ingesting (n = 1).

Two versions of each item were constructed, one with a typical target and one with an atypical target. Examples were given in (2) – (4) above, with the targets underlined. Both typical and atypical targets had to meet the minimum semantic constraints that the verb placed on the participatory role in question. For example, both rocks and spears are potential instruments for killing something, but a rock would be more effective on a bug and a spear would be more appropriate for a tiger. Thus, the typicality of the target was a function of the entire [Subject + Verb (+ direct object)] event. The combinatory nature of the typicality manipulation does not present a complication, because Kamide et al. (2003) have demonstrated that listeners' anticipatory looks are sensitive to combinatory semantics.

The typical and atypical versions of each sentence were rated for acceptability on a scale from 1(very awkward) to 7(very natural). The ratings were completed by two groups of ten students from the same population as those participating in the experiment. Each group rated 48

items, half typical and half atypical, randomly ordered. Sentences were presented in written form and participants circled the appropriate rating on a numerical scale. Each student rated only one version of each item. There were no filler items. The results are summarized in Table 1. Item means were submitted to a 3(verb type) by 2(typicality) ANOVA, which revealed a main effect of typicality [$F2(1,45) = 81.32$, $p < .01$], with no main effect of verb type and no interaction. Sentences with atypical targets were judged less acceptable, regardless of verb type.

Table 1. Mean Acceptability Ratings by Condition.

	Location	Instrument	Recipient
Typical	5.5	4.5	4.9
Atypical	3.3	3.2	2.9

To evaluate the proposal that arguments and adjuncts differ only in terms of the co-occurrence frequency between the head (verb) and the dependent phrase (argument or adjunct), two types of co-occurrence data were collected from the World Wide Web (WWW). Because the dependent measure in the experiment is an anticipatory look to a photograph, prior to hearing the target phrase, the frequency measures that are often used in parsing/reading research are inappropriate. In reading studies, the co-occurrence frequency between a verb and a phrase of a particular class (e.g., a PP beginning with *to* or a phrase that is assigned a particular thematic role) is often used to predict processing difficulty for a phrase of the same class (e.g., Garnsey et al., 1997; Novick & Trueswell, 2001). In contrast, we need to know, given a particular verb, the probability of an object or person like the one in the target picture.

The first analysis used Google to search only English web pages. The total number of web page hits for each of the critical verbs was used as the denominator to compute the proportion of those web pages that also mentioned the target (typical or atypical) noun. For

example, there were 716,000 hits for *slept* and 273,000 of those pages also mentioned *bed*, so the co-occurrence proportion is .38. Note that this proportion does not necessarily reflect co-occurrence in a sentence, or even in a paragraph. And for verbs like *hit* that are homographic with nouns, non-verb tokens are included in the denominator. Another concern is the dynamic nature of the WWW database—the stability of these numbers has not been verified. Despite these limitations, the proportions provide a rough estimate of the strength of the conceptual relationship between each verb and its target pictures. These data are summarized in Table 2a. A 2(typicality) by 3(argument structure) ANOVA revealed main effects of both verb type [$F_2(2,90)$ = 6.92, p < .01] and typicality [$F_2(1,90)$ = 5.64, p < .05], with no interaction. Overall, locations were most likely and instruments were least likely to co-occur with their verbs, and typical targets were more likely than atypical targets. The underperformance of instruments on this measure may be due to the fact that five of the sixteen instrument items had a noun/verb homograph as the verb, the target noun, or both. This was true for only one item in each of the location and recipient sets.

Table 2a. Mean (and standard deviation of) co-occurrence proportion between verbs & targets on WWW.

	Location	Instrument	Recipient
Typical	.18 (.17)	.05 (.05)	.11 (.08)
Atypical	.09 (.10)	.06 (.08)	.04 (.05)

The data in Table 2b represent the proportion of matching verb phrases (VPs) in a Lycos exact phrase search. The adverbs used in the critical stimuli were ignored in counting tokens of the critical verbs used with the critical PP, both with and without the pronoun. For example, when searching on the typical version of the *suggest* item illustrated in (4), the number of hits for *suggested to her teenager* and *suggested it to her teenager* were summed and then divided by the total number of hits on *suggested*. Thus, the proportions in Table 2b represent only those co-

occurrences with the same structural configuration as in the critical stimuli. There was no effect of verb type [$F2(2,90) = 1.94$, $p > .10$], but there was a main effect of typicality [$F2(1,90) = 5.51$, $p < .05$], with more hits for typical VPs. The two variables did not interact. Taken together, the analyses of co-occurrence frequency indicate that co-occurrence is confounded with typicality rather then verb type.

Table 2b. Mean proportion (st dev) of verb tokens heading the critical verb phrase on WWW.

	Location	Instrument	Recipient
Typical	.000328	.000092	.000093
	(.000646)	(.000185)	(.000226)
Atypical	.000036	.000005	.000039
	(.000108)	(.000011)	(.000143)

The sentences were recorded by a female assistant in the local southern Louisiana accent. The recordings were digitized and excess silence was removed from the beginning and end of each file. The typical and atypical versions of each item were then compared to ensure that the prosody was consistent across the two conditions. Measurements were taken to find the acoustic onset of the main verb, the PP, and the target noun. Table 3 provides the mean onsets, relative to the verb onset, of the PP and the target noun for each item group. The intransitive location set had the shortest inter-stimulus-interval (ISI) and the dative recipient set had the longest. Participants received only one version of each item, and equal numbers of typical and atypical items within each verb type. Only typicality was manipulated within items.

Table 3. ISI (ms) between onset of verb, onset of PP, & target noun.

	Location	Instrument	Recipient

PP onset	1433	1515	1672
NP onset	1794	1985	2189

Picture Stimuli and Associated Norms. Four pictures illustrated each story. In most cases, each picture was a photograph; in a few cases it was a realistic cartoon. Each of the four pictures was placed in a quadrant of a PowerPoint slide, and enlarged to fill as much of the quadrant as possible. The background was gray. Each picture was used in one or two experimental items. If a picture was used in two items, it was not used in the same condition on both items.

The picture in the upper left quadrant was the first-mentioned NP.[1] This entity was usually in primary focus. For the intransitive items, it was the subject of the sentence and was most often a person. For the dative and transitive sentences, the first mentioned NP was coreferential with the direct object pronoun in the main clause. The target picture appeared in one of the other three quadrants (sixteen times in each quadrant). Atypical and typical conditions differed only in the picture used for the target, which matched the location/instrument/recipient in the auditory sentence. The other two pictures either depicted other entities from the sentence or were neutral filler pictures. Brief descriptions of all of the pictures, along with the quadrant in which they appeared, are included in Appendix A.

A norming study was conducted to evaluate partipants' expectations that a pictured entity would be mentioned. Ten participants from the University of Michigan community each completed a questionnaire accompanied by a slide show. The questionnaire consisted of sentence fragments for the participant to complete. The sentence fragments were written versions of the experimental sentences, except that they were truncated just prior to the prepositional phrase that mentioned the target object. For example, the sentence fragment for the item in (2) was *The girl slept for a while…*. Each sentence fragment was accompanied by the slide used in Experiment 2. As explained later, these slides differ from those used in Experiment 1 only in that both typical

---

[1] There was one dative item in which the first-mentioned NP was *idea*, so it deviated from the normal pattern.

and atypical targets are pictured on each slide. Participants were told that the pictures were presented to make the sentences more interesting. They were not explicitly advised to use the pictures in their completions, but they were told to look at the pictures before completing each sentence.

The responses were coded by an undergraduate assistant. References to the typical and atypical target pictures were coded as "hits" regardless of the role they played in the sentence and regardless of the specific lexical label used. Only explicit NP references were counted. For example, an instrument that was incorporated into a verb (e.g., nailed, glued) was not counted. If both typical and atypical targets were mentioned, both were coded as a "hit". Because the critical verb was part of the sentence fragment that was provided to the participants, its argument structure was expected to be an important factor governing expectations about which pictures were most relevant. As shown in Table 4, argument structure was important for typical targets only. Participant and item means were submitted to a 3(verb type) by 2(typicality) ANOVA. There was an interaction of verb type and typicality [$F1(2, 18) = 10.46$, $F2(2,45) = 5.77$, $p < .01$], as well as a main effect of each variable [verb type: $F1(2, 18) = 7.97$, $F2(2, 45) = 5.72$, $p < .01$; typicality: $F1(1, 9) = 20.42$, $F2(1, 45) = 40.41$, $p < .01$].

Table 4. Percentage of Target Picture Completions

	Location	Instrument	Recipient
Typical	.36	.46	.68
Atypical	.19	.21	.16

Procedure & Equipment. The auditory sentences and their corresponding pictures were presented in a fixed random order. There were two versions of the stimulus list, each having one version of each item and equal numbers of typical and atypical items for each verb type. Half of the participants were given each list. There were four practice sentences at the beginning of the stimulus list to acquaint the participant with the flow of events.

Eye position was measured as participants listened to each sentence, using an ISCAN ETL-500 Pupil/Corneal Reflection Tracking System. The eye and scene cameras were mounted on a baseball cap, which participants wore throughout the experiment. The sampling rate was 60 Hz, but all point of regard data was analyzed via frame-by-frame examination of a video-tape (of the scene image, superimposed eye position, and all auditory stimuli), providing an effective sampling rate of 30 Hz. Video recording and coding used a SONY DSR-30 frame-accurate digital video recorder. A five-point calibration procedure was conducted before beginning the experiment proper. Data acquisition software was controlled via a 200 MHz Pentium.

The auditory and visual stimuli were presented within a PowerPoint presentation on a separate computer. Participants were told that they would listen to sentences and answer questions about them; pictures would be presented to make the sentences more interesting. Participants were given no specific instructions about looking at the pictures, nor were they required to point at or make a judgment about the pictures. At the beginning of each trial, the PowerPoint screen was incremented to show the four pictures and begin auditory presentation of the sentence for that trial. Simultaneously, the experimenter pressed a button to begin data collection on the eye-tracker. Data collection terminated automatically after ten seconds, time enough for completion of each sentence. After the ten seconds elapsed, the experimenter pressed a button to present a visual comprehension question, if applicable. Yes/no comprehension questions occurred after 20 of the sentences. Participants responded to the comprehension questions verbally. The next trial began when the experimenter pressed the keys to increment the PowerPoint screen and begin data collection from the eye tracker again. The entire experiment, including calibration and practice trials, took less than thirty minutes.

During debriefing, participants were encouraged to give their impressions of the experiment. They usually thought the experimental goal was related to the comprehension questions. No participant guessed that the object of study was the pattern of eye movements precipitated by the verb or the typicality manipulation.

Results

None of the participants missed more than one comprehension question. Two undergraduate coders examined the videotaped data files, frame by frame. The VCR's synchronized audio and video channels allowed the coder to detect the frame containing the acoustic onset of the verb in each trial. Two seconds worth of data from verb onset were analyzed. The participant's gaze was coded as a "hit" on a particular picture if a fixation was either on the picture or judged to be within a few pixels of its edge. For the figures and the analyses, looks to a given picture were not counted if the participant had maintained gaze on that picture since verb onset or before.

