Paying Attention to Attention: Perceptual Priming Effects on Word Order

Rebecca Nappa (nappa@sas.upenn.edu)
Department of Psychology, 3401 Walnut St.
Philadelphia, PA 19104 USA

David January (djanuary@sas.upenn.edu)
Department of Psychology, 3401 Walnut St.
Philadelphia, PA 19104 USA

Lila Gleitman (gleitman@sas.upenn.edu)
Department of Psychology, 3401 Walnut St.
Philadelphia, PA 19104 USA

John Trueswell (trueswel@sas.upenn.edu)
Department of Psychology, 3401 Walnut St.
Philadelphia, PA 19104 USA

Abstract

Two experiments are reported which examine how manipulations of visual attention affect adult speakers' linguistic choices regarding word order and verb use when describing simple visual scenes. Participants in Experiment 1 were presented with scenes designed to elicit the use of one of two perspective verbs (e.g., "A dog is chasing a man"/"A man is running from a dog"). Speakers' visual attention was manipulated by preceding the display with a crosshair positioned on one or the other character. Cross-hair position affected word order and verb choice in the expected direction. Experiment 2 replicated this effect with a subliminal attention-capture cue, and results were further extended to the order within conjoined noun phrases in sentential subjects ("A cat and dog are growling…"). The findings have important implications for incremental theories of sentence planning and suggest some specifics for how joint-attention might serve as a useful cue to children learning verbs.

Introduction

What makes people say what they say? This is a complex question, which has been the source of much investigation and dispute over the past several decades. Early on in the generative linguistic tradition, the emphasis on the productive and creative power of structural expression led many researchers to assume that properties of a visual stimulus can be related to a speaker's linguistic choices in only vague and theoretically uninteresting ways (e.g., Chomsky, 1957). Currently, though not disputing that one can say – or not say – many different things under the same environmental conditions, investigators doing experimental research on word order and structural choices in sentence production have concluded that some combination of perceptual, conceptual and linguistic accessibility contribute in a dynamic way to utterance planning. In particular, questions of word order have received much attention, and prompted much debate, as this issue must be richly intertwined with the planning of both an utterance's overarching message and the syntactic structure carrying that message. In advance of speaking, one must somehow decide where the upcoming utterance is to start, and much of how it is to proceed. A number of factors seem to contribute to this process.

First, studies have found a crucial role for preferred (i.e. primed or otherwise accessible) syntactic structures in the form a message ultimately takes (e.g. Bock & Loebell, 1990). Additionally, lexical/conceptual factors (e.g. accessibility, animacy) have been shown to affect word order materially, even at the expense of a preferred syntactic structure (Tversky, 1977; Bock, 1986; MacDonald, Bock & Kelly, 1993).

The role perceptual prominence plays in word and/or constituent order, however, seems a bit more nebulous. Within the literature on visual attention, it is quite clear that perceptual cues are involved in the interpretation of visual stimuli; research on perception of ambiguous figures (e.g. duck/rabbit, wife/mother-in-law) has shown that the perception of such stimuli can be driven by localizing eye gaze on critical features of a given interpretation (Georgiades & Harris, 1997). And perceptual factors (e.g. size, color) are clearly involved in ordering within simple conjoined noun phrases (e.g. *A bear and a dog*) (Osgood & Bock, 1977; Gleitman, Gleitman, Miller & Ostrin, 1996), but the role of perceptual prominence in constituent order remains unclear. Some find no relationship between initially fixated stimuli and subject role assignment (Griffin & Bock, 2000), while others find evidence supporting a role for attention (perceptual prominence) in constituent order (Tomlin, 1997; Forrest, 1996).

Some have interpreted these latter results as evidence for an incremental account of language production, in which a speaker builds an utterance as it is produced, and is apt to begin with whichever sentential elements are most salient at the time of speech onset. This account contrasts with more structuralist version of sentence planning, in which the underlying message of an utterance must be wholly planned prior to the onset of speech, and which accounts for the robust and reliable effects of syntactic priming (see Bock, in press, for discussion).

