OHIO STATE UNIVERSITY WORKING PAPERS IN LINGUISTICS 38 (54-66)

Illicit Acceptability in *picture* NPs[1]

Wayne Cowart

University of Southern Maine

Abstract: Four experiments examine the interaction between extraction and specificity in **picture** NPs. The results indicate that the acceptability judgements of naive speakers show highly robust patterns that do not conform well to widely held assumptions about the relative acceptability of several theoretically important kinds of sentence. There is also evidence that the difference between argument and non-argument extractions has a marked impact on acceptability (though no such acceptability difference has figured in linguistic theory). Further, the paper argues that there are circumstances in which ungrammatical sentences may be rendered acceptable via the intrusion of extragrammatical mechanisms in comprehension. Thus, the acceptability of these sentences is 'illicit'.

1. Introduction

The empirical generalization addressed by Chomsky's subjacency principle seems to apply to sentences such as (1) (see Chomsky 1973, 1981, 1986; also Riemsdijk and Williams 1986, Lasnik and Uriagereka 1988).

(1) Who did the Duchess sell a portrait of?

Nevertheless sentences such as this seem fully acceptable and intelligible to many speakers. In consequence, such cases have been regarded as fully grammatical in the linguistic literature.

Broadly speaking, there are two approaches by which a grammatical theory that incorporates some form of the subjacency constraint might accommodate these facts. First, the components of the grammatical theory that capture the subjacency constraint might be formulated in such a way that they do not apply to cases such as (1). Second, the grammar might incorporate principles that have the effect of shielding cases like (1) from the subjacency mechanism(s).

This study takes a different tack, looking more closely at the intuitions about that underlie the assumption that it is grammatical. The study begins from the

---

[1] This paper originally appeared in Wiltshire, Gracyk and Music, eds. (1989). *Papers from the 25th Annual Regional Meeting of the Chicago Linguistic Society, Part One: The General Session.* Chicago: Chicago Linguistic Society. Reprinted with permission of CLS.

tentative suggestion that the acceptability of (1) might be spurious, that (1) is in fact ungrammatical, despite its apparent high acceptability. This requires a more systematic approach to assessing intuitions about (1) and close consideration of various alternative cognitive mechanisms that might account for the acceptability of the cases in the face of possible ungrammaticality.

The work reported below assumes that, all other things being equal, it is reasonable to hold the grammar accountable for patterns of judged acceptability. Where observed patterns do not coincide with those predicted by some grammar, it is appropriate to seek an account of the discrepancy in the larger ensemble of mental resources that speaker/hearers bring to language comprehension, in the cognitive processes that realize specifically grammatical knowledge, in the grammar itself, or in some combination of these. These issues are discussed further at the end of the paper.

2. Experiment 1: Possible Subjacency Effects

The first experiment was designed to determine whether the patterns of judged acceptability obtained with sentences similar to (1) are in reasonable accord with the patterns of grammaticality commonly assumed in the formulation of grammatical theory.

2.1. Materials

The target cases are presented in Table I. The Control cases are taken to be uncontroversially acceptable and grammatical. The Specified Subject cases are equally uncontroversially unacceptable and ungrammatical. The status of the other two cases is unclear. The definite cases have sometimes been taken to be acceptable and sometimes unacceptable, with corresponding assumptions about grammaticality. The Indefinite cases are patterned on (1), discussed above. Prevailing assumptions about grammaticality in these cases predict that the Indefinite cases will pattern with the Controls, that the Specified Subject cases will be distinctly less acceptable, and that the Definite cases will pattern with either the first set or the second, depending upon whether they are in fact grammatical and acceptable.

Control:	Why did the Duchess sell Turner's portrait of her father?
Indefinite:	Who did the Duchess sell a portrait of?
Definite:	Who did the Duchess sell the portrait of?
Specified Subject:	Who did the Duchess sell Turner's portrait of?

Table I: Materials for Experiment 1.

