



Incremental interpretation at verbs: restricting the domain of subsequent reference

Gerry T.M. Altman,
Yuki Kamide



Motivation: Background

- Incremental Processing

Referring expressions (*the ornate red vase*)

Adjectives as constraints (narrowing) on the range of entities denoted by the subsequent noun *vase*

Set of available referents narrowed down as soon as info becomes available, even before head noun

(Altman & Steedman, 1988; Sedivy, Tanenhaus, Chambers & Carlson, 1999)

(Eberhard, Spivey-Knowlton, 1995; Sedivy, Tanenhaus, Chambers & Carlson, 1999)



Motivation: Assumption

- Assumed that aspects of grammatical knowledge are lexicalized and accessed together with other aspects of a verb's meaning (Bresnan, 1982; MacDonald, Pearlmutter & Seidenberg, 1994)



Motivation: Research Question

- The primary interest was in determining when, relative to verb onset, the participant's eyes first moved to the target object



Motivation: Hypothesis

- Sentence processing is driven by the predictive relationships between verbs, their syntactic arguments, and the real-world contexts in which they occur.
- Semantic information extracted at the verb is able to guide visual attention towards an appropriate object in visual context (as determined by that semantic information) even before the semantic properties of the direct object become available.



Motivation: Why test this hypothesis?

- Information extracted at verbs identifies roles and positions within the sentence where the recipients of those roles can be identified as well as identifies directly the (real or mental world) entities that play out those roles
- Although referring expressions do generally convey sufficient information with which to uniquely identify the intended referent, there are occasions when verbs convey such uniquely identifying information also.



Motivation: Why this is interesting

- Reference to (visual world) context can be achieved in an **INCREMENTAL**, piecemeal manner, using information to narrow down the set of available referents as soon as that information first becomes available — even before the head noun (*vase*, in the example above) is itself encountered.



Methodology

- Both experiments consisted of 24 University students who were either given £2 or school credit. None had corrected vision.
- 16 sets of stimuli (audio recording, and picture) two sentences for each visual scene
- 1: had a verb with selectional restrictions limited the object of the sentence to one visual item
- 2: had a verb with selectional restrictions allowing four visual items.
- Both referred to the target item in the subject position of the sentence.
- 16 sets sentences were constructed as filler items
 - two sentences with the same verbal structure as the first set, but the direct objects were unrelated to the visual scene



Methodology

- They wanted a natural reaction time from verb onset and offset and determiner onset and noun onset
- No indication of when and how eye-movement was to respond to the sentence given.
 - Indeed a distractor task was presented with each scene.
- Eye movement to the presented visual items were consciously divorced from the mental task at hand.
- Saccadic eye-movement was intended to be non-anticipatory



Implementation

- Participants sat in front of a monitor wearing a head mounted binocular eye-tracker.
- Head movement was unrestricted.
- For each visual scene participants were asked to press a button responding yes or no as to whether or not the sentence they heard was plausible the sentence
- There was no mention of speed needed to respond to each task
- The eye-tracker was re-calibrated in between each set of stimuli



Implementation

- Each sequence was terminated after each response and in no instance was the sequence terminated before the audio track was completed
- Completion of each participant took approximately twenty minutes.
- Temporal markers were used in each recording to reference points of interest in each sentence in order to time eye-movements
- In Experiment 2 all else was identical except for direction given to not pay particular attention to the sentence dictated



Results

- Eye fixation on target item during verb onset was discluded from analysis
- Participants data was analyzed only if eye-movement went away from target item and back towards it
- Data was analyzed and charted based on an average of cumulative probability
- Eye fixations beyond 3 or 4 pixels beyond determined contouring coordinates of visual item were not deemed fixations.

Table 1

Experiment 1 (Section 2): onset of first saccade to the target object in both the `eat' and `move' conditions, relative to verb onset, verb offset, determiner onset and noun onset (timings in ms)

	Eat	Move	Difference (move - eat)
Verb onset	611	838	227
Verb offset	228	415	187
Determiner onset	37	234	197
Noun onset	-85	127	212

Table 2

Word durations for the `eat' and `move' sentences (timings in ms)

Duration	Eat	Move	Difference (move - eat)
Verb	383	423	40
Post-verbal break	192	180	-12
Determiner	122	107	-15
Verb + break + determiner	697	710	13

Table 3

Experiment 2 (Section 3): onset of first saccade to the target object in both the `eat' and `move' conditions, relative to verb onset, verb offset, and noun onset (timings in ms)


	Eat	Move	Difference (move - eat)
Verb onset	988	1246	258
Verb offset	605	823	218
Determiner onset	413	643	230
Noun onset	291	536	245

[Link to Paper/Graphs](#)



Results: Experiment 1

- Participants looked at the target object after the verb onset %90 of the time, (%88 of the time for eat, %92 of the for move)
- Participants looked at the target object before the onset of the noun %38 of the time for the move condition and %54 for the eat condition




QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.



Results: Experiment 2

- Participants looked at the target object 93% of the time (equal percentages for both *eat* and *move* conditions).
- The first saccade to the target object occurred before the noun 18% of all trials for the *move* condition and 32% of *eat* trials.



QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.