A troubling issue with all prior investigations into perceptual prominence and word ordering arises, however, if one examines the methodology. Manipulations have all been overt attention-getting devices (raising demand characteristic concerns), and have often had rigid task demands allowing for minimal generalization.

The current research investigates the question of perceptual contributions to word and constituent order, drawing on the attention and perception literature for more suitable methods. In two experiments, subjects' attention was directed subtly (Experiment 1) and then subliminally (Experiment 2) to scene participants, to determine whether perceptual cues under these covert conditions have any effect on the linguistic choices speakers must make. If such perceptual factors lead subjects to differing descriptions of identical scenes, a clear role can be established for attentional factors in sentence planning and constituent order. Such results may also provide evidence for an incremental approach to production, or perhaps, rather, to message planning. Finally, as we describe later, these effects may rebound on aspects of word learning.

Experiment 1

In the spirit of the afore-mentioned perceptual attention research on ambiguous figure resolution, our first investigation of attentional effects on event interpretation used a simple crosshair fixation point – prior to stimulus presentation – to direct a subject's eye gaze to a scene participant (analogous to directing gaze to a set of critical features in the ambiguous figure literature). Stimuli were designed to elicit one of two word order and verb choices on the part of the speaker, thereby making one or the other character in a scene the subject of the sentence. If initial visual attention subtly alters a speaker's perspective on the scene, we should expect that the speaker's choice of sentential subject and verb would be influenced by our attentional manipulation.

Methods

Norming and Stimuli Prior to initiating data collection on an attention-manipulating task, the specific stimuli to be used were normed, to identify baseline rates of verb selection for these particular items. Twenty-one monolingual English-speaking University of Pennsylvania Intro Psychology students participated for course credit. Subjects were presented with the 52 pictures to be used in experiment one, and asked to describe the event that was taking place in the scene using a simple sentence. No other manipulations or cues were introduced. Of these 52 pictures, twelve depicted pairs of so-called Perspective Verbs (e.g. chase/flee, see Figure 1), and these were the critical items (PVs).

Rates of verb use for these twelve items varied (see Table 1), but for each verb pair, subjects showed some degree of bias towards one interpretation and/or verb choice; there was a preferred verb and a dispreferred verb, and hence a corresponding preferred subject and dispreferred subject (passives were rare, occurring only 6 times across all 252 items). Overall, preferred subjects and verbs were used 69% of the time, dipsreferred subjects and verbs were used 27% of the time.

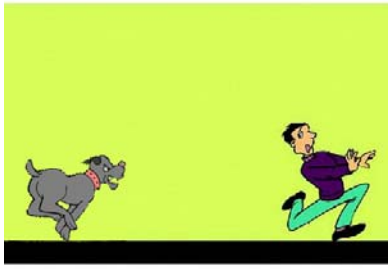

Figure 1: Sample Perspective Verb item from Experiment 1.

Table 1: Norming study baseline rates of verb usage in PV stimuli in Experiment 1. Percentage of total usage across all utterances in parentheses.

Item	Preferred Verb	Dispreferred Verb
Buy/sell	Sell (62)	Buy (38)
Chase/flee (dog/man)	Flee (57)	Chase (43)
Chase/flee (rabbit/elephant)	Chase (71)	Flee (29)
Eat/feed (puppies/dog)	Feed (76)	Eat (24)
Eat/feed (child/mother)	Feed (95)	Eat (5)
Give/receive	Give (71)	Receive (29)
Listen/talk (office)	Talk (76)	Listen (24)
Listen/talk (phone)	Talk (19)	Listen (29)
Perform/watch (singer)	Perform (67)	Watch (33)
Perform/watch (speaker)	Perform (86)	Watch (14)
Win/lose (boxing match)	Win (95)	Lose (5)
Win/lose (race)	Win (48)	Lose (24)

Participants and Design Eighteen monolingual English-speaking Introductory Psychology students at the University of Pennsylvania participated in this study for course credit. There were three conditions, defined by the location of the crosshair fixation point prior to scene presentation: Dispreferred (where the dispreferred subject would appear), Preferred (where the preferred subject would appear), and Middle (a neutral middle region, as a control). Manipulations were within-subjects, with each subject's gaze directed to the dispreferred subject on four of the twelve critical items, to the preferred subject on four of the

twelve critical items, and to a neutral middle region on the remaining four items.