There were 24 sets of materials modeled on those in Table I. There were also 96 filler sentences of diverse kinds. Four presentation lists of sentences were constructed so that only one member of each set appeared in each list and so that six items of each of the four types shown in Table I appeared in each list. Thus each subject saw equal numbers of items of each type distributed throughout a much larger list of fillers and no subject saw more than one member of any set.

2.2. Methods

For this and all subsequent experiments, the materials were presented to subjects as printed lists. Subjects were asked to read each sentences and to

> "...indicate whether the item seems like a fully normal, understandable sentence to you. If it does, please check the box on the far right. If, on the other hand, the sentence seems very odd, awkward or difficult to understand, please check the box on the far left. If your feelings about the sentence are somewhere between these extremes, check one of the middle boxes. THERE ARE NO 'RIGHT' OR 'WRONG' ANSWERS. Please base your responses solely on your personal judgments, not on rules you may have learned about what is 'proper' or 'correct' English." (emphasis in original)

For some experiments a separate machine-scored answer sheet was used. Subjects responded by way of a four point scale whose extremes were marked "Odd" and "OK". Subjects were encouraged to respond rapidly and typically finished the list of 120 sentences, plus additional background questions, in less than 15 minutes.

Subjects were undergraduate students at The Ohio State University.

2.3. Results and Discussion

The results of two separate runs of Experiment 1 with a total of 228 subjects are combined and summarized in Figure 1. The most important result is that the pairwise difference in acceptability between the Indefinite and Control cases is highly significant, $F(1,227)=459.27$, $p<.001$, and indeed is the single largest difference observed. The pairwise differences between the Indefinite and Definite cases and between the Definite and Specified Subject cases, though numerically smaller, are also highly robust, $p<.001$. The overall decline in acceptability across the four cases is significant, $F(3,681)=619.69$, $p<.001$, as is the decline across the Indefinite, Definite and Specified Subject cases, $p<.001$.

The pattern of results seen in Figure 1 does not fit well with typical assumptions about the relative acceptability of these kinds of cases. The Indefinite cases are worse than expected and the difference between them and the Definite and Specified Subject cases is smaller than expected. One possible view of the pattern

in Figure 1 posits that the Indefinite cases are ungrammatical in virtue of being covered by the subjacency generalization and that the decline in acceptability across the Indefinite, Definite and Specified Subject cases reflects increasing specificity in the determiners of the **picture** NPs (Fiengo and Higgenbotham 1981, Fiengo 1987). Two alternative hypotheses must be considered. First, the reduced acceptability of the three most impaired cases might result simply from the presence of a preposition at the end of the sentence, regardless of other aspects of the structure of the sentence. Perhaps the common prescriptivist ban on sentences ending in prepositions exerted some influence (despite the instructions to subjects to ignore such considerations). Second, somewhat similarly, it might be that the apparent effect of specificity is not sensitive to the structure in which the determiner occurs but is somehow induced by mere surface variation in the form of the determiner. Neither hypothesis is especially interesting from a linguistic point of view, but they cannot be dismissed out of hand. The second experiment addresses these issues.

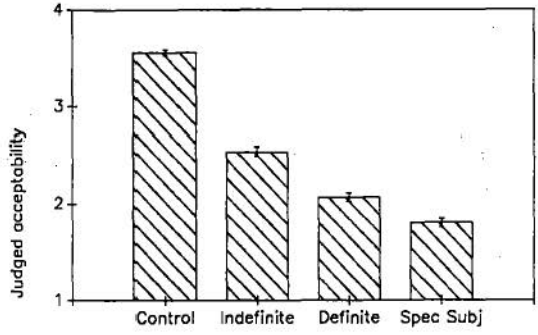

Figure 1: Mean judged acceptability for the four materials conditions of Experiment 1.

3. Experiment 2: *of* vs. *to*

The strategy of Experiment 2 is to compare cases similar to those used above to others where the prepositional phrase containing the extraction site has to as its head and is a sister of the NP to its left, rather than embedded within it. Such cases test the claim about sentence-final prepositions and provide an alternative control case against which to compare the Indefinite cases of Experiment 1.