Criticism: opposing view that they are arguing against

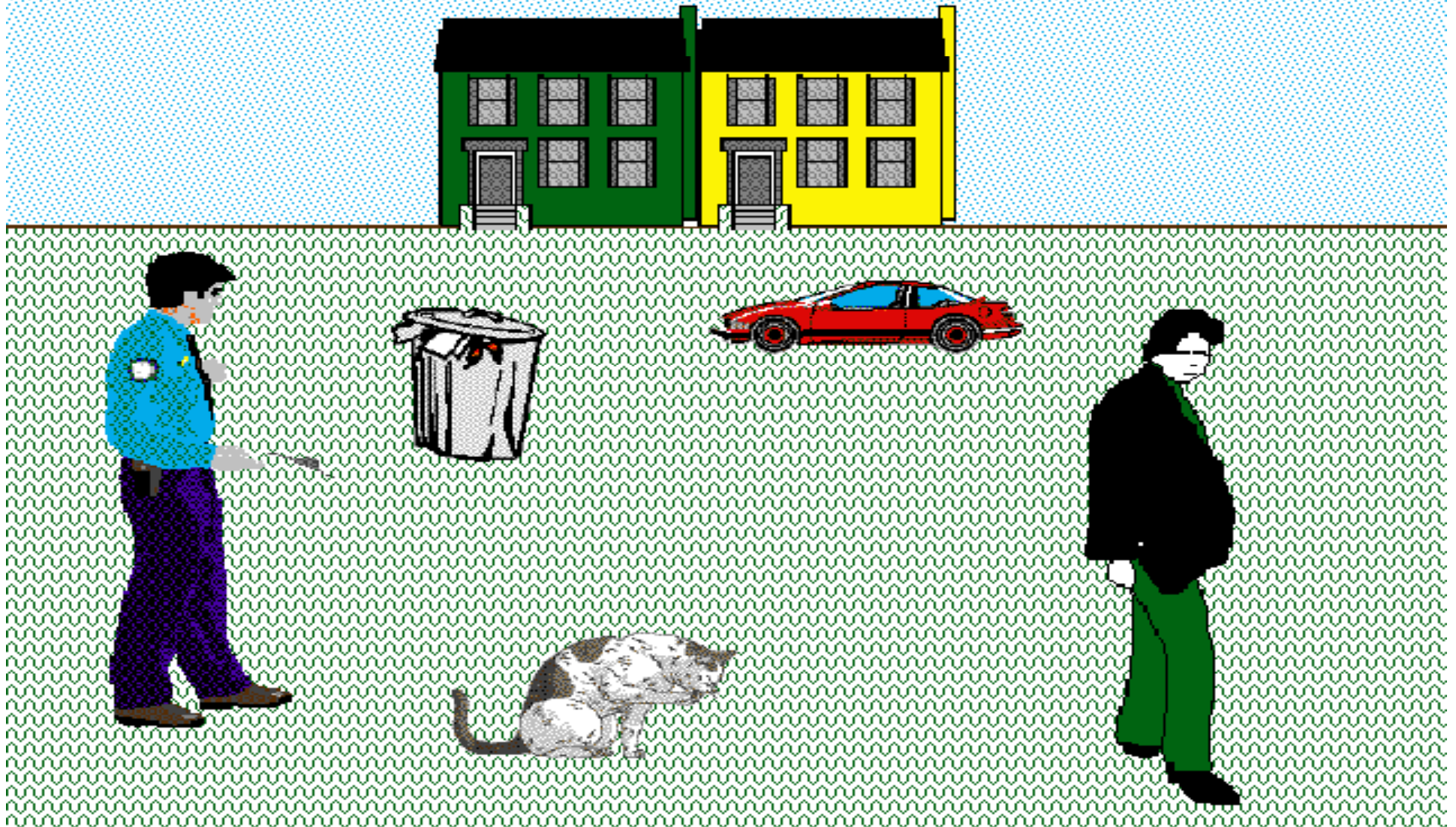
Unlike the word-by-word stop-making-sense judgment task employed by Altmann (1999), the experiments we report below did not require any artificial segmentation of the linguistic input, and in the case of Experiment 2, did not require participants to do anything other than look-and-listen.



Criticism

- Colors used
- Systematic Target Factor

The policeman will arrest / search the man. [car, dustbin, houses, cat]



The woman will bathe / touch the baby. [plant, rocking-horse, stool]





Criticism: problems with their methods or conclusions

- It is conceivable, however, that our data reflect the exigencies of the judgment task – normal language comprehension does not usually require meta-linguistic judgments, and the requirement to make such judgments may have induced anticipatory processing strategies which do not reflect normal processing.
- We cannot rule out the possibility that (some) participants may have interpreted the task as requiring some form of (implicit) meta-linguistic judgment which may have contributed to the overall pattern of results we observed



Criticism: opponents would say

- Concerning the consequences of verb-mediated reference, it is possible that information at the verb does nothing more than restrict the domain of (subsequent) reference. Thus, although the programming of eye movements in our study towards the appropriate visual referent was initiated within perhaps as little as 30ms of verb offset, this need not mean that the processor had already assigned to that visual entity the role associated with whatever was about to be referred to in grammatical object position



Criticism: Future Research

- Visual then audio
- Audio then visual



Questions/Concerns



Thanks,

Brian Robinson

Michael Levine