Procedure Subjects in this experiment were presented with 52 scenes depicting participants engaged in a given activity (e.g. a picture of a boy swimming), including the twelve critical items, depicting perspective verb pairs. Subjects were instructed to describe each picture using one simple sentence, and subjects' utterances throughout the task were recorded.

A crosshair fixation point preceded presentation of each of the 52 scenes. This fixation point was presented on-screen for approximately 500 msec, then immediately followed by presentation of the scene (either filler or trial). (Earlier pilot work with an eyetracker confirmed that subjects followed directions and routinely fixated the cross prior to stimulus presentation.) Subjects in the current study were misled to believe that position of the crosshair was random and irrelevant to their task, so as to prevent their eyes from inspecting scenes in the same fashion on each trial. Position of the crosshair in fact corresponded directly to position of an upcoming scene participant. Although some subjects noted that the fixation marker frequently had been where an object appeared, no subject reported noticing the correlation between the location of scene participants and the crosshair. And in post-experimental interviews, most subjects who bothered to posit a guess as to the experiment's purpose speculated that it pertained to color brightness and/or interpersonal relationships of scene elements.

Results and Discussion

Rate of preferred verb usage was highly influenced by cue location in the expected direction (see Figure 2). In particular, when the preferred subject (e.g., the dog) was visually cued, speakers uttered on 77% of the trials sentences like "A dog is chasing a man." When the dispreferred subject was cued, however, speakers produced such utterances only 61% of the time (and showed a corresponding increase in utterances like "The man is running from the dog"). Analyses of Variance (ANOVAs) on participant and item means revealed that the effect of cue location was significant (both p's<.05).

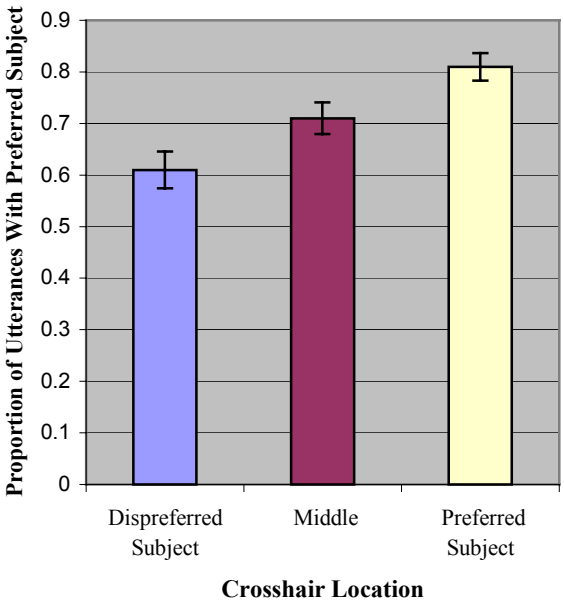

Figure 2: Proportion of utterances beginning with the preferred subject and verb, by condition, in Experiment 1.

Experiment 2

Following the results of Experiment 1, a couple of questions arose. First were concerns regarding demand characteristics of the crosshair fixation point manipulation. Although subjects did not seem to sense the specific purpose of the experiment (namely, subject and verb selection), many noticed that the crosshair's position frequently corresponded to an object in the upcoming scene. We worried that this knowledge alone might have subtly influenced their linguistic choices. To this end, we developed an attention-capture cue (see Jonides & Yantis, 1988), as discussed below, which successfully directed subjects' attention to a particular region of the scenes, without being consciously perceptible.