The probability of a look to the target picture in each of the six conditions is illustrated in Figure 1. Looks initiated at, or prior to, verb onset were excluded. Beginning about 500 ms after the onset of the verb, it appears that participants were more likely to be looking at recipients than either locations or instruments. There also appears to be a preference for looking at typical, as opposed to atypical locations about 600 to 1000 ms after verb onset, with an advantage for typical instruments arising slightly later. In contrast, there is never a tendency to look at typical recipients more than atypical recipients. The observed pattern, in which people tended to look at recipients more than instruments and locations, and also tended to look at typical adjuncts more than atypical adjuncts, is consistent with a view in which argument structure and real world plausibility both guide expectations during sentence understanding.

Figure 1. Proportion of trials on which participants are looking at the target in each condition, in 100 ms bins. The labels on the X-axis represent the starting point of each bin.

The interval 500 – 1000 ms after verb onset was selected for analysis. Based on Figure 1, this is the earliest region in which an argument status effect might be observed. This interval ends prior to the onset of the target noun for all items; thus, looks during this interval are clearly anticipatory. The probability of making at least one look to the target during this window was computed by participants and by items and subjected to 2(list) by 3(verb type) by 2(typicality) ANOVAs. The means are provided in Table 5. There was a main effect of verb type [$F1(2, 36)$ = 3.71, $F2(2, 42)$ = 3.93, p < .05], with dative recipients garnering the most looks. There was no effect of typicality [$F1(1,18)$ = 1.40, $F2(1, 42)$ = 1.53, p > .10], nor was there an interaction between

verb

type
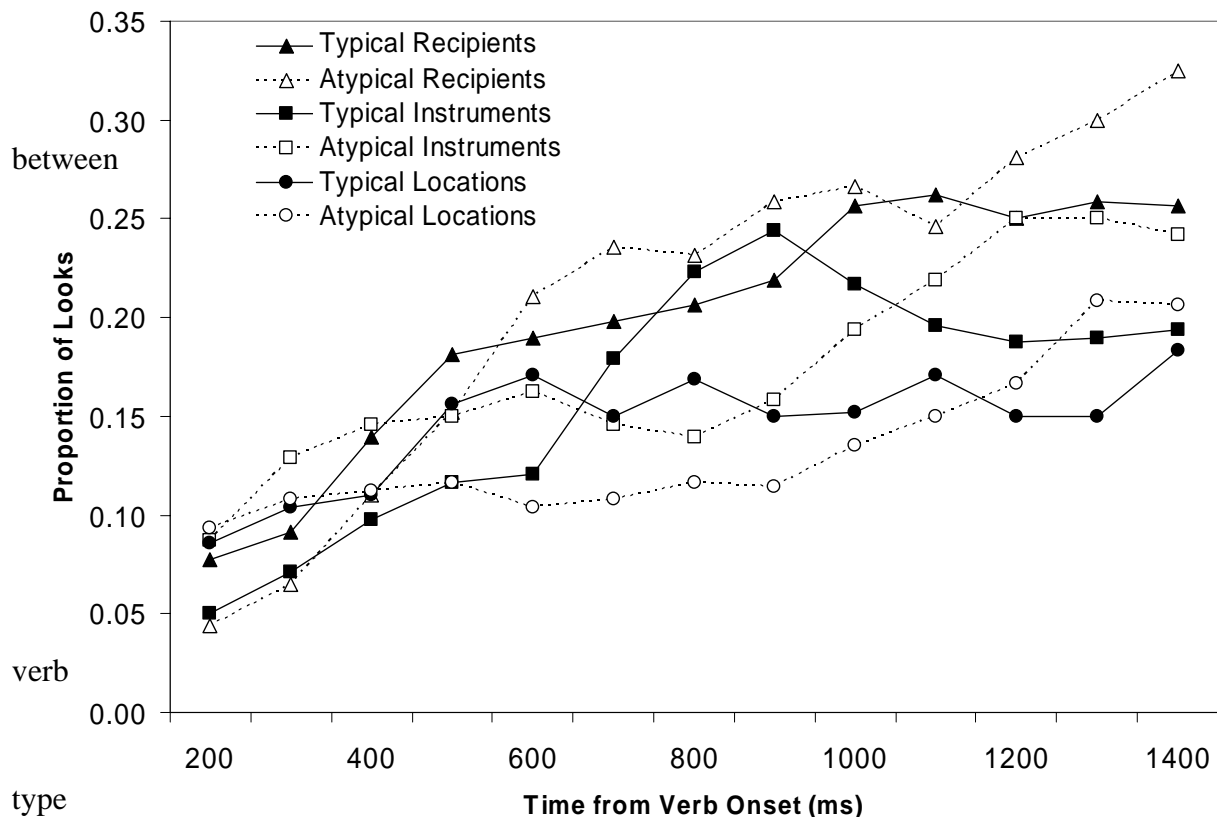
and typicality [Fs < 1.0]. In planned two-tailed comparisons (α = .05, by participants and by items), the atypical recipients received more looks than the atypical locations. The contrast between atypical recipients and atypical instruments was reliable by items and marginal by participants. Atypical instruments and locations did not differ from each other. There were no reliable differences in the probability of a look among the typical targets. This pattern suggests

that people looked at the potential recipients, even if it was implausible in the current situation. In contrast, they were unlikely to look at a potential instrument or location unless it was typical of the type of event being described.

Table 5. The probability of a look to the target picture 500 – 1000 ms after verb onset.

	Location	Instrument	Recipient
Typical Target	0.29	0.34	0.39
Atypical Target	0.24	0.27	0.39

The proportion of looks to typical targets was correlated with the likelihood of mention from the illustrated sentence completion norms summarized in Table 4 [$F2(1,46) = 6.49$, $p < .05$, $r^2 = .12$]. However, looks to atypical targets were uncorrelated [$F2(1, 46) = 1.51$, $p > .05$]; although atypical recipients received more looks than atypical locations and instruments, they were no more likely to be mentioned than atypical instruments and atypical locations.

For more detailed analyses, the data were divided into 100 ms bins, as in Figure 1. The temporal window 500 – 1000 ms after verb onset was again utilized. Within each bin and condition, the average proportion of trials in which the participant was looking at the target was calculated by participant and by item. These values were submitted to 2(list) by 3(verb type) by 2(typicality) by 5(bin) ANOVAs. There was a three-way interaction of verb type, typicality, and bin [$F1(8,144) = 2.73$, $F2(8,168) = 2.54$, $p < .05$]. The expected main effect of verb type was not quite significant [$F1(2,36) = 3.18$, $p = .05$, $F2(2,42) = 2.56$, $p < .10$]. There was no main effect of typicality [$Fs < 1.00$].

The three-way interaction was explored by performing 2(list) by 3(verb) by 5(bin) ANOVA's on each level of typicality. Within the typical condition, there was no main effect of verb type [$Fs < 1.00$], but rather an interaction of verb type and bin in the analysis by participants

only [$F1(8, 144) = 2.50$, $p < .05$; $F2(8, 168) = 1.76$, $p > .10$]. In contrast, the atypical condition produced a main effect of verb type [$F1(2,36) = 3.96$, $F2(2, 42) = 3.61$, $p < .05$]. Verb type interacted with bin [$F1(8, 144) = 2.01$, $p < .10$; $F2(8, 168) = 2.77$, $p < .05$], though the interaction was marginal by participants. This pattern of results is consistent with the planned comparisons from the initial analysis. Atypical recipients received more looks than atypical locations and atypical instruments throughout much of the 500 to 1000 ms window. However during some bins within this window, typical instruments and/or typical locations received just as many looks as typical recipients.

Discussion

The primary finding in Experiment 1 was the main effect of verb argument structure during the temporal window 500 to 1000 ms after verb onset. As predicted by the argument structure hypothesis, recipient arguments received more looks than instruments or locations, prior to mention of the target item. This contrast suggests that verb argument structure provides a mechanism for introducing new entities into the discourse. This result is striking because it suggests a relatively tight link between visual attention and the activation of linguistic representations during sentence comprehension. It is all the more remarkable because robust argument structure effects were found without using any secondary task, such as pointing to the target objects.

Notably, the effect of verb type was strongest among looks to the atypical targets. Thus, while there were no reliable effects of target typicality in the overall analyses, the results suggest that real world associations often prompted looks to likely adjuncts. The influence of real world associations was most apparent in the correlation between the probability of a look to a typical target and the probability of mention in the illustrated sentence completion norms. In contrast, the pattern of looks to atypical targets deviated strongly from the gaze pattern that would have been predicted on the basis of the illustrated completion norms.

It was stated at the outset that instruments were a borderline case within the linguistic categories of arguments and adjuncts. The current data provide some support for this notion. For example, in Table 5, the probability of looking at an instrument falls in between the probability of looking at a recipient and looking at a location.

The tendency to look at typical locations, typical instruments, and both typical and atypical recipients is consistent with a view in which both argument structure and real world knowledge influence gaze, and by implication, online sentence understanding. Nonetheless, the lack of a typicality effect for recipients is difficult to reconcile with prior results such as Kamide et al. (2003), who analyzed looks to (direct object) arguments. Recall that listeners viewing a carnival scene made anticipatory looks to a motorcycle upon hearing *The man will ride the…* and to a carousel upon hearing *The girl will ride the...* If Kamide et al. found typicality effects for the theme of *ride* why shouldn't there be comparable typicality effects for the recipients in the current study?

Experiments using a directed action task have also produced typicality effects (Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Sedivy et al., 1999). However, it is not clear how the requirement to carry out linguistic instructions impacts gaze. The desire to act quickly may increase anticipatory looks to relevant objects and may lead to strategic processing that enhances typicality effects. For example, for some of the critical trials in Chambers et al.'s Experiment 2, participants were instructed to *Put the X inside the can*, with two cans available—only one of which was large enough to contain the first object. By 300 ms after the onset of *can*, participants were making looks to the larger can, indicating that they had integrated the relevant linguistic and nonlinguistic constraints on the instruction. Half of the 60 trials instructed participants to put one object *inside* another (the other half used *beside* instead), so it is not surprising that participants were attuned to the relative size of the containers. In fact, they may have strategically evaluated the containers' sizes with respect to a possible "put inside" action, as soon as they heard the first part of the instruction, *Pick up the X*. It would be informative to determine whether the same rapid integration of non-linguistic knowledge would hold in an experiment that

contained very few *put inside* items and that did not depend upon the participant performing the action.

For purposes of comparison with the current experiment, Kamide et al.'s (2003) passive looking study provides the best example, because a similar task and a similar variety of linguistic stimuli were used. There are at least two possible reasons for the lack of a typicality effect for recipients in the current experiment. First, different pictures were used as typical and atypical targets in the current study, and it is possible that the atypical recipient pictures were more effective at capturing visual attention, thereby washing out a typicality effect for the recipients. Alternatively, the discrepancy may be due to a difference in the way the visual stimuli were presented. In Kamide et al. (2003), hearing "rode" introduced an abstract ride-able object into the discourse. In the visual scene, there were two ride-able objects (a motorcycle and a merry-go-round), so real world knowledge guided the viewer to the most plausible one. In the current experiment, the argument structure of the dative verbs introduced an abstract recipient, but there was only one potential referent pictured. In both the typical and atypical conditions, the potential referent met the lexical constraints on recipients for that particular verb. Therefore, plausibility had no opportunity to play a role if we assume that verb argument structure is first used to identify possible referents.[2] This account is consistent with prior findings that pragmatic constraints influence ambiguity resolution but not the generation of linguistic structure (see Boland, 1997b). This explanation for the absence of a recipient typicality effect was tested in Experiment 2.