3.1. Materials

A sample set of materials is displayed in Table II. The Control and Indefinite/of cases are drawn from Experiment 1. The Indefinite/to case should be identical to the Indefinite/of case if the depressed results in the former case results merely from the presence of a preposition at the end of the sentence. The contrast between the Indefinite/to and Specified Subject/to cases likewise tests whether the

apparent effect of specificity in Experiment 1 is sensitive to the structural relation of the NP and PP.

Control:	Why did the Duchess sell Turner's portrait of her father?
Indefinite/of:	Who did the Duchess sell a portrait of?
Indefinite/to:	Who did the Duchess sell a portrait to?
Specified Subject/to:	Who did the Duchess sell Turner's portrait to?

Table II: Materials for Experiment 2.

3.2. Results and Discussion

The most critical result of Experiment 2 is that there is a statistically robust difference between the acceptability of the Indefinite/of and Indefinite/to cases, $F(1,40)=12.93$, $p<.001$. Furthermore, the decline in acceptability from the Control case to the Indefinite/of case is reliably larger than the decline from the Control to the Indefinite/to case, $F(1,40)=12.93$, $p<.001$. This suggests that the unacceptability of the Indefinite/of cases is not to be explained solely by the presence of some preposition at the end of the sentence. This picture is somewhat clouded, however, by the finding that the Indefinite/to cases are also significantly less acceptable than the Control, $F(1,40)=10.20$, $p<.005$. This fact is not explained by either the subjacency or specificity proposals discussed above.

Some support for the specificity proposal is evident in the fact that the Specified Subject/to cases are more, rather than less, acceptable than the Indefinite/to cases. This is opposite to the pattern found in Experiment 1, though the difference here is not robust, $F(1,40)=1.04$, NS. When the PP is outside of the NP, increasing specificity does not compromise acceptability.

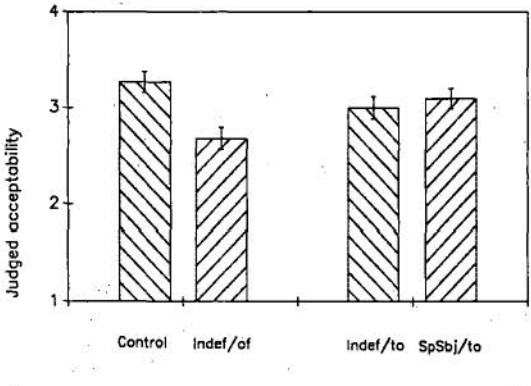

Figure 2: Mean judged acceptability for the four cases of Experiment 2.

The difference between the Control and Indefinite/of cases is robust, $F(1,40)=35.23$, $p<.001$, which replicates Experiment 1. The observed acceptability of the Control cases in Experiment 2 is somewhat lower than that obtained in Experiment 1. Such interexperiment

differences and, in general, the absolute numerical values of the acceptability means will not be considered here. There were small differences in the filler sentences used in different experiments, and other minor differences of technique, that may account for any such differences.

In sum, these results are consistent with the claim that the Indefinite cases of Experiment 1 are affected by subjacency and that the acceptability of similar extraction cases is sensitive to the specificity of the NP within which the PP is embedded.

4. Experiment 3: Depth of Embedding

Another possible confound relevant to the results of Experiment 1 is depth of embedding. Though there is no very clear and generally accepted metric of depth of embedding, nevertheless, it seems clear that the Indefinite/of cases involve extraction from a site that is more deeply embedded in the hierarchical structure of the sentence than is the comparable site in the Control sentences or the Indefinite/to cases. Counting only S and NP boundaries, the of cases involve extraction from two levels down, while the others require extraction from no more than one level down, as illustrated in Table III.

