

ARE THERE PREFERENCE TRADE-OFFS IN ATTACHMENT DECISIONS?

Lenhart K. Schubert

Department of Computing Science
University of Alberta, Edmonton

Abstract. The paper argues for an affirmative answer to the question, against the view that correct attachment decisions can be made by a serial process that considers alternatives in some order and accepts the first "satisfactory" alternative. The pitfall in serial strategies is that they are apt to finalize their choice while "the best is yet to come".

1. Background

Given the increasingly comprehensive *competence* frameworks for grammar developed within linguistics in recent years, computational linguists have been able to formulate increasingly specific *performance* theories for human (and machine) parsing. In particular, there has been a growing interest in the theory of lexical disambiguation and phrase attachment (e.g., see Frazier & Fodor 1978, Wanner 1980, Marcus 1980, Ford et al. 1981, Shieber 1983, Hirst 1984, Schubert 1984, Wilks et al. 1985). These studies are motivated in part by an interest in psycholinguistics, and in part by a desire to construct practical parsers which emulate human choice behaviour, producing only "preferred" analyses of sentences rather than all possible analyses. As well, such studies feed back into the grammatical frameworks within which they are conceived, confirming or disconfirming those frameworks to the extent that they make it easy or hard to embed convincing attachment theories within them.

The point of departure for most of the studies has been Kimball's principles, especially Right Association (RA) and Minimal Attachment (MA) (Kimball 1973, Frazier & Fodor 1978). Both principles are purely syntactic. RA states that a newly postulated phrase is attached as low in the tree structure to its left as possible. This explains, for example, why the prepositional phrase (PP) for Mary in

(1) John bought the book which I had selected for Mary

is understood as modifying selected, rather than bought. MA (as strengthened by Frazier & Fodor, 1978) states that a phrase is to be attached into the tree structure to its left using the smallest possible number of additional nonterminal nodes. Thus in

(2) John carried the groceries for Mary

the PP for Mary is attached to the VP headed by carries rather than to the NP the groceries, since (it is claimed) VP's can accommodate a PP directly via rule $VP \rightarrow V NP PP$, while NP's can accommodate a postmodifying PP only via a rule that creates an additional NP node, $NP \rightarrow NP PP$. This accounts for the fact that most readers interpret the PP in (2) as modifying carried, even though RA would appear to favour attachment to groceries.

Another principle which has played an important role in recent discussions of attachment priorities is Lexical Preference (LP)

(Ford, Bresnan & Kaplan 1981). In essence, LP says that lexical verbs and other lexical items may prefer one pattern of complementation to another. For example, the verb want is said to prefer the pattern of complementation V NP to the longer pattern V NP PP, while the verb position has the opposite preference, and this accounts for the contrast between

(3) Mary wants the dress on that rack
and

(4) Mary positioned the dress on that rack

(Note that in (3) LP must be assumed to override MA to account for preferred attachment of the PP to the dress.)

There is an older, more "semantic" version of LP due to Wilks (1975a). According to this version, particular senses of lexical verbs (or other items) prefer certain complements to others, not because of the syntactic features of those complements, but because of their semantic categories; i.e., the preferences correspond to *selectional restrictions*. (Another class of semantic preferences is associated with certain words, chiefly prepositions - see below.) This notion was at the heart of Wilks' theory of Preference Semantics, according to which sentences are interpreted in such a way as to maximize the density of preferences satisfied (Wilks 1975a). Wilks' ideas did not find their way into the above theories of attachment, since those theories were concerned for the most part with attachment preferences in "semantically neutral" contexts.

2. Preference Trade-offs

In Schubert (1984) (henceforth Sch84) syntactic theories of attachment were criticised on several grounds:

- (i) They often depend on ill-specified or implausible principles of parser operation.
- (ii) They often depend on questionable assumptions about syntax.
- (iii) They lack provision for integration with semantic/pragmatic preference principles.
- (iv) They admit counterexamples even when (i)-(iii) are discounted.

An alternative approach was sketched, involving numerically weighted preferences and allowing trade-offs among syntactic and semantic/pragmatic preferences. Syntactic preferences were to be captured by the following two principles.

- (a) A graded distance effect: immediate constituents of a phrase prefer to be close to the *head lexeme* of the phrase. The effect is mediated by an "expectation potential" which decreases with distance from the head lexeme and increases with constituent size; as a result, larger constituents admit larger displacements from the head lexeme;¹

¹ The exact form of the distance effect is still somewhat

- (b) A rule habituation effect: there is an inhibitory potential or "cost" associated with each phrase structure rule (including lexical rules), leading to a preference for low-cost rules over high-cost rules. (e.g., ADJ->fat is preferred to N->fat).