Experiment 2

Experiment 2 was designed to further explore the effect of typicality on arguments and adjuncts. The sentence and picture stimuli from Experiment 1 were used again, but with several

---

[2] If verb argument structure and plausibility simultaneously and jointly constrain referent identification, then typicality effects would have been expected in the current experiment even though only one possible referent was pictured. Because the typical target is more strongly associated with the sentence context, it should draw attention more effectively than the atypical target, regardless of whether there is another potential target pictured.

critical changes. First, the sentences were recorded by a new speaker at a substantially faster speech rate. Second, both typical and atypical target pictures were presented on every trial. As discussed above, this provides an opportunity for typicality effects to emerge for recipients, even if verb argument structure initially constrains referent identification.

Presenting both the typical and atypical target pictures on every trial should increase the salience of the target instruments and locations, because half of the pictures will represent concepts within a single category. This salience of the target category may be most obvious for the instruments. Consider the example, *A broken pipe was flooding the laundry room, so the plumber cut it precisely with a saw*…. In Experiment 1, listeners were shown a picture of a pipe, a laundry room, a man in a plumber's uniform, and a saw. As they heard the first words of the sentence—or perhaps even before—they may have deduced that the sentence would be about a plumbing incident. Listeners may even have assumed that the plumber would use the saw to cut a pipe. But there was no reason to focus on the saw in particular. In Experiment 2, the laundry room is replaced by an axe, the atypical target. When the listeners see the pipe, plumber, and two tools, they might deduce that the sentence will be about a man using some tools. If so, as the sentence unfolds, the listener may begin to guess which tool was used for what purpose. In many cases, it will not be clear that the typical and atypical instruments are from the same semantic category (e.g., that both a knife and a pillow can be used as weapons) until well into the sentence. Nonetheless, by the action verb in *A guard was suspicious of a convict who was planning to escape, so the convict attacked...* listeners are likely to be guessing how the knife and/or pillow would be used in the incident being described. The point is that the change in the visual stimuli may alter the listeners' expectations concerning the linguistic stimuli, by drawing attention to the semantic property embodied by the two target pictures.

The predictions concerning typicality effects in Experiment 2 assume that verb argument structure is the most important factor for defining the set of relevant items. This assumption is based, in part, upon the results of Experiment 1. A dative verb requires a recipient, and if there is only one potential recipient pictured (as in Experiment 1), participants should look at it

regardless of whether it is plausible in the current linguistic context. If two potential recipients are pictured (as in Experiment 2), there is a temporary ambiguity concerning which is the actual recipient. Listeners are likely to use world knowledge to resolve that ambiguity in favor of the more typical recipient. Thus, the argument structure hypothesis predicts a recipient typicality effect in Experiment 2, even though such an effect was not predicted in Experiment 1. The argument structure hypothesis does not itself predict typicality effects for potential adjuncts, because the verb does not delimit a relevant semantic domain for potential adjuncts (e.g., places to sleep, instruments to beat with) in the way that it does for potential arguments. However, looks to typical adjuncts might be observed in both experiments due to associative relationships between the verbs or the described events and the typical locations/instruments.

Methods

Participants. Fourteen undergraduates at the University of Michigan participated in the experiment, in exchange for partial credit in an introductory Psychology class.[3]

Materials. The sentence materials and instructions were described in Experiment 1. New recordings of the sentences were produced by a male speaker with a northeastern US accent. Consequently, the speech rate was faster than in Experiment 1. The procedures for digitizing, editing, and measuring critical points in the sentences were identical to those used for Experiment 1. Table 6 provides the mean onsets, relative to the verb onset, of the PP and the target noun for each item group. This table is analogous to Table 3 from Experiment 1.

Table 6. ISI (ms) between onset of the verb, onset of the PP, & target noun.

	Location	Instrument	Recipient
PP onset	997	1031	1130
NP onset	1306	1322	1402

[3] An additional seven participants participated in a music condition that was later dropped. In the music condition, participants looked at the same pictures while listening to music instead of sentences. This condition was dropped upon the advice of a reviewer because it was uninformative.

The picture stimuli were identical to those used in Experiment 1, except that both atypical and typical targets were always presented. The second target picture replaced the filler item for dative and action verb sentences, and replaced one of the two filler pictures in the intransitive sentences. Thus, dative trials usually pictured three animate entities (subject, two potential recipients) and one inanimate (direct object), whereas intransitive and action verb trials usually depicted one animate entity (subject) and three inanimate entities (two potential locations and one filler, or two potential instruments and the direct object). Because both targets were presented on every trial, the slides were identical for the typical and atypical sentence conditions.

Procedures & Equipment. The procedures were roughly equivalent to those of Experiment 1, except that a heavier, high-speed camera was used. This changed our calibration procedures, made calibration more difficult, and led to the shifting of headgear for some participants, as described below.

Three calibration checks were conducted during the experiment to insure accurate tracking throughout. On each calibration check, participants looked at six colored plus signs, arranged in a large pentagon around a central plus sign. Each plus sign was centered within a 2.5 cm circle. The plus signs and circles were presented on the same computer screen as the experimental stimuli. The plus signs and circles also appeared on the experimenter's screen, along with a .75 by 1 cm crosshair representing current gaze location, as recorded by the experimental software. For each participant, there was a set of printouts of the stimulus screen, showing the array of plus signs and their circles. The experimenter drew the location of the fixation cross-hair on the printout, as the participant looked at each plus sign. If the initial calibration was unsatisfactory (meaning that the fixation cross-hairs did not land within the circles), the participant was recalibrated and the new positions of the cross-hair were noted, before beginning the experiment. If a participant could not be calibrated such that four out of six fixation points registered the cross-hair on or within the circle, the participant did not complete the experiment. Ten potential participants were dismissed for this reason. Additional calibration

checks were recorded at the mid-point and the end of the experiment. The data from seven other potential participants were discarded due to a noticeable shift in the headgear. (The experimenter usually noticed this during the experiment because the mid-point or final calibration was problematic. In addition, two research assistants independently examined the data in graphic form (the scan path) using the PRZ analysis software provided by ISCAN® to catch headgear shifts not noted by the experimenter.)

The auditory sentence began playing approximately one second after the mouse press. Fixations were hand-coded from videotape by two undergraduates, as in Experiment 1.

Results

The mean accuracy on comprehension questions was 96%. Participant looks to the target arguments and adjuncts are summarized in Figure 2. The typical and atypical sentence conditions are collapsed, because there was no difference in lexical content prior to the target word. As in Experiment 1, fixations were not included if they began prior to verb onset because they were initiated prior to recovering any information about the verb. In contrast to Figure 1, there is a clear effect of typicality that begins to emerge 400 ms after verb onset and no obvious effects of argument status. Thus, depicting two potential recipients, instruments, or locations on each trial seems to have increased the probability of fixating typical adjuncts and decreased the probability of fixating atypical arguments.

Figure 2. The probability of a look to each target object, 200 – 1500 ms after verb onset.
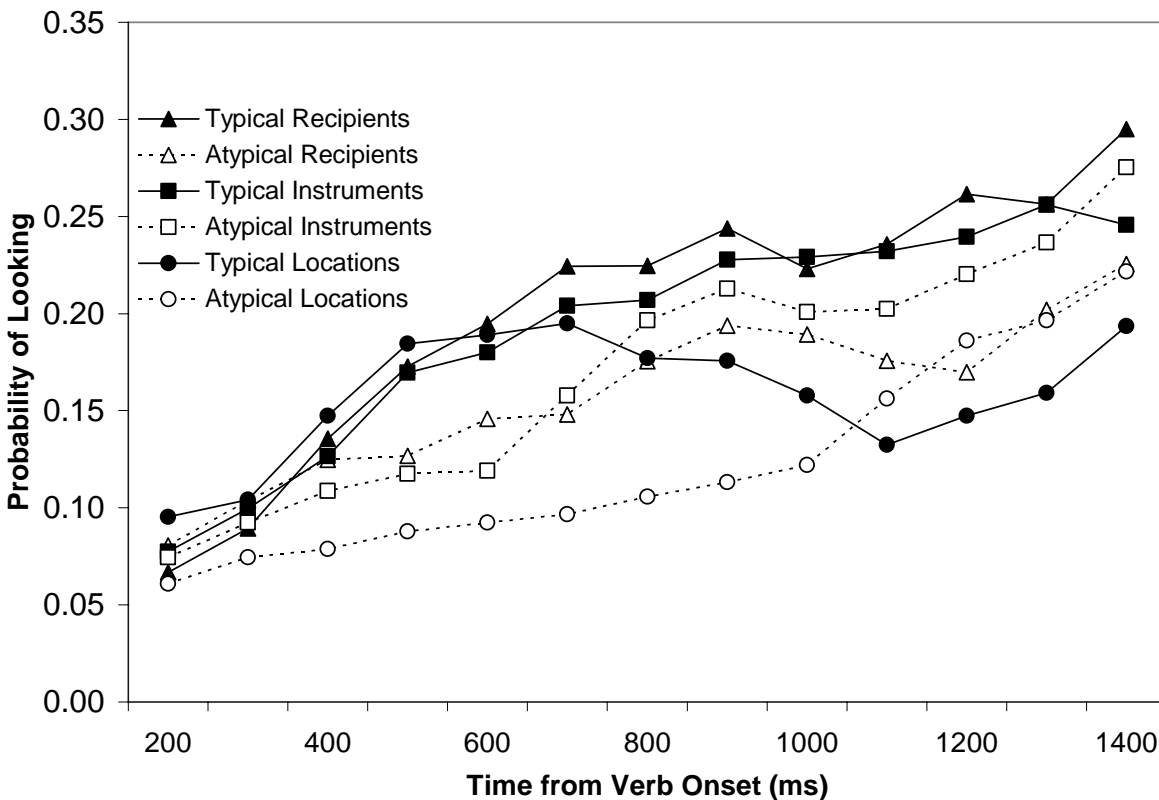
For the initial analysis, the probability of an anticipatory look to each target picture was computed separately by participants and by items for the interval 400 to 733 ms after verb onset[4]. These data are summarized in Table 7. Compared to Experiment 1, the critical interval is shorter and begins earlier, but the interval was chosen using the same criteria as that for Experiment 1. The end of the interval in the current experiment was forced by the onset of the target noun, which began as early as 747 ms after verb onset. The critical PP began as early as 507 ms after verb onset. Assuming that it takes a little more than 200 ms to plan and program a saccade, the current interval does not include looks that were initiated after hearing the beginning of the critical PP.