Why did [the Duchess sell [Turner's portrait of her father] *t*]
Who did [the Duchess sell [a portrait of *t*]]
Who did [the Duchess sell [a portrait] to *t*]
Who did [the Duchess say [Max likes *t*]]

Table III: Depth of embedding, bracketing only relevant NPs and Ss.

As a partial control for the possible influence of depth of embedding, Experiment 3 compared the Control/Indefinite contrast of Experiment 1 with pairs contrasting two degrees of embedding where the more deeply embedded case is uncontroversially regarded as acceptable in the linguistic literature, as in the last example in Table III.

4.1. Materials

One complete set of materials is illustrated in Table IV. The Subjacency Cases are drawn from Experiment 1. The Shallow Depth Cases involve extraction from the subject of the higher clause, while the Deep cases involve extraction of the object of the lower clause.

Subjacency Cases

Shallow: When did the Duchess sell Max's portrait of Bill?
Deep: Who did the Duchess sell a portrait of?

Depth Cases

Shallow: Who said Max likes George?
Deep: Who did the Duchess say Max likes?

Table IV: Materials for Experiment 3.

4.2. Results and Discussion

The results of Experiment 3 are displayed in Figure 3. Most importantly, there is a significant interaction between Depth (Deep vs. Shallow) and Sentence Type (Subjacency Cases vs. Depth Cases), $F(1,27)=5.69$, $p<.05$. This indicates that the difference between the two Subjacency cases is reliably larger than that between the two Depth cases. There was also a reliable main effect of Depth, $F(1,27)=46.69$, $p<.001$.

This pattern of results is not consistent with the claim that the Control vs. Indefinite contrast of Experiment 1 arose because of a depth of embedding difference between the two extraction sites.

Two notable features of these results are that the Shallow Depth cases are significantly less acceptable than the Shallow Subjacency cases (i.e., the Control cases of Experiment 1), $F(1,27)=27.94$, $p<.001$, and that there is no appreciable difference in acceptability between the two Deep cases. These observations suggest a partial alternative account of the pattern obtained in the first experiment. This account draws a sharp distinction between adjunct extractions (questions with **why**, **when**, or **where**) and those from argument positions (questions with **who**, **what**, or **which**), with the latter associated

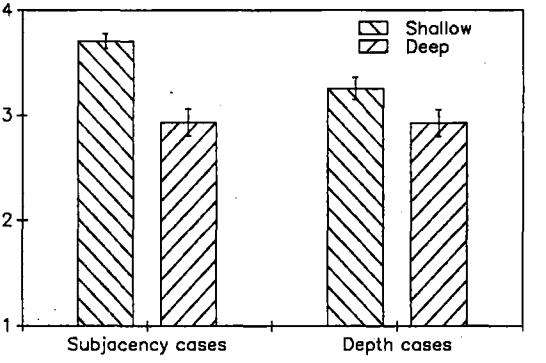

Figure 3: Mean judged acceptability for the four cases of Experiment 3.

with reduced acceptability and, presumably, greater difficulty. On this account, it was inappropriate in Experiment 1 to use non-argument extractions as controls against which to assess the acceptability of the three cases involving extraction from an argument position. It is, in other words, the distinction between argument and adjunct extraction that best explains the difference between the Control and Indefinite cases of Experiment 1, not the engagement of any effect related to subjacency. This, of course, does not explain the differences among the three cases of argument extraction, nor why the scale of those differences was less than that between the Control and Indefinite conditions.

Further experimental work that will help determine the source of the Control/Indefinite difference in Experiment 1, and which will control for extraction-type, is planned.

5. Illicit acceptability and its analysis

There are two puzzles implicit in the various findings discussed above. One is why some cases that are regarded as acceptable in the linguistic literature get such poor acceptability ratings with the methods used here.

