Language and Cognitive Processes

Comprehension of Sluiced Sentences

Lyn Frazier Charles Clifton
Published online: 21 Sep 2010.

PLEASE SCROLL DOWN FOR ARTICLE

be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Comprehension of Sluiced Sentences

Lyn Frazier and Charles Clifton, Jr

University of Massachusetts, Amherst, MA, USA

We report two reading experiments and two questionnaire studies designed to investigate the processing of "sluiced" sentences, like *Somebody left—guess who.* A self-paced reading experiment showed that sentences with explicit (overt) antecedents are read more quickly than sentences with implicit (covert) antecedents, both when the antecedents in question were arguments and when they were adjuncts. An eye movement experiment showed that sluiced sentences containing two potential antecedents were read faster than sentences containing only a single antecedent in matrix subject position. We suggest this is because only the ambiguous sentences contained an antecedent in a normal focus position (embedded object position). Two questionnaire studies suggested that perceivers prefer a focused constituent as the antecedent of the sluiced constituent. Since we argue that the interpretation of a sluiced constituent take place at the representational level of "logical form" (LF), we conclude that implicit arguments are not made explicit at LF but that focus is important in the processing of LF.

INTRODUCTION

Psycholinguistic investigation of ellipsis has focused primarily on the processing of verb phrase (VP) anaphors, contrasting the behaviour of "deep anaphors" (Hankamer & Sag, 1976), such as *did it too*, with surface anaphors, such as *did too* (Carlson & Tanenhaus, 1988; Malt, 1985; Tanenhaus & Carlson, 1990; Ward, Sproat, & McKoon, 1991). Most of these studies have concentrated on effects of the form of the antecedent and where it appears in a preceding discourse (cf. Shapiro & Hestvik, 1995, discussed below, for one exception).

We propose to extend the domain of psycholinguistic studies to include the processing of a kind of ellipsis called "sluicing" (Ross, 1967). Interpreting a "sluiced" constituent, such as the *who*-clause (a "complement phrase", or CP) in (1) requires an inference (where we use the term *infer* to mean adding material that is not explicitly present in the input string).

1a. Somebody left—guess [$_{CP}$ who].
1b. Joan likes someone but I don't know [$_{CP}$ who].

The entire CP apart from *who* is missing in these sentences, and must be filled in by the reader. This filled-in material is in fact copied from the initial portion of the sentence, in a way to be made explicit shortly. There are clear linguistic constraints on what can be copied or filled in. For instance, (2a) is an acceptable example of sluicing, while (2b) is not. This suggests that the material that is filled in or inferred by the reader must somehow incorporate an indefinite noun phrase (NP) (*some girl*) rather than a definite NP.

2a. John likes some girl but I don't know who.
2b. John likes the girl but I don't know who.

Our research focuses on two questions about the inference or copying process underlying the interpretation of sluiced sentences. The first, addressed in Experiment 2, is how does the reader or listener choose among multiple alternatives that could in principle be used as the sluiced material? The second, addressed in Experiment 1, is when the sluiced material itself contains elements that are not overtly present but must be inferred by the reader, how and when is that inferencing done? We will argue that inferring the implicit element in the sluiced material must be done as part of the task of copying the sluiced material from the initial portion of the sentence, which we will argue must be done as part of the process of constructing a particular level of linguistic representation, "logical form" (LF). We will contrast the inferences required to interpret a sluiced sentence with thematic role inferences that Tanenhaus and his colleagues (Carlson & Tanenhaus, 1988; Mauner, Tanenhaus, & Carlson, 1995) have studied. Tanenhaus and his colleagues claim that thematic role inferences are present in a discourse representation. To anticipate our conclusions, we will accept their proposal but also claim that thematic role inferences are not present in the grammatical representation that is required to interpret a sluice.

To make arguments of this degree of precision, we must propose an explicit analysis of just what we think is inferred in sluicing. We describe a

recent analysis by Chung, Ladusaw and McCloskey (1995) that seems to us to be the most adequate existing grammatical account of sluicing. Once we have explicitly described what we claim is inferred, we will return to a discussion of the process by which it might be inferred.

Grammatical Analysis of Sluicing.   Chung et al. (1995) argue that constraints on sluicing apply at Logical Form (LF). Logical form is a structural representation which is abstracted from surface structure and which captures certain semantic properties of a sentence such as the scope of quantifiers (e.g. *every*, *some*, *many*, etc.). The empty structure for the elided constituent (the CP in (1)) is added at the LF representation, not at any surface syntactic representation or at a discourse structure representation. Chung and co-workers' basic proposal is that "if a legal LF can be constructed for the sluice, then the sluice *has* an interpretation, one similar in all relevant respects to the interpretation of the corresponding constituent question" (p. 243).