---

[4] The endpoints of the critical interval reflect the available samples, given that the data were coded from video frames. The effective sampling rate was 30 hz, as in Experiment 1.

Table 7. The probability of a look to each target, 400 – 733 ms after verb onset.

	Locations	Instruments	Recipients
Typical Targets	.30	.32	.32
Atypical Targets	.20	.25	.23

Participant and item means were submitted to a 2(list) by 3(verb type) by 2(sentence typicality) by 2(target typicality) ANOVA. In sharp contrast to Experiment 1, there was a main effect of target typicality [$F1(1,12) = 4.92$, $F2(1,42) = 6.06$, $p < .05$] and no effect of verb type [$Fs < 1.0$]. There was no interaction between the two factors [$Fs < 1.0$]. Thus, upon hearing the critical verb, participants' gaze was drawn to both typical arguments and typical adjuncts. A planned comparison tested the prediction that a typicality effect would be found for recipients in particular, but the contrast was not reliable when the analysis was restricted to recipients alone.

The size of the typicality effect was comparable in the first and second halves of the experiment, with more typical items receiving 8.5% more anticipatory fixations and 10% more, respectively. This is consistent with the assumption that the change in the visual stimuli altered listeners' expectations, rather than listeners developing a strategy of anticipating typical targets over the course of the experiment.

In contrast to Experiment 1, the probability of looking at an atypical target picture was correlated with the probability of mentioning that picture in the illustrated sentence completions summarized in Table 4 [$F2(1,46) = 4.27$, $p < .05$, $r^2 = .08$]. No reliable relation was observed for typical targets [$F2 < 1.0$].

Next, the looking pattern was broken into temporal bins (see Figure 2). The proportion of looks in each cell was computed by participant and item, and the bins beginning 400, 500, and 600 ms after verb onset were used for the analysis. These values were submitted to 2(list) by

3(bin) by 3(verb type) by 2(target typicality) by 2(sentence typicality) ANOVAs. A main effect of target typicality was observed, with typical target pictures receiving the most looks [$F_1(1, 12)$ = 6.19, $F_2(1,42)$ = 4.37, $p < .05$]. There was no main effect of verb type [$Fs < 1.0$], nor did verb type interact with target typicality [$Fs < 1.0$]. Neither verb type nor target typicality interacted with bin [All F's < 1.0, except for target typicality by bin, $F_2(2,84)$ = 2.40, $p > .10$]. Target typicality was not reliable when the analysis was restricted to recipients alone.

In sum, participants were more likely to look at typical targets overall, but the expected recipient typicality effect occurred only as a nonsignificant trend.

Comparison of Experiments 1 and 2

The differences between the looking patterns in Experiments 1 and 2 are illustrated in Figure 3, using all of the temporal bins that were utilized in either experiment. Positive values indicate more looks in Experiment 2 than in Experiment 1. The most striking difference between the two experiments is the 6-8% decrease in looks to atypical recipients in Experiment 2 compared to Experiment 1, 600 to 1000 ms after verb onset. The greater decrease in looks to atypical recipients, compared with the two atypical adjunct conditions, is consistent with the fact that there were relatively few looks to the atypical instruments and locations in Experiment 1. It was only in the recipient (argument) condition of Experiment 1 that there were many looks to both the typical and atypical targets. Thus, this contrast between the two experiments is exactly as predicted by the argument structure hypothesis: verb argument structure was used to identify potential (recipient) arguments and real world knowledge was used to select the most likely candidate.

Figure 3. The bin means from Experiment 2 minus those from Experiment 1, 400 to 1000 ms after verb onset
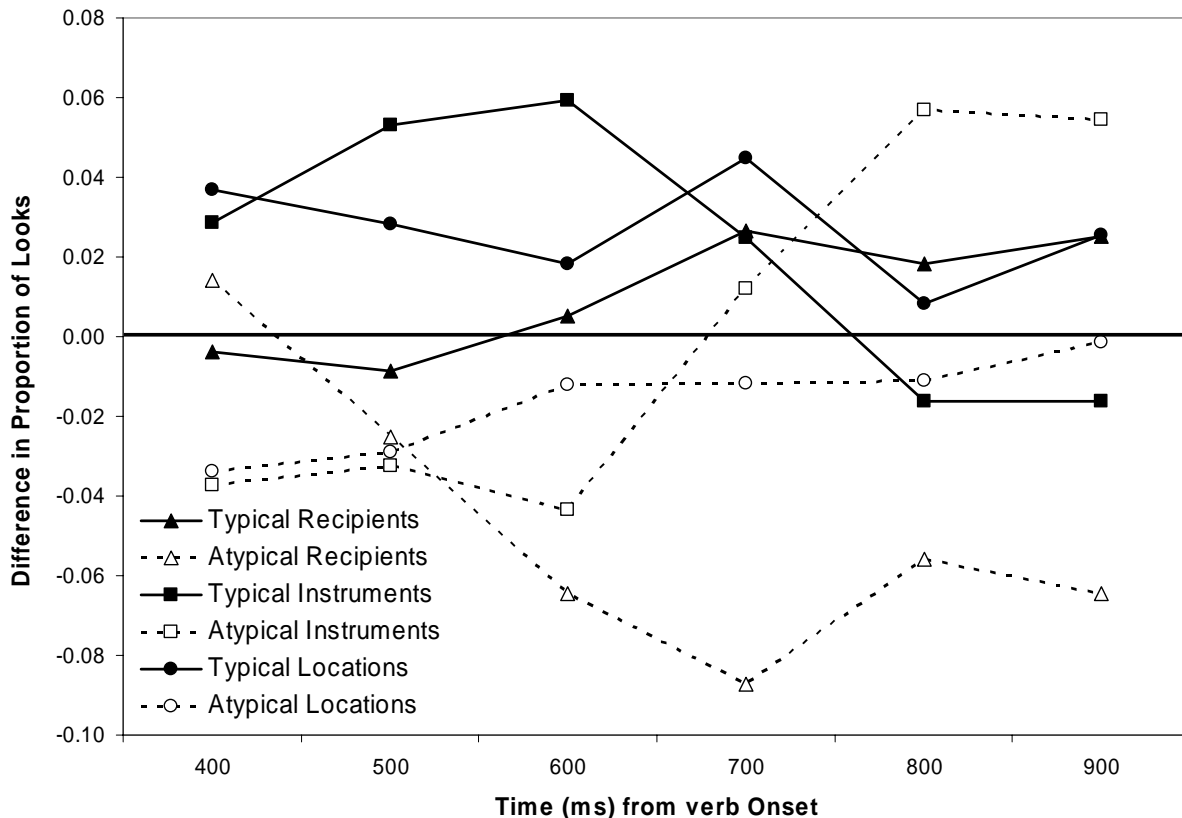
The conditions leading to a typicality effect for recipients also led to earlier and larger typicality effects for adjuncts (locations and instruments). That is, from 400 to 700 ms after verb onset, listeners had more of a tendency to look at a typical instrument or location if an atypical one was also pictured. Correspondingly, they were less likely to look at an atypical instrument or location if a more typical one was depicted.

To better understand how the results of Experiments 1 and 2 differ, the participant and item means from each experiment were combined for a meta-analysis with *experiment* as a between-participants variable. In order to make the data sets from the two experiments comparable, several choices were made. First, the sentence type variable was eliminated in Experiment 2. That is, by averaging over both typical and atypical sentences in Experiment 2, a single probability was obtained for each typical and atypical target condition. Second, because the critical temporal interval differed in the two experiments, the data sets were compared in two ways. The first analysis considered the probability of at least one look during the critical

intervals chosen for each experiment: 500 -1000 ms after verb onset in Experiment 1 and 400 –

733 ms for Experiment 2. This analysis has the advantage of using the data from the temporal

intervals that were deemed most appropriate for each experiment. However, it has the obvious

disadvantage of using intervals that differ in both latency and duration across the two

experiments. The second analysis addressed this concern by comparing the data in the bins 500 –

800 ms after verb onset for both experiments. These bins were selected because this temporal

region exhibited the primary results in both experiments.

The probability of at least one new look to a target picture, during the critical intervals

defined for each experiment, was analyzed in 2(experiment) by 2(list) by 3(verb type) by 2(target

typicality) ANOVAs by participants and by items. These data were summarized in Tables 5 and

7 above. A main effect of target typicality was found [$F1(1,30) = 6.60$, $F2(1, 42) = 7.98$, $p < .05$]

that did not interact with experiment [$Fs < 1.5$]. The main effect of verb type was marginal

[$F1(2,60) = 3.03$, $F2(2, 42) = 2.49$, $p < .10$] and did not interact reliably with experiment [$F1(2,

60) = 1.62$, $p > .10$; $F2(2, 42) = 2.69$, $p < .10$].

For the second analysis, participant and item means for each verb type by target typicality

condition in bins 500, 600, and 700 of each experiment were submitted to 2(experiment) by

2(list) by 3(verb type) by 2(target picture) by 3(bin) ANOVAs. There was a main effect of target

picture typicality [$F1(1, 30) = 5.68$, $F2(1, 42) = 5.64$, $p < .05$] that interacted with experiment in

the items analysis and reached a marginal level of significance in the analysis by participants

[$F1(1, 30) = 3.96$, $p = .06$; $F2(1, 42) = 4.22$, $p < .05$]. Thus, there were more looks to typical

targets overall, but the effect was enhanced in Experiment 2, as shown in Figure 3. Picture

typicality did not interact with verb type, either alone [$F1(2,60) = 1.39$, $F2(1, 42) = 1.02$, $p > .10$]

or in combination with experiment [$Fs < 1.0$]. There was, however, an interaction between verb

type and bin [$F1(4, 120] = 3.70$, $F2(4, 84) = 3.83$, $p < .05$] that did not interact with experiment

[$Fs < 1.0$]. An examination of the means in the verb by bin cells indicates that the probability of

looking at locations did not increase with bin onset time, whereas the probability of looking at

recipients increased with each successive bin, and looks to instruments increased only in the final

bin. Thus, while the main effect of verb type was not significant [$F1(2, 60) = 2.06$, $F2(2, 42) = 1.23$, $p > .10$], an effect of argument status was observed in the verb by bin interaction.

Discussion

The primary finding in Experiment 2 was a main effect of target picture typicality. Although the typicality effect was not reliable across the recipient arguments alone, they exhibited the same pattern as the adjuncts in this experiment. These results replicate the typicality effects found in previous research using both passive listening (e.g., Kamide et al., 2003) and directed action tasks (e.g., Chambers et al., 2002; Sedivy et al., 1999).

The argument structure hypothesis predicted that verb argument structure would be used to identify potential recipient arguments, and if more than one potential recipient was pictured, real world knowledge would be used to select the most likely candidate. Figure 3 supports this prediction by revealing a substantial drop in anticipatory looks to the atypical recipient in Experiment 2. The argument structure hypothesis made no predictions about an adjunct typicality effect. Rather, the stronger adjunct typicality effects in Experiment 2 are consistent with an increased focus on instruments/locations, brought about by representing potential instruments/locations in two of the four pictures.