Secondly, as discussed previously, much prior research on sentence production and linguistic choice has compared the role of many different factors, from animacy to size of entities, on differing constructions. Specifically, different variables seem to contribute differently to linguistic choice in simple conjoined noun phrases (e.g. *the dog and the man* vs. *the man and the dog*) than to linguistic choices involving thematic role assignment (e.g. *the dog chased the man* vs. *the man fled the dog*). In Experiment 2, we wanted not just to replicate our prior result, but also to compare the influence of our covert attention-capture manipulation on these sorts of different constructions. To this end, twelve additional items were added in Experiment 2, depicting events in which two scene participants were engaging in an activity together (see Figure 3, designed to elicit "The cat/dog and the dog/cat are growling at each other"). These Conjoined Noun Phrase (CNP) items were aimed at eliciting descriptions containing a conjoined noun phrase in the sentential subject position (e.g. A dog and a cat are growling), so as to investigate the effect of our covert manipulation on word order in a simple conjoined noun phrase.

Figure 3: Sample Conjoined Noun Phrase item from Experiment 2.

Methods

Norming and Stimuli CNP pictures were first normed for baseline preferences. Twenty-one monolingual English-speaking University of Pennsylvania Intro Psychology students described 64 scenes (52 fillers, and the 12 CNP items), absent any manipulations or cues. In an effort to avoid utterances beginning with uninformative sentential subjects (e.g. "Two people are…"), CNP stimuli consisted of scenes with animal, rather than human, participants (e.g. a dog and a cat growling, see Figure 3).

Baseline rates of first-mentioned scene participants varied less than with the PV stimuli in Experiment 1 (See Table 2). Most items were relatively unbiased, but scene participants with even a slight advantage were dubbed *Preferred First-Mentioned*, and referred to as such from this point onward, for the sake of simplification. Overall, preferred first-mentioned participants were mentioned first 56% of the time, dispreferred first-mentioned only 44%.

Table 2: Norming study baseline rates of first-mentioned participants in CNP stimuli in Experiment 2. Percentage of total usage across all utterances in parentheses.

Item	Preferred First-mentioned	Dispreferred First-mentioned
Biking	Turtle (61.9)	Dog (38.1)
Dancing	Fish (57.1)	Bear (42.9)
Eating	Koala (52.4)	Panda (47.6)
Growling	Cat (52.4)	Dog (47.6)
Juggling	Elephant (52.4)	Seal (47.6)
Jumping	Frog (57.1)	Cat (42.9)
Playing cards	Pig (57.1)	Dog (42.9)
Playing horns	Rhino (52.4)	Snail (47.6)
Rowing	Bear (52.4)	Snowman (47.6)
Skating	Monkey (57.1)	Rabbit (42.9)
Swinging	Elephant (61.9)	Monkey (38.1)
Waiting	Penguin (52.4)	Deer (47.6)

An additional consideration that arises when adding the CNP stimuli is orientation. As previously mentioned, one factor driving word order in conjoined noun phrases is the left-to-right bias, with leftmost participants more likely to be first mentioned. This prediction bore out in the current norming study as well, with leftmost participants mentioned first 78.2% of the time for CNP items (as compared to only 52.8% of the time for PV items in prior norming study).

Participants and Design Forty monolingual English-speaking Introductory Psychology students at the University of Pennsylvania participated in this study for course credit. Both the location of the attention-capture cue and the left-to-right orientation of the scene were systematically varied, creating a 2 X 2 design (cued participant X leftmost participant) and four stimulus lists. Manipulations were within-subjects, with each subject assigned randomly to one of these four lists.

Procedure Subjects in this experiment were presented with 64 scenes: the same 40 fillers and 12 PV scenes used in Experiment 1 and the 12 normed CNP scenes. Subjects were instructed to describe each picture using one simple sentence, and subjects' utterances and eye movements were recorded throughout the task.