Chung et al. argue that sluicing is permissible whenever the sluiced constituent (*who* in (1)) can be used to create an LF similar to the LF of the antecedent constituent (*somebody left*). The LF that is created must follow all constraints on LF, including the fact that an operator such as the wh-expression *who* must be in the same domain as the variable it binds. Thus, the antecedent constitutent must supply a free variable, such as an unbound indefinite NP, for the sluice's wh-operator (the *who* in (1)) to bind. This is illustrated in (3) and (4). In (3), the wh-constituent is in [Spec,CP] of the LF of the sluiced constituent, *guess who*. As an operator, it must bind a variable in the inflection phrase (IP) of this constituent. But this IP is empty. The content of this empty IP is, according to Chung et al., determined by copying the articulated empty structure from the antecedent IP, namely the structure dominating *somebody left* in (3).

3.
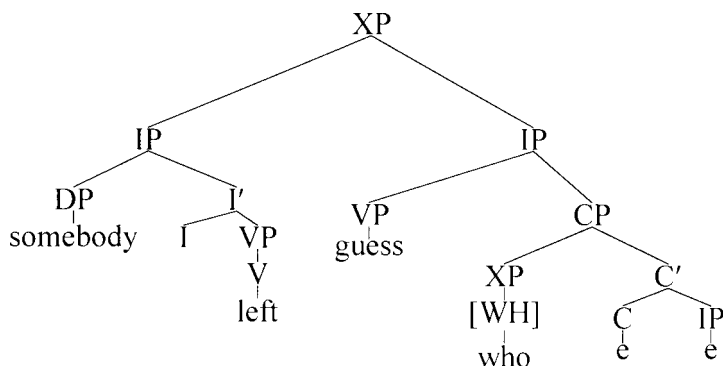
4.
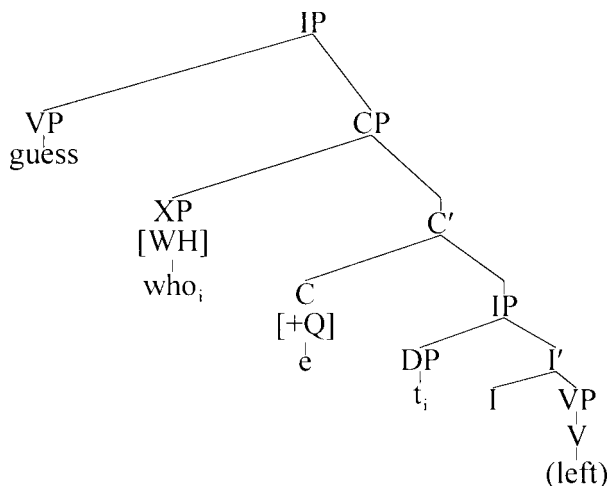
Example (4) shows the result of this copying. In (4), a trace has been inserted as the subject of the IP to provide the variable for the win-constituent (*who*) to bind. This trace is inserted in the LF representation in the position corresponding to a free indefinite (one that is not bound by another operator) in the antecedent IP—*somebody* in (1). The free indefinite in the antecedent IP (somebody) is called the "inner antecedent". Any restrictions on the inner antecedent are inherited by the semantic variable of the trace through an operation of "merger". For example, in *Some girl left—guess which one*, the possible values for the trace variable would be restricted to girls.

In Example (4), the antecedent IP whose LF is copied into the sluiced constituent contained an explicit inner antecedent (*someone*). In (5), however, the antecedent IP *Ella typed her assignment* doesn't contain a free indefinite to serve as the inner antecedent.

5. Ella typed her assignment but I don't know with what.

When no inner antecedent is present, a variable may simply be "sprouted" (added to the LF representation) in a licensed syntactic position. A PP must be added or sprouted in the empty IP to supply a variable for the wh-constituent to bind, as illustrated in (6), where the final PP does not correspond to anything in the initial part of the sentence in (5). It is this operation of sprouting which we investigate in Experiment 1.

6.
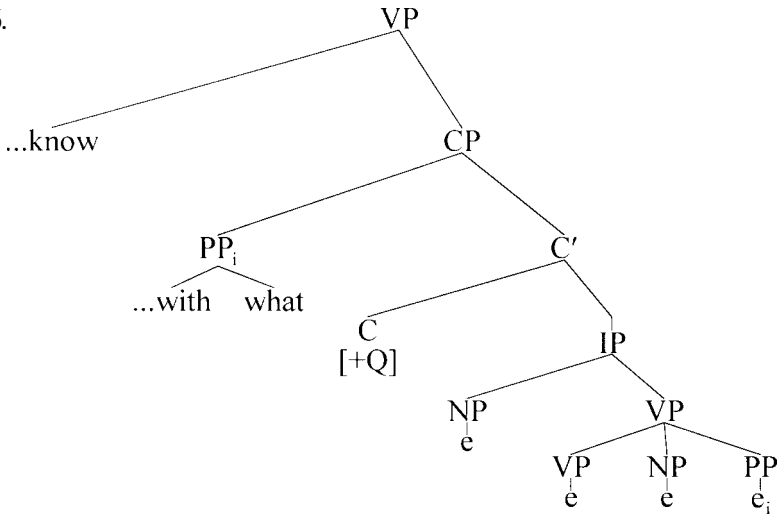
To summarise, we have described two mechanisms for constructing a legal LF for a sluiced constituent. The first copies or recycles an antecedent IP with articulated empty structure which is available in the discourse. The second identifies a variable in the recycled IP that corresponds to a free indefinite in the antecedent, and binds this variable to a wh-operator in the sluice. If the antecedent IP does not contain an appropriate free indefinite (or "inner antecedent,"), a variable is "sprouted" in a licensed syntactic position in the recycled IP of the sluice. It is critical to note that both mechanisms operate at the level of LF and obey the constraints on legal LF.