One consequence of the increased probability of looking at typical adjuncts, coupled with a decrease in looks to atypical recipients, was the disappearance of the verb type effect observed in Experiment 1. To be clear, in **both** Experiments 1 and 2, the potential argument status of <u>typical</u> targets provided no additional benefit above and beyond the typicality of that target for the described events. The finding that listeners are just as likely to make anticipatory looks to typical adjuncts as to typical arguments is consistent with several alternatives to the argument structure hypothesis. One such alternative is that there is no distinction between arguments and adjuncts, and detailed world knowledge or co-occurrence frequency is inextricably intertwined with linguistic knowledge. A second possible alternative is that there is no direct relationship between listeners' initial linguistic representations and their looking patterns. In other words, a

propensity to look at a bed upon hearing the verb *slept* does not entail that the listener has constructed a linguistic representation of a verb phrase with the word *bed* in it.

Importantly, neither of these alternative accounts can explain the effect of argument status on atypical targets in Experiment 1: Participants' attention was drawn to an atypical potential recipient if no typical recipient was pictured, while no comparable pattern was observed for atypical adjuncts. This contrast between arguments and adjuncts can only be explained by an account in which argument structure knowledge guides visual attention within this paradigm, and in which formal linguistic constraints are distinct from co-occurrence patterns. Given the potential significance of the argument status effect for distinguishing among theoretical alternatives, it was replicated in Experiment 3, using a different set of stimuli.

Experiment 3

In this experiment, verb type/argument status was manipulated <u>within</u> sentence and picture set. In order to create maximally similar sentences and identical picture sets, only two levels of verb type were used: dative and action verbs. Example sentences are given in (5). The same picture (corresponding to *owners* in (5)) was used to depict the recipient argument in dative condition and the benefactive adjunct in action-verb condition. An instrument was never mentioned in the critical trials, though a typical instrument (with respect to the action verb) was always pictured.

> (5) *One window was broken, so the handyman…*
>
> a. …***mentioned*** *it right away to the <u>owners</u>.* (recipient-Argument) b.
>
> …***fixed*** *it hurridly for the <u>owners</u>.* (benefactor-Adjunct)
>
> (pictures: broken window, handyman, owners, tools)

The argument structure hypothesis predicts that verb type should influence anticipatory looking rates to the recipients/benefactors, but not to instruments. Between verb onset and the onset of *owners* there should be (i) more looks to recipients in the dative condition than to benefactors in action condition, and (ii) more looks to recipients in dative condition than looks to

instruments in the action condition. Alternatively, suppose that arguments and common adjuncts such as benefactors and instruments are represented in the same way in lexical structure, with both types of constituents implied by verb. In that case, looks to recipients in the dative sentences should be balanced by looks to instruments and benefactors in the action verb sentences.

The argument structure hypothesis does not make any specific predictions about looks to the subject and direct object pictures, except that looks to these pictures may be somewhat reduced after the verb in dative sentences compared to the action verb sentences, due to the increased tendency to look at the recipient picture.

Methods.

Participants. Fourteen students from the University of Michigan completed the experiment. Twenty additional participants were recruited, but were not included in the study for reasons outlined below. All participants were from introductory psychology courses and received partial course credit in exchange for their participation. All participants were native English speakers with normal or corrected-to-normal vision.

Sentence Stimuli. There were 16 pairs of critical sentences similar to (5). The full set is in Appendix B. The sentences in each pair were designed to be maximally similar, except that one contained a dative verb and a subcategorized PP recipient, while the other contained an action verb and a non-subcategorized PP benefactor. All the recipient NPs were in PPs beginning with *to*, as in Experiment 1. The benefactor NP was always in a PP beginning with *for*. The action verbs could all be used with either a benefactor or an instrument or both, but for which it was unlikely (in the author's judgment) that either a benefactor or an instrument was subcategorized. An instrument was not explicitly mentioned in any of the critical sentences. There were, however, eight filler sentences that all contained *with*-PPs that served as instruments. In addition, there were eight filler trials that did not mention an instrument, benefactor, or recipient. When including filler trials in the count, participants heard eight sentences with recipients, eight with benefactors, and eight with instruments.

The verbs were matched for frequency; the mean frequency of the action verbs was 38.3 and the mean frequency for dative verbs was 38.6 (Francis & Kucera, 1982). All critical sentences had a pronominal direct object and an adverbial phrase between the direct object and the recipient/benefactor. For the sake of naturalness, the adverbial phrase differed in the dative and action verb versions, but was matched for syllable length. In all cases, the adverbial phrase was between three and five syllables. Mean syllable length was 4.25 for the dative condition and 4.19 for the action verb condition. Sentences were spoken by a woman with a Midwestern accent and digitally recorded for presentation. The mean temporal parameters for each condition are provided in Table 8.

Table 8. ISI (ms) between onset of the verb, onset of the PP, & target noun.

	Action verb	Dative Verb
PP onset	1243	1307
NP onset	1567	1594

As in Experiment 1, co-occurrence information was collected from the WWW. These data are summarized in Table 9. There was no difference in co-occurrence frequency among the recipient, benefactor, and instrument in either search. Three two-tailed t-tests for groups with unequal variance were conducted on the item proportions from each search.

Table 9. WWW mean co-occurrence proportions (standard deviations)

	Google	Lycos Exact Phrase
Dative Recipient	.14 (.15)	.000047 (.000173)
Action Benefactor	.14 (.11)	.007293 (.029170)
Action Instrument	.09 (.10)	.000300 (.000887)

Picture Stimuli. Each critical or filler sentence was accompanied by a set of four photographs, as in Experiments 1 and 2. The photographs corresponded to the subject/agent, the direct object/theme, a potential recipient or benefactor, and a potential instrument. The photograph in the upper left quadrant was always the referent for the first NP in the sentence. The other three photographs were rotated across the other three quadrants, so the instruments and recipients/benefactors were equally likely to appear in each of the three positions. The same visual display was used for the dative and action verb versions of each item. The visual displays in the filler trials were similar arrays of four pictures.

Procedure & Equipment. The auditory sentences and their corresponding pictures were presented in a fixed random order. There were two versions of the stimulus list, each having one version of each item and equal numbers of typical and atypical sentences for each verb type. Half of the participants were given each list. There were four practice sentences at the beginning of the stimulus list to acquaint the participant with the flow of events.

Both eye position and head position were tracked, allowing for automated data analysis. The data were collected with an ISCAN ETL-500 head-mounted camera, running at 120 hz, with the Polhemus magnetic head-tracking system. The position of the stimulus screen, relative to a magnetic transmitter was defined prior to the collection of any data, and was fixed for the entire set of experimental participants. When each participant arrived at the laboratory, the experimental procedure was explained and they were seated in a comfortable chair located between the stimulus screen and the magnetic transmitter. The eye tracking headgear was placed securely on the participants' head and the eye camera and illumination level were adjusted for optimal tracking. Head position was tracked via a magnetic receiver on the headgear. A five-point calibration routine was used to calibrate head and eye position with fixation targets on the stimulus screen. Before the experiment, there was an initial calibration check and additional checks occurred throughout the experiment, using the procedures outlined in Experiment 2. Ten potential participants were dismissed because the initial calibration failed. The data from ten other potential participants were discarded due to a noticeable shift in the headgear.

There were three analog records of calibration accuracy. For each fixation target, accuracy was considered to be adequate if some portion of the fixation cross-hair was inside the circle or touching the circle. It was rare for this criterion to be met for all six points in all three calibration checks (an accuracy score of 18), but it had to be met for 12 out of 18 fixation targets, and at least three out of six on each calibration check. The average calibration score was 14.4.

During the experiment, each trial was initiated by a mouse press. This simultaneously started data collection from the eye tracking system and incremented the Power Point presentation to the next slide. The auditory sentence began playing approximately 340 ms after

the mouse press. There were no practice sentences before beginning the experiment, but the first item was not a critical trial. Thirteen of the sentences were followed by a yes/no comprehension question.

Fixations were automatically categorized in terms of four spatial regions of interest: the sentential subject, the direct object, the recipient/benefactor, and the instrument. On average, the direct object and the recipient/benefactor were the largest figures. They did not differ in average size. The figures of the subjects and the instruments averaged 78% and 56% of the direct object/recipient figures, respectively. Those fixations that fell within a critical spatial region and occurred during the temporal window of interest underwent further analysis.

The one temporal variable that was not controlled in this experiment was the time to load the sound file and initiate playback within PowerPoint. Oddly, the delay was not positively correlated with the size of the sound file (R = -.19), and for a given item, the delay varied from one run to another. In three test runs of the experiment, the average delay was 338 ms, with a standard deviation of 235. Therefore, the start of the auditory file was always assumed to be 340 ms after the onset of data collection. Obviously, this adds some undesirable noise to the data, flattening potential effects over a longer time interval.

Results.

None of the participants missed more than one comprehension question; mean accuracy was 97%. One item (#11 in Appendix C) was excluded from the analysis because the sentence did not distinguish the sentential subject picture and the benefactor picture in the action verb condition.[5] New looks to each picture were coded automatically, beginning at verb onset. The time of verb-onset was determined separately for each item/condition.

---

[5] The action verb condition of this item was the only case in which there were no looks to the subject during the critical region by any participant. Listening to the sentence, it became clear why no one looked at the subject picture: "The bacteria were complex, and the technician studied them painstakingly for the professors." Because professors and technicians look similar, the singular/plural distinction was intended to distinguish the subject from the recipient. Unfortunately, coarticulation between the end of "technician" and the beginning of "studied" resulted in the impression that "technicians" had been uttered. A careful review of the stimuli verified that the problem was isolated to this item.

The probability of at least one look to the recipient/benefactor and instrument pictures, 500 to 1000 ms after verb onset, was computed by participant and by item. These data are summarized in Table 10. The time window was selected for three reasons. First, it is the same time window used in Experiment 1 and the goal of the current experiment is to replicate the argument status effect observed there. Second, Figure 4 below suggests that effects of verb type began around 500 ms after verb onset. Finally, this interval precedes the onset of the critical NP in every case (the earliest NP onset was 1248 ms after verb onset) and any influence of the preposition should be very minimal. The earliest PP onset was 960 ms after verb onset.

Table 10. The percentage of trials with at least one look to the specified picture, 500 to 1000 ms after verb onset.

Verb	Recip/Benef	Instrument
Dative	.49	.16
Action	.34	.15

Table 10 reveals a high probability of fixating dative recipients, as expected. In contrast, the instrument pictures were fixated infrequently. In fact, there appear to have been substantially fewer looks to the action verb instruments compared to Experiments 1 and 2 (see Tables 5 and 7). The instruments in the current experiment were selected as highly typical for the action described by the action verb. Therefore, the action-instrument cell of Table 10 should be comparable to the typical instrument cell of Table 5. Surprisingly, in Experiment 3 there were no more looks to the instruments in the action verb condition than there were in the dative verb condition, in which the described event was very unlikely to involve such an instrument.