An additional effect suggested in the verbal presentation of the paper was the following.

- (c) Inhibition by errors: "mild errors" such as concord errors contribute inhibitory potentials to the phrases in which they occur.

Note that (a), although nominally a distance effect, also accounts for RA. Note further that (b), though primarily intended to account for (syntactic) lexical preferences, can also lead to MA-tendencies, if all rules have non-zero cost. (However, I regard it as an open question whether there are any MA-tendencies to be accounted for.)

Like syntactic effects, semantic and pragmatic effects are assumed to influence attachment choices through potentials contributed to phrases, called semantic potentials. The following are some possible principles governing semantic potentials (somewhat elaborating the very sketchy proposals in Sch84).

- (d) Salience in context: the potential of a word sense or phrase is high to the extent that the denotation of that word sense or phrase is salient in the current context.
 (e) Familiarity of logical-form pattern: the potential of a phrase is high to the extent that its logical translation instantiates a familiar pattern of function-argument combination.
 (f) Conformity with scripts/frames: the potential of a phrase is high to the extent that it describes a familiar kind of object or situation (such as might be specified in a script or frame).

Principle (d) is intended to allow for "semantic priming by spreading activation", which is the postulated basis for the contrast between the following sentences:

- (5) The Hollywood producer married the star
 (6) The astronomer married the star

In addition, the principle permits implementation of the idea that the parser prefers phrases interpretable as references to previously introduced entities to phrases that introduce new entities into the discourse context (cf. Crain & Steedman 1981).

Examples of the sorts of patterns subsumed under (e) might be

KICK(THE-BUCKET)
 predicate-of-locomotion(temporal-term)

The first pattern is assumed to match such predicate-argument combinations as

kick'(<the' bucket'>)
 kick'(<the' (old'(bucket))>),

while the second pattern is assumed to match such predicate-argument combinations as

flies'(KIND (time'))
 creep by'(<the' minutes'>).

The latter two formulas indicate in an approximate way the logical translations of the sentences "Time flies" and "The minutes creep by", respectively (see Schubert & Pelletier 1982, Pelletier & Schubert 1984). The claim implicit in (e) is that such idiomatic and quasi-idiomatic patterns induce preferences

for the phrases whose translations they match. (This, I would claim, is part of the reason why "Time flies like an arrow" is not normally perceived as ambiguous.)

Principle (f) accounts for the contrasting PP-attachment preferences in

- (7) John saw the bird with the yellow feathers
 and
 (8) John saw the bird with the binoculars

The idea is that "feathers" matches a slot in a "bird" frame, so that "bird with feathers" is recognized as a familiar combination. Similarly "binoculars" matches a viewing-instrument slot in a "viewing" frame activated by "see", so that "seeing with binoculars" is recognized as a familiar combination. More subtle frames, capturing stereotyped social situations, would be needed to account for the PP-attachment preferences in

- (9) John met the girl he married at a dance
 and
 (10) John married the girl he met at a dance.

The "potentials" contributed by (a)-(f) are cumulative, i.e., they are transmitted upward in phrase structure trees, adding to (or, in the case of inhibitory potentials, subtracting from) the overall potential of superordinate phrasal nodes. The parser is assumed to operate left-to-right, maintaining a set of (overlapping and in general incomplete) phrase structure trees, each of which completely covers the material seen so far. Whenever there are three complete, ambiguous phrases within these trees, the tree with the lowest global potential is discarded. (A phrase is ambiguous if it has more than one parent within the set of overlapping trees.) The three-phrase limit, as in Marcus' PARSIFAL (Marcus 1980), accounts for garden-path phenomena.

How do Wilks' semantically based "lexical preferences" fit into this scheme? I think they correspond roughly to preferences of category (e), i.e., to preferred patterns of function-argument combination. However, as was pointed out in Sch84, preferred function-argument patterns may violate selectional restrictions. For example, the predication flies'(KIND(time')) generates a positive type-(e) potential even though "flies" (and similarly "creeps by", etc.) presumably selects for a moveable physical object as argument.

Apart from the resemblance between type-(e) potentials and Wilks' semantic preferences, the preference for structures with high global potential is not unlike the preference for "semantically dense" structures in Preference Semantics. The more recent theory of Wilks et al. (1985) that I will focus on, however, uses a serial decision making process rather than parallel evaluation of alternatives.