Prior to stimulus presentation, subjects fixated a crosshair fixation point (equidistant from the two scene participants) for 500 msec. Subjects were misled to believe that position of the crosshair was randomized, to assist the experimenters in maintaining eyetracker calibration accuracy (no subject reported suspecting anything otherwise). The fixation point was then followed by a brief, covert attention-capture manipulation. This manipulation consisted of a small black target area (subtending an area of approximately 0.5X0.5 degrees of visual angle) against a white background, with a duration of 60-80 msec, followed immediately by the stimulus. Although no subject reported noticing the subliminal cue, it was highly effective in capturing attention. Subjects looked first to the cued location a median of 76% of the time.

Results

Table 3 shows rates of mentioning the Preferred First-Mentioned participant first for the CNP stimuli, and Table 4 shows rates of using the Preferred Subject for the PV stimuli for all four conditions in the 2X2 design. Collapsing across sentence types, significant effects of Left-Right Position and Attention-Capture were observed; leftmost and cued entities were more likely to be first-mentioned (p's<0.01). Further analyses showed that Left-Right orientation was significant only for word order in CNP stimuli ($p<0.01$), not for subject selection in PV items. Both sentence types, however, showed significant, stable effects of Priming, with primed characters more likely to appear first in CNPs ($p<0.05$) and to be the subject of a perspective verb ($p<0.01$).

Table 3: For all four conditions of Conjoined Noun Phrase stimuli, proportion of utterances in which subjects mentioned Preferred First-Mentioned participant first

	Preferred First-Mentioned Primed	**Dispreferred First-Mentioned Primed**	Average
Preferred First-Mentioned on Left	79.3%	63.8%	**71.6%**
Dispreferred First-Mentioned on Left	58.1%	41.4%	**49.7%**
Average	**68.7%**	**52.6%**	

Table 4: For all four conditions of Perspective Verb stimuli, proportion of utterances in which subjects mentioned used Preferred Subject

	Preferred Subject Primed	**Dispreferred Subject Primed**	Average
Preferred Subject on Left	87.4%	66.4%	**76.9%**
Dispreferred Subject on Left	77.3%	60.1%	**68.7%**
Average	**82.3%**	**63.2%**	

General Discussion

Language Production Overall, our results show a role for perceptual prominence in constituent ordering, and may be taken as support for a more incremental approach to sentence production.

It is important, however, to keep these results in the context of the current literature on the subject of speech production. Although Griffin and Bock (2000) found no correlation between first-fixated scene participants and first-mentioned participants, in an extensive investigation into the time course of message extraction from a visual scene, they *did* show tightly linked eye movement and speech patterns once an utterance was to begin; subjects looked reliably to an object less than a second before producing the corresponding word. This, and other research in this vein (Bock, Irwin, Davidson & Levelt, 2003), implies a system that begins with an initial, message-planning stage, followed by a more incremental process of retrieving the necessary lexical elements to construct an utterance (see Bock, in press, for discussion).

Our result is in no way inconsistent with this model of speech production. It is quite possible that subjects in our studies, rather than beginning to incrementally code their final utterance at the onset of the stimulus, begin with an information-extracting, message-planning stage, and that the perceptual priming effects we see take effect in this early stage. In the analogous ambiguous-figure literature, such attentional manipulations seem to affect the way subjects *perceive*, or interpret a stimulus. This may well be what's resulting from our similar attention-driving tools: a different perception, or interpretation of the stimulus. Ongoing research will investigate the effects of the same perceptual prime on both transitive verbs – where subjects must shift to an infrequent, passive structure to alter subject role assignment – and symmetrical predicates – where prominent information tends to appear in the object role/position (e.g. "I met Meryl Streep" vs. "Meryl Streep met me") (Gleitman et al., 1996). These explorations into the underlying nature of the perceptual prime should begin to determine where and how it is having its effect.