The decrease in looks to typical instruments may be explained by an animacy bias that is strongly apparent in Table 10. Compare the cells for action verb instruments and action verb

benefactors. Both of these adjuncts were plausible, but the potential benefactors received double the number of looks during the critical interval. Such an animacy bias was not apparent in Experiment 2, where typical instruments received as many looks as typical recipients (see Table 7). The animacy bias may be present in the current experiment because every critical sentence (half of all trials) used a PP containing an animate NP--either a recipient or benefactor—within the verb phrase. In contrast, the probability of an animate NP in that part of the sentence was .33 in Experiments 1 and 2, and half of those animate NPs were atypical for the described event. In other words, listeners may have learned to look at animate NPs over the course of Experiment 3, even when the verb's argument structure did not specify an animate argument. Some support for this possibility is found in the pattern of looks to benefactors and recipients across the sequence of trials. Anticipatory looks to benefactors increased over the course of the experiment. There were none on the first critical trial. Averaging over the first eight critical trials, listeners looked at benefactors 29% of the time, and the percentage rose to 41% over the last seven trials. In contrast, the likelihood of an anticipatory look to a recipient was relatively stable: 43% on the first trial; 48% over the first eight trials, and 49% over the last eight trials.

Because of the animacy bias observed in Table 10, the critical comparisons were made across verb type, while holding picture constant. Planned comparisons revealed that the recipient/benefactor picture received more anticipatory looks during the critical region following a dative verb than following an action verb [t1(13) = 2.22, t2(14) = 2.35, p < .05]. In contrast, there was no effect of verb type on looks to the instrument picture. This pattern is predicted by the argument structure hypothesis, because only the dative recipient is a subcategorized argument.

Looks to each picture were next analyzed across time, as a function of verb type. The eye movement record was divided into 100 ms bins, beginning 200 ms after verb onset and ending 1600 ms after verb onset, around the time of the target noun onset. Figures 4 and 5 each illustrate looks to two of the four pictures during each bin, for each verb type.

Figure 4. The proportion of new looks to the recipient/benefactor and direct object pictures across time, 200 – 1600 ms after verb onset.
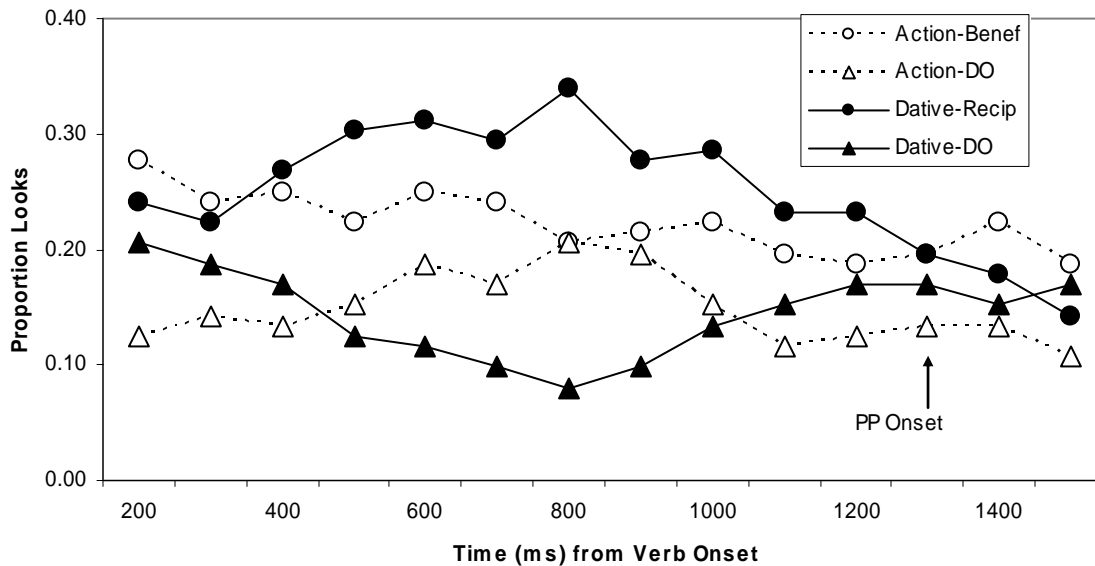

Figure 4 illustrates the proportion of looks to the recipient/benefactor and direct object pictures. There appear to be more looks to the recipient/benefactor picture for the dative condition, relative to the action verb condition, beginning about 500 ms after verb onset and lasting until the PP onset. During this period, looks to the direct object fall dramatically in the dative condition. Recall that in both verb conditions, the direct object was focused at the beginning of the sentence and referred to with a pronoun immediately following the verb. The proportions for participants and for items were submitted to 2(list) by 2(picture) by 2(verb type) by 5 (bin) ANOVAs, encompassing the interval 500 - 1000 ms after the onset of the verb. There were more looks to the recipient/benefactor than the direct object during this period [$F1(1,12)$ =12.69, $F2(1,13) = 13.02$, $p < .01$], consistent with the animacy bias observed in Table 10. More importantly, this effect interacted with verb type [$F1(1,12) = 12.26$, $F2(1,13) = 14.68$, $p < .01$]. A second ANOVA was performed on just the recipient/benefactor picture and confirmed the prediction that there should be more looks to the recipient picture, following a dative verb as opposed to an action verb [$F1(1,12) = 11.13$, $F2(1,13) = 7.34$, $p < .05$]. This finding replicates

the argument status effect from Experiment 1 during the same time interval.

Figure 5. The proportion of new looks to the subject and instrument pictures across time, 200 – 1600 ms after verb onset.
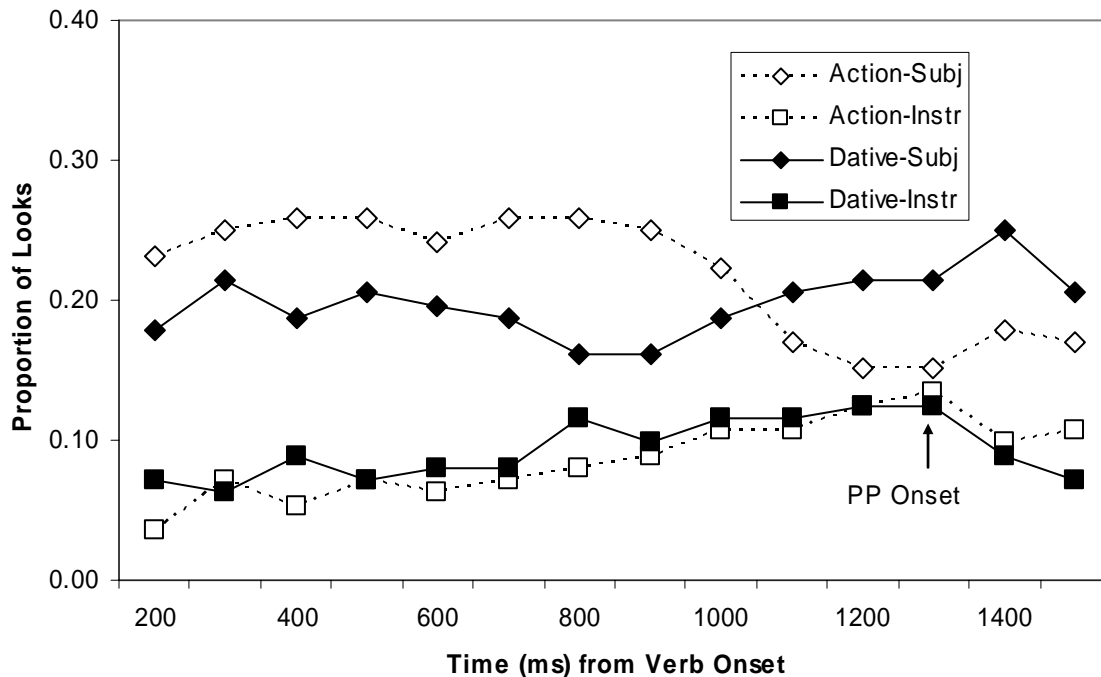

Figure 5 illustrates looks to the subject and instrument pictures across time. There was no time period during which there were more looks to the instrument in the action verb condition compared to the dative condition. A 2(list) by 2(picture) by 2(verb type) by 5 (bin) ANOVA, using the same bins as above, revealed only an effect of picture, with more looks to the subject picture than the instrument picture [$F1(1,12) = 11.59$, $F2(1,13) = 9.65$, $p < .01$], consistent with the animacy bias described above. The interaction of picture and verb type was nearly significant [$F1(1,12) = 5.73$, $p < .05$; $F2(1,13) = 4.67$, $p = .05$], with a tendency for more looks to the subject picture after action verbs than after dative verbs.

Discussion

Experiment 3 replicated the crucial argument status effect that was observed in Experiment 1. The effect occurred during the same time interval, despite a faster speech rate. The

argument status effect in the current experiment is particularly compelling because the same picture received more looks when the linguistic context set it up to be a potential argument than when the linguistic context set it up to be a potential adjunct. In contrast to the argument status effect for the recipient/benefactor, there were no more looks to the instrument picture following the action verb than there were following the dative verb, suggesting that instruments are not arguments for the set of action verbs used here.

In Experiment 3, the argument status effects are overlaid on top of a general preference to look at animate pictures: both the recipient/benefactor picture and the subject picture attracted a large proportion of the looks compared to the direct object and instrument pictures. This may explain the otherwise surprising drop in the probability of looks to typical instruments, compared to the previous experiments. As noted above, the drop in looks to instrument adjuncts was balanced by the relatively high probability of looking at a potential benefactive adjunct, especially in the second half of the experiment. Participants in the current experiment found pictures of animate entities highly salient during the temporal interval of interest, possibly because an animate NP (either benefactor or recipient) was mentioned in a post-verb PP 50% of the time.

The argument structure hypothesis did not make predictions about looks to the subject and direct object pictures, except that looks to these pictures might be reduced after the verb in dative sentences compared to the action verb sentences due to the increased tendency to look at the recipient picture. They were analyzed in the current experiment primarily to explore the animacy bias observed in Table 10.

General Discussion

The argument status effects in Experiments 1 and 3 support the view that linguistic arguments, such as recipient indirect objects, have a special status in language comprehension. As we are listening to a sentence, new entities are implicitly introduced into the discourse via the argument structures of verbs and other phrasal heads. In the studies presented here, recipient

arguments were contrasted with locative adjuncts, benefactive adjuncts, and instruments--which have long been considered a borderline case in the argument/adjunct distinction. In addition, the typicality effects from Experiment 2 demonstrate that typical event participants will draw visual attention, regardless of their linguistic status as arguments or adjuncts, under some conditions. Thus, both argument structure and real world knowledge can and do influence the likelihood of anticipatory looks to pictures that are relevant to a spoken sentence.

The following discussion is divided into three parts. The first section evaluates the claim that argument structure knowledge can be distinguished from general world knowledge, and that linguistic arguments draw attention during the passive listening paradigm because verbs introduce their arguments into the discourse. The second section discusses the evidence suggesting that instruments are, in fact, adjuncts. The final section evaluates the utility of the experimental paradigm for investigating the cognitive components of sentence comprehension and summarizes the current findings.