Another puzzle is that, on either analysis of why the Indefinite cases of Experiment 1 are rated so poorly, some sentence types seem much better than might be expected from one or another point of view. If the results of Experiment 1 are interpreted as evidence that the Indefinite cases are covered by the subjacency generalization, it is surprising that these sentences have been seen as entirely acceptable in the literature. On the other hand, the low ratings of the Indefinites may be due simply to the fact that they involve extraction from an argument position, not to ungrammaticality. On this view, it seems surprising that the Definite and Specified Subject cases were not judged more negatively in Experiment 1. A proponent of this extraction-type analysis must somehow explain why there should be greater differences in judged acceptability when grammatical sentence types are compared to other grammatical types than when grammatical types are compared to ungrammatical types. Thus, from either point of view there seem to be sentences that enjoy a degree of illicit acceptability, i.e., surprisingly high degrees of acceptability associated with ungrammaticality.

Findings such as these suggest as yet undescribed complexities in the relation between strictly grammatical phenomena and the judgments of acceptability on which this study is based. Experiment 4 constitutes a preliminary attempt to identify one potential source of that complexity.

There is a widely held view of the relation between performance phenomena, such as acceptability judgments, and matters of competence, as represented by some grammatical theory. On this view, the standard approach to accounting for many apparent discrepancies between the two is to advert to features of the psychological mechanisms that implement and deploy grammatical knowledge in support of speech

behavior and language comprehension. A classic example of this approach appears in the account of doubly center-embedded sentences in Chomsky and Miller (1963). Here it is maintained that the sentences are grammatical despite their manifest unacceptability on the grounds that their unacceptability arises from characteristic limitations of the memory structures used by the parser. Any discrepancy that can be analyzed in these terms is properly seen as of little or no relevance to grammatical theory. For example, in so far as there might be evidence of informal heuristic mechanisms playing some role in sentence comprehension, this is seen as arising in some way within the mechanisms that implement the grammar.

Another much less widely discussed view situates the linguistic system as but one of several cognitive resources that might be engaged in the course of language comprehension. On this view, the presentation of an utterance will typically engage several discriminable mental competencies and the utterance's effect on the listener might best be viewed as a negotiated result that integrates effects arising from two or more of the participating systems. On this view there are several competing and collaborating kinds of competence and an associated performance theory for each. For example, there are clearly extralinguistic phenomena of deixis. If the competence theory of general deixis turns out to subsume the theory of deixis in language, then the deictic system could be seen as another competency, closely linked to, but also partly independent of, the linguistic system. A similar account might be given of a mechanism that exploits pragmatic knowledge to infer what roles a given set of nouns might play relative to some verb. In that humans clearly have some ability to discern possible sentential interpretations where lexical material is presented with few or no syntactic cues, this too might be seen as another associated but independent competency.

Part of the importance of this second view is that it can lead to a quite different treatment of apparent discrepancies between observed performance and grammatical theory. It becomes possible in this context to ask whether an utterance that is, say, ruled ungrammatical by the grammar-based linguistic component, might nevertheless become acceptable through the intervention of some extralinguistic mechanism (see Bever 1974 for a discussion of some possible instances where ungrammatical forms may nevertheless be acceptable). On this view the grammatical implications of acceptability are more difficult to discern. The grammar ought to be held accountable only for those utterances whose acceptability does not arise through extra-grammatical means. Thus studies of performance interpreted in this frame have a potential to bear on linguistic theory somewhat more directly than can results interpreted in terms of the more common frame. Experimental observations of performance that argue that a given sentence type comes to be acceptable by way of the involvement of some extragrammatical competency allow grammatical theory to set aside certain sentence types that it would otherwise have to cope with.