3. The Puzzling Case of the Partisan Informants

Wilks et al. (1985) (henceforth WHF85) concur with the criticisms of the syntactic preference principles in Sch84, but go on to reject both the preference trade-off approach and a rather different synthesis by Hirst (1984). They describe what they believe to be a simpler and more powerful "semantics-based" approach. Their proposal merits close examination, in view of the claims made for it. I will first discuss their criticisms of the preference trade-off approach and then apply the critical perspective of Sch84 to their own proposal (in Section 4).

Among the sentences in Sch84 were the following. (The fractions shown after them will be explained shortly.)

- (11) Mary saw the man who had lived with her while
 on maternity leave

¹(cont'd) open. In the verbal presentation of Sch84, the effect was formulated in terms of an inhibitory "delay potential" directly proportional to distance (in words) and inversely proportional to constituent size (in words).

- (12) John met the tall, slim, auburn-haired girl that he
married at a dance 17/29
(13) John was named after his twin sister 20/29

These sentences were intended to illustrate that syntactic preferences can cause a semantically less coherent alternative to prevail over a more coherent one. My informants did indeed find the sentences confusing - at least momentarily. The authors of WHF85, however, insist that their informants were not confused by (11) or (12), and that the ambiguity they perceived in (13) has a non-syntactic explanation. This, they believe, clears the way for their putatively non-syntactic theory.

These opposing claims are puzzling. Whose informants are to be trusted? The puzzle is not hard to unravel, however. By "prevail over a more coherent alternative", I did not necessarily mean "prevail irrevocably", but only long enough to cause a momentary sense of anomaly or confusion. Often, the informants experienced a double-take, something that might be termed a "huh? oh!" experience. This suggests that for these sentences normal parser operation leads to an anomalous interpretation, prompting re-analysis. To be sure of the data, I have re-tested (11)-(13) more formally and extensively, also including the sentences

- (14) Mary moved in with her uncle in New York who
had fallen ill while on maternity leave 25/29

(15) John said that he will definitely leave yesterday 20/29
The fractions alongside (11)-(15) show the proportion of subjects who reported initially arriving at an anomalous reading. In most cases re-analysis led them to a sensible analysis, but some subjects did not recover; e.g., 6 in the case of (12), and 8 in the case of (15). (A sample test form is reproduced in the Appendix.) The results attest to the reality of the phenomenon at issue.²

The informants used in WHF85 presumably were asked only about their *eventual* interpretation of (11)-(13). That they were able to recover from any initial confusion they may have experienced is entirely consistent with my data, at least if their number was small. One point in the discussion of (11)-(13) in WHF85 calls for comment, however, namely the alleged role of "information vacuity" in (12) and (13). The authors say that their informants resolutely attach the *at-PP* in (12) to met; they still do so when dance is replaced by wedding, even though, the authors assert, this requires them to discount the information vacuity of married at a wedding (or Grice's maxim of quantity). But an informant who rejects a vacuous combination is not discounting Grice's maxim; on the contrary, he is assuming that the speaker conformed with it! The authors' assertion that in (13), information vacuity tells against the interpretation of named after .. as named later than .. is equally groundless. The proposition that John was given his name later than his twin sister was given hers is informative, and that is precisely why I chose (13) instead of John was named after his father (from Wilks 1973). The point is that the perfectly sensible named later than .. reading is blocked (at least temporarily) by a powerful lexical preference (which I am inclined to regard as syntactic, though this is perhaps a matter of terminology).

Wilks et al. may wish to argue that the "huh? oh!" experience provides no evidence for separate phases of parser operation,

² Similar examples have often been discussed in the literature. For example, the reality of the distance effect illustrated by (12) is already rather well documented (see Frazier & Fodor 1979, Ford et al. 1981).

one normal and the other involving re-analysis.³ It may simply indicate that in the normal course of parser operation, rejection of a semantically anomalous combination registers consciously. However, one would then expect at least one of (7), (8) to elicit the "huh? oh!" experience. Experimentally, it turned out that 1/37 subjects (3%) reported such an experience for (7) and 6/39 (15%) for (8). These fractions are rather low compared with those for (11)-(15), tending to disconfirm such an explanation.