Argument status vs. Real World Knowledge. The argument status effects observed here are consistent with prior findings from the reading literature suggesting that argument structure knowledge guides sentence comprehension by implicitly introducing new entities into the discourse. For example, Carlson and Tanenhaus (1988) used a whole sentence reading task and found that participants read sentences like (7) faster following (6a) than after (6b), despite the fact that a suitcase is plausible in both scenarios. The explanation was that the argument structure for *unloaded* includes an agent (John), a theme (unspecified) and a source (trunk). Even though the theme was not explicitly mentioned, it was implicitly introduced by the verb. Because the theme was already in the discourse model, it was easy to refer back to it with a definite NP anaphor, *the suitcase*. In contrast, a bridging inference is required to interpret *the suitcase* after (6b) (see Haviland & Clark, 1974).

      (6)   a. *John had difficulty unloading his car.*

             b. *John had difficulty running fast to catch his plane.*

      (7)   *The suitcase was heavy.*

As discussed in the introduction, implicit arguments have also been investigated by Mauner et al. (1995; Mauner & Koenig, 1999, 2000), using short passives. Crucially, Mauner and Koenig found immediate anomaly effects even using passive/intransitive contrasts like *was sold/had sold*, where both events logically require an agent. This supports their contention that implicit agents are derived from lexical rather than conceptual sources.

The current study builds upon the earlier reading research, as well as prior listening paradigms that investigated anticipatory looks to arguments (e.g., Altmann & Kamide, 1999). The current study investigated a different class of verb-argument relations and is unique in its explicit contrast between arguments and adjuncts. The results presented here add further support to previous claims that verbs implicitly introduce their arguments as the verb is recognized (e.g., Altmann & Kamide; Mauner & Koenig, 2000). And as maintained by Mauner and Koenig, the current results suggest an early and automatic mechanism, driven by lexical access, rather than a high-level conceptual process.

Before concluding that the argument status effects reported here are due to the use of argument structure knowledge, consider the possibility that the argument status effects in Experiments 1 and 3 were caused by co-occurrence frequency. It is likely that, within a large corpus of text, the dative verbs co-occur with a recipient more frequently than the transitive verbs co-occur with an instrument or the intransitive verbs co-occur with a location. At this level, co-occurrence makes the same predictions as the argument structure hypothesis, because a particular class of argument generally has a higher text co-occurrence frequency with a given verb than a particular class of adjunct. Thus, argument status could be viewed as an emergent property of a frequency-driven system that makes generalizations that correspond to distinctions in linguistic theory. The current data do not distinguish between the argument structure hypothesis and the co-occurrence hypothesis if co-occurrence is computed over phrases of a particular linguistic class (e.g. the co-occurrence of *suggest* and any recipient, as opposed to the co-occurrence of *suggest* and *toddler*).

Under a more fine-grained frequency account, co-occurrence patterns among specific words or entities should matter. With this in mind, sentence-level acceptability (as opposed to verb-level acceptability) was manipulated via the typicality contrast of Experiments 1 and 2. The typical targets in the first two experiments were intended to be a prototypical recipient, instrument, or location for a given subject+verb(+object) context, whereas the atypical targets satisfied only the verb-specific constraints. For example, when killing a bug, a rock would be a better instrument than a spear, but when killing a tiger, a spear would be more suitable. In fact, the typical targets attracted more looks than the atypical targets in Experiment 2, just as a fine-grained frequency hypothesis would predict. However, a fine-grained frequency account cannot explain the argument status effects found in Experiments 1 and 3.

An explanation based upon the co-occurrence of a particular verb and its complements would also have trouble accounting for the results from Kamide et al.'s (2003) third experiment. An English translation of an example sentence pair is provided in (8). Using Japanese, a verb-final language, Kamide et al. found that a dative particle on the second noun (in 8a) prompted anticipatory looks to a potential direct object (a hamburger). Clearly, these anticipatory looks were not driven by the co-occurrence of "bring" and "hamburger". Kamide et al. suggest that the case-marking particles facilitate thematic role assignment, which in turn constrains the class of verbs that can occur. In the case of (8a), a verb of transference can be inferred from the dative particle. A verb of transference requires that something be transferred, and real world knowledge about waitresses and customers may lead to the expectation that some food will be the transferred substance as opposed to a trash can, the other pictured item. Crucially, real world expectations about waitresses and customers alone did not prompt looks to the hamburger, demonstrated by the lack of looks to the hamburger in (8b). Thus, argument structure constraints introduced a new entity into the discourse (the transferred substance) and real world knowledge prompted looks to the most plausible argument.

(8)   a. Waitress-nominative customer-dative merrily hamburger-accusative bring.

       b. Waitress-nominative customer-accusative merrily tease.

Pictures: waitress, customer, hamburger, trash can

In short, while some of the reported findings clearly reveal the influence of real world knowledge on anticipatory looks to potential arguments and adjuncts, the results of the current study (as well as some previous findings) support a distinction between real world knowledge and formal linguistic constraints. The argument status effects reported here can only be explained by an account in which argument structure knowledge guides sentence understanding.

Are instruments adjuncts? One motivation for investigating instruments in the current study was their ambiguous linguistic status with respect to the argument/adjunct distinction. On balance, these data suggest that instruments are not lexically specified arguments for action verbs like *attack, eat*, and *clean*. While there was some evidence from Experiment 1 that instruments pattern in between definitive arguments (such as recipients) and definitive adjuncts (such as locations), there was no evidence that action verbs implicitly introduced an instrument in Experiment 3. (Because there were no argument status effects in Experiment 2, that experiment provided no evidence on this point.) The pattern of results for instruments across Experiments 1 and 3 could be easily explained if a subset of the action verbs used in Experiment 1 took instrument arguments, while few or none of the action verbs used in Experiment 3 took an instrument argument. However this possibility seems unlikely. The action verbs that were used in Experiment 1 but not Experiment 3 were *bump, beat, attack, eat, assault, damage, smash*, and *clean*. None of these verbs meet the semantic obligatoriness criterion used by Koenig et al. (2002) to determine whether a given verb specifies an instrument argument.

In apparent contrast with the current results, Ferretti et al. (2001) found that action-instrument pairs patterned differently from intransitive-location pairs: *Served* primed *platter*, but *danced* did not prime *ballroom*. Ferretti et al. concluded that instruments are arguments, while locations are not. It is likely that a small set of action verbs do specify an instrument argument, but semantic priming in a lexical decision task is a strange type of evidence for argument status: The fact that *dog* primes *cat* cannot make *cat* an argument of *dog* even if some other related word does not produce comparable amounts of priming. There is undoubtedly a great deal of

conceptual knowledge about individual words that falls outside the realm of argument structure, and some aspects of this knowledge will generate stronger priming effects than others, depending upon the structure of semantic memory. The point is that our knowledge about action verbs and the events described by them does include information about the instruments that are normally used, regardless of whether the verb subcategorizes for an instrument as an argument.

Summary. The eye-tracking paradigm used here appears to be well suited to the investigation of how lexically specified entities are introduced into the discourse by the verbs that specify them. Listeners naturally looked at entities that were introduced by the verb's argument structure as soon as they recognized the verb. Importantly, argument structure knowledge was not the only factor guiding anticipatory looks in this paradigm; typical adjuncts were also often fixated prior to explicit mention. However, the probability of anticipatory looks to typical adjuncts rose and fell, depending upon experimental conditions. For example, typical instruments were often fixated in Experiment 2, in which the experimental stimuli made instruments salient. In contrast, typical instruments were rarely fixated in Experiment 3, where the experimental sentences highlighted animate entities. The evidence that the experimental stimuli made some pictures more salient than others demonstrates that it can be dangerous to compare looks to different pictures, as in Experiments 1 and 2. In fact, this practice has been quite common in directed action versions of the paradigm (e.g., Allopenna et al., 1998; Chambers et al., 2002; Dahan et al., 2001).

In sum, these results support an expectation-driven account of sentence comprehension in which arguments have a privileged status. Verbs can implicitly introduce new entities into the discourse model via their argument structure and the associated thematic roles. Despite ongoing debates about the relationship between thematic role knowledge and world knowledge, the current results suggest that the thematic roles associated with subcategorized arguments are distinct from general world knowledge.

Author Note

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory & Language*, **38**, 419-439.

Altmann, G. T. M. (1999). Thematic role assignment in context. *Journal of Memory & Language*, **41**, 124-145.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, **73**, 247-264.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S. & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition,* **76**, 13-26.

Boland, J.E. (1997a). The relationship between syntactic and semantic processes in sentence comprehension. *Language and Cognitive Processes*, **12**, 423-484.

Boland, J. E. (1997b). Resolving syntactic category ambiguities in discourse context. *Journal of Memory and Language,* **36**, 588-615.

Boland, J.E., Tanenhaus, M.K., Garnsey, S.M., and Carlson, G. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*, **34**, 774-806.

Boland, J.E., Tanenhaus, M.K., and Garnsey, S.M. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language*, **29**, 413-432.

Carlson, G. & Tanenhaus, M. (1988). Thematic roles and language comprehension. In W. Wilkins (Ed.) *Thematic relations*. New York: Academic Press.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory & Language*, **47**, 30-49.

Cooper, R. M. (1974). The control of eye fixations by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, **6**, 84-107.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology,* **42**, 317-367.

Eberhard, K., Spivey-Knowlton, M., Sedivy, J. & Tanenhaus, M. (1995). Eye movements as a window into real-time spoken language processing in natural contexts. *Journal of Psycholinguistic Research*, **24**, 409-436.

Ferreira, F. & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, **25**, 348-368.

Ferreira, F. & McClure, K. K. (1997). Parsing of garden-path sentences with reciprocal verbs. *Language and Cognitive Processes*, **12**, 273-306.

Ferretti, T. R., McRae, K., & Hatherall, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, **44**, 516-547.

Francis, W. N. & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin Co.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, **37**, 58-93.

Haviland, S. E. & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, **13**, 512-521.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Jackendoff, R. (1987). The status of thematic relations in linguistic theory. *Linguistic Inquiry*, **18**, 369-412.

Kamide, Y., Altmann, G. T. M. & Haywood, S. (2003). Prediction and thematic information in incremental sentence processing: Evidence from anticipatory eye-movements. *Journal of Memory and Language*, **49**, 133-156.

Koenig, J-P, Mauner, G, & Bienvenue, B. (2002). Class specificity and the lexical encoding of participant information. *Brain & Language*, **81**, 224-235.

Konieczny, L., Hemforth, B., Scheepers, C. & Strube, G. (1997). The role of lexical heads in parsing: Evidence from German. *Language and Cognitive Processes*, **12**, 307-348.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation.* Chicago: University of Chicago Press.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, **101**, 676-703.

Marantz, A. P. (1984). *On the nature of grammatical relations.* Cambridge, MA: MIT Press.

Mauner, G. & Koenig, J-P. (1999). Lexical encoding of event participant information. *Brain & Language*, **68**, 178-184.

Mauner, G. & Koenig, J-P. (2000). Linguistic vs. conceptual sources of implicit agents in sentence comprehension. *Journal of Memory and Language*, **43**, 110-134.

Mauner, G., Tanenhaus, M. K., & Carlson, G. N. (1995). Implicit arguments in sentence processing. *Journal of Memory and Language*, **34**, 357-382.