6. Experiment 4: *Which*-effects

The role of Experiment 4 is to examine one suggestion as to how an extragrammatical mechanism might be involved in some cases similar to those in Experiment 1. If the Indefinite cases of Experiment 1 are ungrammatical, this is presumably because grammatically based mechanisms for linking the wh-element and the gap are somehow impaired by the hierarchical configuration of the sentence. If such sentences were to have their interpretability and thus acceptability restored by some extragrammatical mechanism, it is apparently the filler-gap relation that this other mechanism must address. It does not seem far fetched to suggest that a relevant mechanism could be defined along the following lines: it would maintain only a flat (i.e., non-hierarchical) lexical representation of the utterance and would simply look for overt cues to fillers and gaps. On finding a filler and a gap, it would associate them in some way that would facilitate recovery of an analysis for the entire utterance. If this mechanism were to operate in an informal and heuristic fashion, its performance would likely improve with surface features that somehow made the elements of the filler-gap more salient or conspicuous. Thus, the essential idea of Experiment 4 is simply to manipulate the saliency of the wh-element in sentences like those used in Experiment 1 to determine whether more salient wh-elements are associated with higher acceptability.

6.2. Materials

The materials for Experiment 4 were in part similar to those of Experiment 1. Pairs of sentences of the Indefinite and Specified Subject types of Experiment 1 were matched to other pairs that were identical in every respect except that in the second pair the **who** or **what** was replaced by a **which** phrase. The **which** phrase was identical for the two members of the pair. A sample set of items appears in Table V. The expectation is that if the **which** phrase makes the filler more conspicuous, acceptability will improve in the cases including this structure.

What Cases
What did Sue resent a comment about?
What did Sue resent Tom's comment about?
Which Cases
Which of the new pledges did Sue resent a comment about?
Which of the new pledges did Sue resent Tom's comment about?

Table V: Sample Materials for Experiment 4.

6.3. Results and Discussion

The results of Experiment 4 are displayed in Figure 4. The most important result is that there is no significant improvement in acceptability in the presence of the **which** phrases, $F(1,66)=2.18$, NS, though the experiment did replicate the

difference between the Indefinite and Specified Subject cases of Experiment 1, $F(1,66)=140.90$, $p<.001$.

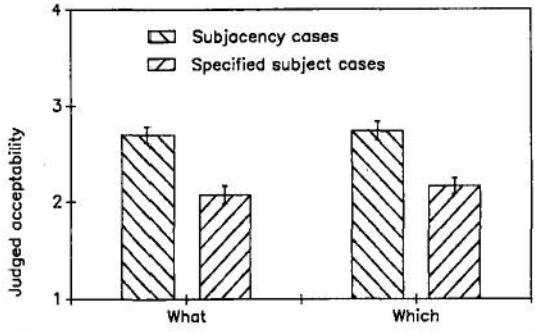

Figure 4: Mean judged acceptability for the four cases of Experiment 4.

These results of course do not support the suggestion that more salient wh-elements are associated with improved acceptability. Close informal examination of the by-sentence results of Experiment 4 suggests, however, a somewhat different picture. Though further experimentation will be required to assess the generality of these effects, it appears that there were numbers of specific materials sets within which the sentences with **which** were more acceptable. The aim of further investigation will be to identify factors that distinguish such sentences from others that did not show a **which** effect.

8. A multimodal view of language comprehension

The theoretical frame of this study is somewhat different than that conventionally assumed in linguistic and psycholinguistic work that touches on these issues. In the standard view, linguistic competence is imperfectly represented in language behavior due to a variety of phenomena that arise in the psychological and neurological mechanisms that implement the speaker/hearer's knowledge of language. Usually, such phenomena do not motivate changes in the theory of competence any more than the observation that people often make mathematical errors, even systematic ones, would motivate a change in the principles of mathematics. The underlying model of language comprehension might be termed unimodal; it assumes that all utterances are interpreted via the grammar and that all departures from grammatically defined norms must be accounted for by reference to the character and limitations of the mechanisms that apply the grammar.

There is, however, another potential source of discrepancies between the form of language behavior and the principles underlying that behavior, as captured by linguistic theory. It seems quite plausible, especially in language comprehension, that there may be more than one cognitive system that can participate. Thus, while there surely is a parser that implements a grammar, there may also be one or more other cognitive systems that can be involved in comprehension, either in collaboration with the parser/grammar or more independently in cases where the parser/grammar fails to provide an analysis. The availability of such a mechanism

is suggested by the seeming facility with which humans can sometimes cope with ill-formed utterances in the speech of immigrants, very young children, and individuals with anatomical or neurological impairments of speech.