The evidence, therefore, seems to favour my point about the potency of syntactic effects in certain instances. However, I would emphasize that the case for preference trade-offs certainly does not hinge on this point. I will now show that the serial, semantics-based strategies in WHF85 (which actually contain more syntax than the authors imply) suffer from some of the shortcomings that the trade-off strategy was designed to remedy.

4. Rules A and B

The strategies in WHF85 are claimed to achieve wide coverage without "syntactic rules or complex syntactico-semantic weighting". The first strategy is the following.

Rule A: Moving leftwards from the right hand end of a sentence, attach (word or phrase) X to the first entity to the left of X that has a preference that X satisfies. Assume also a pushdown stack for inserting such entities as X into until they satisfy some preference. Assume also some distance limit (to be empirically determined) and a default rule such that, if any X satisfies no preferences, it is attached locally, i.e., immediately to its left.

"Preferences" here mean verb and noun complement preferences, such as that want prefers a physical object as its object case and a human recipient as its recipient case, and that ticket prefers a place as its direction case (as in ticket to London).

Rule A is claimed to be intentionally naive, being stated only to demonstrate "the wide coverage of the data by a single semantics-based rule".

The first and most obvious observation about rule A is that while it is a single rule in the sense that *any* algorithm is a single rule, it makes tacit use of 3 separate principles:

- (i) RA, other things being equal
- (ii) N/V selectional preferences
- (iii) a distance limit for attachment

I can see no justification for the claim that this method dispenses with syntactic preferences - (i) and (iii) are patently syntactic - or that it achieves greater coverage with simpler means than, say (a)+(e) as stated earlier (or for that matter, RA+MA+LP, despite the difficulties enumerated in Sch84). My (a) covers (i) and (iii) (and in addition allows for the size of the constituent being attached), and (e) is comparable to (ii). Note that (b), (c), (d), (f) are concerned with phenomena (such as lexical disambiguation) not covered by rule A.

Moreover, the method is subject to some of the same criticisms as the methods discussed in Sch84. It depends on ill-specified principles of parser operation, lacks provision for integration with preferences determined by context and world knowledge, and is susceptible to classes of counterexamples even within the limits of its intended coverage.

³ That there *is* a separate recovery mechanism is widely conjectured; e.g., Wilks (1975b:132), Milne (1982).

Concerning the first point, rule A seemingly relies on a preprocessor that performs lexical disambiguation and packages a sentence into a sequence of disjoint phrases, leaving only certain phrase attachment decisions to be made. But how plausible is it that lexical disambiguation can be decoupled from phrasal disambiguation in general? One would expect phrasal combinations which are made possible by particular lexical choices to influence those very choices. One could conceivably feed all possible phrase sequences, generated by the various combinations of lexical choices, into rule A and select the best result; but it is an open question whether the resulting lexical and phrasal choices will match those of people, and whether the number of alternative phrase sequences that need to be parsed by rule A will be moderate (keeping in mind that before phrases are combined, one cannot in general tell whether a phrase sequence is grammatically possible).

Concerning the second point, it is not hard to make up contexts in which the PP associates with the noun in (2) or even in (8). Rule A is simply wrong in such cases, as is any rule that describes particular parser *choices* without regard for context. The advantage of the trade-off theory is that it lets us have our say about syntactic, lexical, or semantic preferences once and for all; allowing for context is a matter of *adding* something to the theory, not *revising* it. Much the same can be said about the role of world knowledge. The following are some sentence pairs in which shifts in attachment preferences are induced by subtle kinds of world knowledge.

- | | | |
|---------|---|-------|
| (16) a. | The women discussed the children in the kitchen | 23/29 |
| b. | The women discussed the dogs on the beach | 15/29 |
| (17) a. | John broke the vase in the kitchen | 15/29 |
| b. | John broke the vase in the corner | 7/29 |
| (18) a. | Mary talked with a man on her front porch | 26/29 |
| b. | Mary talked with a man on a park bench | 13/29 |

As indicated by the fractions, the PP is more likely to be attached to the verb in the a-sentences than in the b-sentences. It is hard to see how a serial algorithm like rule A could be modified to allow for such effects, especially without recourse to weights of some kind.