McElree, B. & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 134-157.

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, **12,** 137-176.

Novick, J. & Trueswell, J. C. (2001). *Lexical priming of verb argument structure during auditory language comprehension.* The 42nd Meeting of the Psychonomic Society, Orlando.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, **71**, 109-147.

Spivey-Knowlton, M. & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, **55**, 227-267.

Stowe, L., Tanenhaus, M. & Carlson, G. (1991). Filling gaps on-line: Use of lexical and semantic information in sentence processing. *Language and Speech*, **34**, 319-340.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M. & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, **268**, 1632-1634.

Trueswell, J., Tanenhaus, M. & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, **33**, 285-318.

Trueswell, J., Tanenhaus, M. & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 528-553.

Van Valin, R. & Lapolla, R. (1997). *Syntax: Form, meaning, and function*. Cambridge: Cambridge University Press.

APPENDIX A: STIMULI FOR EXPERIMENTS 1 AND 2

The typical and atypical target items are separated by a slash, with the typical target listed first. Filler pictures for Experiment 1 are listed within parentheses after each sentence. The superscripted numbers indicate the quadrant of the picture on the PowerPoint slide (1= upper left, 2 = upper right, 3 = lower left, 4 = lower right); the location specified for the target applies to both the typical and atypical conditions in Experiment 1. In Experiment 2, both typical and atypical targets were presented, replacing the one filler picture for Instrument and Dative items, and replacing one of the second of the two filler pictures for Intransitive items.

Instrument Items

1. A pedestrian^1 was crossing the parking lot when a truck^2 bumped her absentmindedly with its tailgate^3/antennae on Friday. (church ^{4})

2. One student^1 was very naughty in class, and the teacher^4 hit him angrily with a yardstick^2/dictionary during recess. (basketball court^3)

3. The donkey^1 would not move, so the frustrated farmer^3 beat it vigorously with a stick^4/hat every day. (grass^2)

4. A guard^1 was suspicious of a convict^2 who was planning to escape, so the convict attacked him viciously with a knife^3/pillow earlier today. (truck^4)

5. One window^1 was especially dirty, so the janitor^4 wiped it carefully with a cloth^2/jacket this afternoon. (medical staff^3)

6. A broken^1 pipe was flooding the laundry room, so the plumber^3 cut it precisely with a saw^4/an axe Monday evening. (laundry room^2)

7. The child^1 reached for a steak^2 and he ate it anxiously with some silverware^3/tools right away. (woman^4)

8. The table^1 was broken, but the handyman^4 fixed it very quickly with several nails^2/some glue this morning. (woman^3)

9. A cop^1 was harassing a teenager^3, so the teen assaulted him violently with a rock^4/pencil Sunday evening. (street corner^2)

10. The walls^1 in the stairway were old and crumbly, and a young man^2 damaged them carelessly with his pocketknife ^{3}/backpack last night. (stairway^4)

11. The window^1 to the office was closed so the man^4 smashed it in frustration with a chair^2/book last night. (office^3)

12. A tiger^1 was bothering the village, so tribesmen^3 killed it skillfully with a spear^4/rock last week. (alligator^2)

13. The corridor^1 was dirty again, so the housekeeper^2 cleaned it right away with a broom^3/a sock this afternoon. (medical staff^4)

14. The sock^1 had developed a hole, so the girl^4 repaired it once again with a needle^2/stapler last night. (boot^3)

15. The toy soldier[1]was a favorite toy, but the boy[3]broke it mischievously with his boot[4]/a bottle this morning. (woman[2])

16. The door[1]was looking shabby, so the woman[2]painted it diligently with a roller[3]/sponge this weekend. (truck[4])

Intransitive Items

17. The girl[1] slept throughout the night on the bed[2]/bus after the party. (toy car[3], pillow[4])

18. The toddler[1] sneezed very loudly into the tissue[4]/soup at lunchtime. (teacher & class[3], toy car[2])

19. The grandfather[1] coughed very softly into his hand[3]/wallet during church. (church[2], toddler[4])

20. The dancer[1] stumbled awkwardly on the floor[2]/table at the party. (candle[4], couple[3])

21. The magician[1] waited patiently near the stage[4]/sidewalk before the show. (audience[3], doll[2])

22. The actress[1] sighed rather rudely at one of her fans[3]/cars last week. (pencil[2], bus[4])

23. The cancer patient[1] died without any pain at the hospital[2]/library last weekend. (microscope[4], priest[3])

24. The American Indian[1] dozed quietly near the fire[4]/airplane at midnight. (sky[3], alligator[2])

25. The baby[1] burped a little on his bib[3]/spoon during dinner. (sponge[2], backpack[4])

26. The puppy[1] yelped excitedly at the mailman[2]/couch at Bobby's house. (house[4], bottle[3])

27. The fireworks[1] glowed very brightly in the sky[4]/water last night. (audience[3], woman[2])

28. The audience[1] laughed continuously at the clown[3]/cripple this afternoon. (stage[2], tissue[4])

29. The water[1] disappeared completely from the bowl[2]/hat this morning. (fabric[4], mom & son[3])

30. The candles[1] flickered for a while on the table[4]/refrigerator on Thursday. (mom & son[3], toddler[2])

31. The ghost[1] appeared out of nowhere in the old house[3]/grocery store at midnight. (candle[4], actress[2])

32. The ballerina[1] emerged gracefully out of the dancers[2]/prisoners last night. (doll[4], boot[3])

Dative Items

33. The newspaper[1] was difficult to read, but the mother[2] suggested it anyway to her teenager[4]/toddler last week. (dictionary[3])

34. The woman[1] wanted to start dating again, so her brother, who is a chef[2], introduced her graciously to his doctor[3]/priest Thursday evening. (candles[4])

35. An expensive[1] necklace was stolen during the night, and a female[4] witness described it in detail to the cop[2]/chef Friday morning. (fire[3])

36. The book[1] was short, and the girl[3] returned it with some others to the library[4]/the hospital this afternoon. (teacher & class[2])

37. The equation[1] was complex and the teacher[4] explained it carefully to her class[3]/doctor this morning. (ruler[2])

38. The couch[1] was very comfortable and the saleswoman[3] recommended it and another one to the executive[2]/beggar this morning. (tv[4])

39. The beer[1] was delicious and the thoughtful man[2] brought some from the bar to his wife[4]/baby Sunday afternoon. (fireworks[3])

40. The dolls[1] were wrapped when the man[4] sent them first class to his daughter[3]/grandparents this Wednesday. (fabric[2])

41. The shirt[1] was well-made and the tailor[3] showed it enthusiastically to the customer[2]/janitor last weekend. (needle[4])

42. The raw meat[1] was chopped before the zookeeper[2] offered it as breakfast to the lions[4]/fish each morning. (tiretool[3])

43. The book[1] was challenging and well-written, so the teacher[4] assigned it for homework to the class[3]/fireman Friday morning. (tools[2])

44. The house[1] was new, so the realtor[3] sold it for a high price to the couple[2]/children fairly quickly. (fireworks[4])

45. The idea was complex, but the scientist[1] mentioned it to the student[4]/gorilla right away. (microscope[2], tiretool[3])

46. The grass[1] was freshly cut and the boy[4] fed it generously to the cow[3]/cat in the afternoon. (sky[2])

47. The money[1] was stolen, and the guilty child[3] confessed it under pressure to the police officer[2]/convict on Tuesday. (tools[4])

48. The television[1] was not the right size, so the frustrated customer[2] mailed it carefully to the store[4]/restaurant last week. (ruler[3])

Appendix B: Stimuli for Experiment 3

For each pair of stimuli, the first version uses a dative verb with a prepositional recipient and the second version uses an action verb with a prepositional benefactor. After the sentence number is a brief description of the pictures used. The order indicates their screen location: upper left, upper right, lower left, lower right. Where there are multiple terms listed for the instrument picture, all of terms were used in the WWW co-occurrence analyses. Item number 11 was presented to participants, but omitted from the data analysis.

1. (pasta, teenage boy, oven mitt/potholder, girl)
The pasta was piping hot, so John recommended it wholeheartedly to his sister.
The pasta was piping hot, so John held it securely for his sister.

2. (house, broom, newlyweds, realtor)
The house was dirty inside, but the realtor sold it easily to some newlyweds**.**
The house was dirty inside, but the realtor swept it herself for some newlyweds**.**

3. (apartment, soldier, landlord, paint roller)
The apartment had been unoccupied, so the landlord offered it free of charge to a soldier who needed a place to stay.
The apartment had been unoccupied, so the landlord painted it graciously for a soldier who needed a place to stay.

4. (window, handyman, tools, owners)
One window was broken, so the handyman mentioned it right away to the owners.
One window was broken, so the handyman fixed it hurriedly for the owners.

5. (steak, knife, dog, child)
The steak was tough so the child fed it furtively to the dog.
The steak was tough so the child cut it up clumsily for the dog.

6. (orchestra, elderly couple, Tim, video camera)
The symphony was extremely good, and Tim described it vividly to his parents.
The symphony was extremely good, and Tim recorded it successfully for his parents.

7. (piñata, principal, stick, children)
The piñata was unusually sturdy, so the elementary school principal assigned it tentatively to the second graders.
The piñata was unusually sturdy, so the elementary school principal broke it himself for the second graders.

8. (diagram, pencil, wife, convict)
The plan for the robbery was nearly complete, but the convict confessed it nervously to his wife.
The plan for the robbery was nearly complete, but the convict changed it anyway for his wife.

9. (cow, vet, woman in scrubs, thermometer/stethoscope)
The cow was feverish, so the intern showed it with concern to the veterinarian.
The cow was feverish, so the intern examined it carefully for the veterinarian.

10. (blazer, seamstress , sewing machine/needle, tailor)
The blazer didn't fit properly, so the seamstress sent it once again to the tailor.
The blazer didn't fit properly, so the seamstress repaired it hastily **f**or the tailor.

11. (bacteria, microscope, professors, scientific technician)
The bacteria were complex, and the technician explained them thoroughly to the professors last Thursday.

The bacteria were complex, and the technician studied them painstakingly for the professors last Thursday.

12. (desk, customers, salesman, dolly/handtruck/cart)
The desk was very spacious and the salesman suggested it confidently to the customers.
The desk was very spacious and the salesman moved it expertly for the customers.

13. (tiger, tribesmen, spear, tourists)
The tiger was a constant threat, but the tribesmen displayed it guardedly to the tourists.
The tiger was a constant threat, so the tribesmen killed it triumphantly for the tourists.

14. (dolls, soap, girl, woman)
The dolls were old but the mother presented them lovingly to her daughter.
The dolls were old but the mother washed them delicately for her daughter

15. (vase, businessman, maid, paper towel/cloth,)
The expensive vase had a chip in it, so the maid brought it nervously to the executive.
The expensive vase had a chip in it, so the maid wiped it fretfully for the executive.

16. (baseball, Barry Bonds, bat, boy)
The baseball had been used when Barry Bonds made his 70th homerun, and he mailed it generously to the little boy with cancer.
The baseball had been used when Barry Bonds made his 70th homerun, and he hit it tearfully for the little boy with cancer.