Thus, an alternative view is available whenever some discrepancy between grammar and behavior is apparent. On this view, the discrepancy may arise because of some interaction between the core linguistic system and some other system, viewed as an alternative kind of competence, i.e., not as some manifestation of the mechanisms that implement grammatical knowledge. This might be termed a multimodal account of comprehension.

For any given problem, the matter can be put as a question: What is it about this particular discrepancy between grammar and behavior that shows that it is best attributed to implementational aspects of the linguistic system rather than to some interaction between the linguistic system and some other component of the cognitive system?

Where there is no compelling reason to take the implementational view, it seems necessary to consider whether the result in question may bear on linguistic theory, in particular, competence theory. That is, where discrepancies can be attributed to some other mechanism, then they need not be addressed by the theory of grammar. But where there are persistent conflicts between observed performance and what a grammar implies, and no convincing basis for attributing the difference to some other cognitive mechanism, it may be reasonable to hold the grammar responsible for those discrepancies and to consider whether it might be appropriate to modify the grammar to account for the obtained results.

As noted above, one of the arguments for a multimodal account of comprehension is that it offers some hope of accounting for the apparent resiliency of the language comprehension system. It seems plausible to suggest that a collaborating ensemble of mechanisms, each exploiting a different aspect of the information available in the utterance and its context, should, in general, be able to cope with anomalies in the input better than a more unified system. It is worth emphasizing, however, that any resiliency gained in this fashion is available only where the several mechanisms enjoy a high degree of autonomy. The more tightly the work of a given module is linked to that of another, the more vulnerable it will be to anomalies in the input that disrupt the work of the module on which it depends. It is just in so far as each module can make a useful contribution to the analysis of an utterance in the face of failures elsewhere that the multimodal model of language comprehension offers a better account of resiliency.

Acknowledgements

The work reported in this paper was supported in part by a grant from the National Institutes of Health (1 R01 NS22606-01). I am grateful to Tom Bever, Helen Cairns, Bob Fiengo, Bob Levine, Dana McDaniels, Craige Roberts, and

Michael Studdert-Kennedy for much wise counsel relevant to these matters. All responsibility for errors of fact or judgment that may be found here of course remains entirely with the author. Diana Smetters, Ken Williams, Rick Neighbarger, Andy Gudgel, and Grace Kuo have provided invaluable assistance with various aspects of the experimental work reported here.

References

Bever, T. (1974) The ascent of the specious, or, There's a lot we don't know about mirrors. In Cohen, D. (Ed.) *Explaining Linguistic Phenomena.* Washington, D.C.: Hemisphere Publishing. Pp. 173-200.

Chomsky, N. (1973) Conditions on transformations. In S.R. Anderson and P. Kiparsky (Eds.) *A Festschrift for Morris Halle.* New York: Holt, Rinehart and Winston. Pp. 232-286.

_____. (1981) *Lectures on Government and Binding.* Dordrecht: Foris.

_____. (1986) *Barriers.* Cambridge: MIT Press.

Chomsky, N. and G. Miller. (1963) Introduction to the formal analysis of natural languages. In Luce, Bush and Galanter (Eds.), *Handbook of Mathematical Psychology*, Vol. 2, Ch. 11. New York: John Wiley & Sons.

Fiengo, R. (1987) Definiteness, specificity, and familiarity. *Linguistic Inquiry* **18:** 163-166.

Fiengo, R. and J. Higgenbotham. (1981) Opacity in NP. *Linguistic Analysis* **7:** 395-421.

Lasnik, H. and J. Uriagereka. (1988) *A Course in GB Syntax.* Cambridge, MA: MIT Press.

Riemsdijk, H. and E. Williams. (1986) *Introduction to the Theory of Grammar.* Cambridge, MA: MIT Press.