Third, concerning the existence of classes of counterexamples, rule A is acknowledged in WHF85 to be ill-equipped for recognizing standard V/N+PP patterns, because selectional preferences do not in themselves indicate what prepositions may be used to introduce complements. This is remedied in rule B, but at a cost in complexity. More seriously, rule A fails for a-b pairs of the following sort:

- | | |
|---------|--|
| (19) a. | Joe lost the ticket to Paris |
| b. | Joe mailed the tickets to London |
| (20) a. | Mary hires men who have worked on farms as cowboys |
| b. | Mary describes men who have worked on farms as cowboys |

In order to account for (19)a, WHF85 posit a locative case for ticket, as already mentioned. But then, given the right-to-left strategy of rule A, the PP will be attached to tickets in (19)b as well. However, experiments show that readers of (19)b almost always attach the PP to the verb (28/29), in most cases without awareness of the alternative (24/28). Similarly in (20)a, to explain the tendency to attach the final PP to worked (22/29), we need to assume a corresponding preference; but this assumption leads rule A to predict attachment to worked in (20)b as well, whereas experimentally most readers attach to describes (18/29). Thus pairs like (19) and (20) show that rule A is apt to make premature attachment decisions while "the best is yet to

come". Such pairs are unproblematic for the trade-off theory, since in that theory different verbs or nouns are allowed to "compete" in parallel for a postmodifier.

Let us now turn to the more subtle rule B. One refinement is that lexical entries for verbs and nouns list the prepositions which may introduce their complements. As well, lexical entries for prepositions list the patterns of verb or noun postmodification that these prepositions prefer to participate in. For example, one of the patterns (called preplates) for on is [*do-dynamic, loc-static, point, on4] which is satisfied by a phrase like position on the rack.

Rule B is intended only for attachment of PPs immediately following an object NP, and works roughly as follows.

1. Try to attach the PP as in rule A, minus the default rule and distance effect. If this fails move upward to the next sentential level and restart.
2. (Still no attachment at top sentential level) Attempt PP attachment using preferences of the preposition (preplates), starting at the main verb of the sentence and working rightward.
3. (Default) Try to attach the PP to the verb using "relaxed" versions of the preplates. Similarly try to attach to the object NP.

For example, step 2 is decisive in the sentences

- | | |
|---------|-----------------------------------|
| (21) a. | John stabbed the girl in the park |
| b. | John loved the girl in the park |

where, according to WHF85, stabbed in the park satisfies a preplate of in, allowing attachment to the verb in (21)a. In b, loved in the park fails to satisfy any preplate of in, so that attachment to girl is tried next, and this does satisfy a preplate of in.

Does rule B escape the criticisms of rule A? Evidently not. The claim that it dispenses with syntactic information is still unwarranted; like A it has *proceduralized* the syntactic preferences, not eliminated them (any more than Shieber's Shifting Preference has done so, for example). Essentially the implicit principles are

- (i) "Strong" (N/V-based) and "weak" (P-based) semantic preferences
- (ii) RA (or low attachment) of the unattached PP to constituents with a "strong" preference. (This takes precedence over (iii) when both apply.)
- (iii) High attachment to constituents with a "weak" preference for the unattached PP (cf. Ford et al.'s Invoked Attachment)

Step 1 of rule B corresponds to (ii), and steps 2 and 3 to (iii). (i) is a semantic principle but (ii) and (iii) are syntactic (once semantic preferences have been assigned in accordance with (i)).

As with rule A, the exact role of rule B within the parser remains unclear, particularly the interaction with lexical disambiguation. The rule is still incorrect when context or world knowledge come into play, and it is still susceptible to the kinds of counterexamples noted for rule A.

Further, it should be noted that the intended coverage of rule B is much more limited than the intended coverage of the preference trade-off account, making complexity comparisons somewhat irrelevant. Some of the phenomena not covered by rule B are attachment of PPs following a sequence of VPs, attachment of adverbs, participles, infinitives, relative and subordinate clauses, noun premodification, lexical ambiguity, garden path phenomena, the effect of concord errors, and distance effects.

I think it quite unlikely that a serial approach like rule B can be expanded so that, functioning as part of a parser, it will account correctly for all such phenomena, and in addition allow for the effects of context and world knowledge. The great advantage of the trade-off approach is that it allows various preference principles to be formulated more or less independently, adding their effect through potentials to a parser whose computational structure is fixed. Moreover, the ideas contained in the original version of Preference Semantics have a rather natural place within such a scheme.

Finally, an ever-present phenomenon in sentence comprehension tests such as those I have cited is extensive individual variation. This seems rather easy to account for within a model based on competing preferences of various strengths, and hard to explain in a serial model whose behaviour can be modified only by adding or deleting patterns to which the parser is sensitive. Is it not more natural to assume that there are individual differences in the *degree* of sensitivity to various syntactico-semantic patterns?