Language Acquisition These results have interesting implications for word learning studies as well. It has been noted that perspective verb pairs should be specifically very difficult for children acquiring a language to learn, as in many cases both members of these pairs necessarily co-occur under the same situational circumstances (Gleitman, 1990; Fisher, Hall, Rakowitz & Gleitman, 1994); for instance, a child is not apt to be presented with a situation that involves chasing but, at the same time, does *not* involve fleeing, and vice versa. How can the young learner figure out, then, whether the mother was saying "chase" or "run away?" These studies showed that syntactic information can inform the listener/learner as to the speaker's intended meaning. By varying the syntactic frame in which a novel verb appeared while referring to a perspective verb stimulus (e.g. "The man is glorping the dog" vs. "The dog is glorping the man," with regard to Figure 1) Fisher et al. showed that young listeners are quite adept at using this syntactic input, or "zoom lens," to arrive at the same interpretation intended by the speaker.

Another "zoom lens" that is more closely related to the present studies, is joint visual attention of speaker and listener. Infants as young as 2-months-old engage in such gaze-following activities (Bruner, 1998), looking where an adult is looking, during conversation. Moreover, by 12 to 18 months of age, the infant can successfully use this gaze-

direction information as a cues for how to label new objects (Baldwin, 1993). Contributions of attentional cues to word learning have not been as broadly or rigorously investigated for the case of verb learning. Given our current result on the relationship between attention-direction and variation between subject and verb choice, we suggest that similar attentional cues are available to the young language learner in successfully parsing and interpreting speech as well, even in the especially difficult case of perspective verbs.

Conclusion

Taken together, the results of these two experiments as they interface with relevant prior investigations clearly demonstrate a relationship between attention and language production. Further investigation will be necessary to delve into the detailed nature of this relationship, and explore the way it fits into a model of language production. These results, though, and the implications they have for attentionally-aware young language learners trying to interpret the speech stream, open exciting new investigative doors in both language production and acquisition.

References

Baldwin, D. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*, 395-418.

Bock, J.K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 575-586.

Bock, J.K., Irwin, D., Davidson, D. & Levelt, W.J.M (2003). Minding the clock. *Journal of Memory and Languag*e, *48*, 653-685.

Bock, J.K., Irwin, D.E. & Davidson, D.J. (in press) Putting First Things First. In J. M. Henderson & F. Ferreira (Eds.), *The integration of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.

Bruner, Jerome S. (1998). Routes to reference. *Pragmatics & Cognition*, *6,* 209-227.

Chomsky, N. (1957). *Syntactic Structures.* Oxford, England: Mouton.

Fisher, C., Hall, D.G., Rakowitz, S. & Gleitman, L. (1994). When It Is Better to Receive Than to Give: Syntactic and Conceptual Constraints on Vocabulary Growth. *Lingua, 92*, 333-375.

Forrest, L. B. (1996). Discourse goals and attentional processes in sentence production: The dynamic construal of events. In A. E. Goldberg (Ed.), *Conceptual structure, discourse and language*. Stanford, CA: CSLI Publications.

Georgiades, M. & Harris, J.P. (1997). Biasing effects in ambiguous figures: Removal or fixation of critical features can affect perception. *Visual Cognition, 4,* 383-408

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1,* 3-55.

Gleitman, L., Gleitman, H., Miller, C. & Ostrin, R. (1996). Similar and similar concepts. Cognition, 58, 321-376.

Griffin, Z.M. & Bock, J.K. (2000). What the eyes say about speaking. *Psychological Science, 11*, 274-279.

Jonides, J. & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics, 43,* Apr 1988, pp. 346-354

MacDonald, J.L., Bock, J.K. & Kelly, M.H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology, 25,* 188-230.

Osgood, C.E. & Bock, J.K. (1977). Salience and Sentencing: Some Production Principles. In Sheldon Rosenberg (Ed). *Sentence Production: Developments in Research and Theory*. Hillsdale, N.J.: Erlbaum.

Tomlin, R.S. (1997). Mapping conceptual representations into linguistic representations: The role of attention in grammar. In Nuyts, Jan & Pederson, Eric (Eds). *Language and conceptualization. Language, culture and cognition.* New York: Cambridge University Press.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352