Acknowledgements

I am grateful to Jeff Pelletier of the Philosophy and Computing Science Departments for conducting the experiments reported herein. I have also benefited from discussions with him and other members of the Logical Grammar Study Group, especially Matthew Dryer, who suggested some of the example sentences. The research was supported by NSERC Operating Grant A8818.

References

- Crain, S. & Steedman, M. (1981). The use of context by the Psychological Parser. Paper presented at the Symposium on Modelling Human Parsing Strategies, Center for Cognitive Science, Univ. of Texas, Austin.
- Ford, M., Bresnan, J. & Kaplan, R. (1981). A competence-based theory of syntactic closure. In Bresnan, J. (ed.), *The Mental Representation of Grammatical Relations*, MIT Press, Cambridge, MA.
- Frazier, L. & Fodor, J. (1979). The Sausage Machine: a new two-stage parsing model. *Cognition* 6, 291-325.
- Hirst, G. (1983). Semantic interpretation against ambiguity. Tech. Rep. CS-83-25, Dept. of Computer Science, Brown Univ., RI.
- Hirst, G. (1984). A semantic process for syntactic disambiguation. Proc. AAAI-84, Austin, TX, 148-152.
- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition* 2, 15-47.
- Marcus, M. (1980). *A Theory of Syntactic Recognition for Natural Language*, MIT Press, Cambridge, MA.
- Milne, R. (1982). Predicting garden path sentences. *Cognitive Science* 6, 349-373.
- Pelletier, F. J. & Schubert, L. K. (1984). Two theories for computing the logical form of mass expressions. Proc. COLING-84, July 2-6, Stanford, CA, 247-250.
- Schubert, L. K. (1984). On parsing preferences. Proc. COLING-84, Stanford, CA, 247-50.
- Schubert, L. K. & Pelletier, F. J. (1982). From English to logic: context-free computation of 'conventional' logical translations. *Am. J. of Computational Linguistics* 8, 26-44.
- Shieber, S. M. (1983). Sentence disambiguation by a shift-reduce parsing technique. Proc. IJCAI-83, Aug. 8-12, Karlsruhe, W. Germany, 699-703. Also in Proc. of the 21st Ann. Meet. of the Assoc. for Computational Linguistics, June 15-17, Cambridge, MA., 113-118.

- Wanner, E. (1980). The ATN and the sausage machine: which one is baloney? *Cognition* 8, 209-225.
- Wilks, Y. A. (1973). Understanding without proofs. Proc. IJCAI-73, Stanford, CA, 270-277.
- Wilks, Y. A. (1975a). A preferential pattern-seeking semantics for natural language inference. *Artificial Intelligence* 6, 53-74.
- Wilks, Y. A. (1975b). Methodology in AI and natural language understanding. Proc. Workshop on Theoretical Issues in Natural Language Processing (TINLAP 1975), Cambridge, MA, 130-3.
- Wilks, Y. A., Huang, X., & Fass, D. (1985). Syntax, preference, and right attachment. Proc. IJCAI-85, Aug. 18-23, Los Angeles, CA, 779-784.

APPENDIX: Example of Linguistic Quiz

Your cooperation is requested in completing the following simple psycholinguistic quiz. This is an anonymous quiz -DO NOT PUT YOUR NAME ON IT.

The purpose of the quiz is to gain some insights into the processes of syntactic analysis and interpretation of sentences. There is a test sentence, which is followed by a question. PLEASE READ THE SENTENCE AT NORMAL SPEED, AND THEN IMMEDIATELY ANSWER THE QUESTION THAT FOLLOWS IT AS HONESTLY AS YOU CAN.

Test sentence: John saw the bird with the yellow feathers.

Question: Which of the following statements best describes your impressions upon reading this sentence? (Indicate your choice with an "X"):

- You took "the yellow feathers" as referring to a viewing instrument used by John, rather than as part of the bird (however odd this may have seemed).
- You eventually took "the yellow feathers" as referring to part of the bird; however, you initially took it as referring to a viewing instrument used by John, and were prompted to reanalyze the sentence because of the oddity of that interpretation.
- You became conscious that "the yellow feathers" could in principle refer to either part of the bird or a viewing instrument used by John, and you chose the former (more plausible) interpretation without any sense of correcting an initial misunderstanding.
- You took "the yellow feathers" as referring to part of the bird without becoming conscious of another interpretation.
- Other (explain):