Inferring the Sluiced Material.   The first set of questions we address concerns the process of inferring, or "sprouting", an inner antecedent. If no explicit inner antecedent exists, as in (5) or in (7b,d), one might be able to determine when the inner antecedent is inferred by observing when increases in processing time occur, compared to cases like (7a,c) where the inner antecedent is explicit. Presumably, the operation of inferring a non-existing constituent will cost some processing time.

    7a. The secretary typed something but I don't know what.
    7b. The secretary typed but I don't know what.
    7c. The secretary typed somewhere but I don't know where.
    7d. The secretary typed but I don't know where.

One existing analysis of inference in sentence comprehension suggests that (7b), where the missing inner antecedent is an argument of the verb, may behave differently from (7d), where it is an adjunct. Carlson and

Tanenhaus (1988) and Mauner et al. (1995) propose that implicit arguments, or their theta-roles, are immediately incorporated into some linguistic representation—presumably a discourse representation—whenever a theta-assigner is encountered. In the original study showing the effects of implicit arguments, Tanenhaus, Carlson and Hudson (reported in Carlson & Tanenhaus, 1988) examined the time taken to read sentences like (8) to show that a phrase *the suitcases* could be integrated into discourse rapidly if its role in the event described had already been introduced as an implicit argument (e.g. in *John unloaded the car* the theme is an implicit argument). For example, the continuation sentence in (8b) is read faster following (8a) than following (8a′):

8a. John unloaded the car.
8a′. John hurried to the airplane.
8b. The suitcases were very heavy.

Later Mauner et al. (1995) showed that the implicit agent in a short passive is immediately available to serve as the controller for the empty subject of a rationale clause. Thus, the rationale clause in (9a), which contains an implicit agent of *sink* as controller of the PRO that is presumed to be in the subject position of *collect*, was processed just as rapidly as the rationale clause in (9b), which has an overt agent to serve as controller of PRO:

9a. The ship was sunk to collect the insurance money.
9b. The ship was sunk by the captain to collect the insurance money.

In (7b), the direct object argument of *type* is implicit, not explicit. The claim of Mauner et al. seems to imply that, at some level of representation, a theme argument will be inferred. This inference could be made at (or before) the level of LF (as we are assuming that the operation of copying or inferring the sluiced material is done). If this is so, then the sluiced representation of (7b) will contain an inner antecedent, just as (7a) does. No extra processing time should be needed in (7b) compared to (7a) to sprout a variable for the *wh*-constituent of the sluiced constituent to bind.

The situation is quite different in (7c) and (7d). Here the inner antecedent is an adjunct, not an argument, of the matrix verb. As an adjunct, its relation to the verb *type* is not specific to that verb. Rather, it can modify any verb of a wide semantic class (roughly, action verbs) by specifying a time, or instrument, or manner, or, in this case, a location (etc.). In contrast, the relation of an argument (like the direct object of *type*, *something*) to the verb is specified by the verb. Since the presence of an adjunct is permitted by general rules of the language, not by being specified by the verb itself, the mechanism that Carlson and Tanenhaus (1988) and Mauner et al. (1995) propose for inferring implicit arguments does not apply to adjuncts. This mechanism emphasises how thematic roles are specified by verbs (or by the

event denoted by the verb), and how implicit role-fillers are inferred when no explicit role-filler is present. Since adjuncts generally do not fill thematic roles, the proposed mechanism does not apply to them.

We assume that the requirement to infer a variable when one does not exist in the representation of a sentence (e.g. the requirement to sprout a bound variable in the sluiced constituent) slows the time to read the sluiced constituent. Therefore, if an implicit argument is already inferred as the inner antecedent of a sluice but an implicit adjunct is not, then reading the sluiced constituent should be slowed in cases like (7d), with a missing adjunct, but not in cases like (7b), with a missing argument.

This proposal assumes that the level of representation at which Carlson and Tanenhaus (1988) and Mauner et al. (1995) observed their effects is the same as the level involved in sluicing, or that the same inferences are made at these two levels. The level of representation involved in sluicing is, we have argued, LF. The level of representation at which Tanenhaus and colleagues have demonstrated that implicit arguments are inferred is probably a discourse representation or discourse model (Mauner et al., 1995, p. 376). A discourse representation and LF are distinct representations. A discourse model incorporates material from outside the current sentence or utterance, including information about the structure of events and the identity of entities involved in the events, and its construction presumably depends on and follows the construction of an LF representation. Thus, if we find evidence that implicit arguments are inferred prior to interpreting a sluice, we will be able to claim that implicit arguments are inferred at LF. If our evidence indicates that neither implicit arguments nor implicit adjuncts are so inferred, we will combine our evidence with that of Carlson and Tanenhaus (1988) and Mauner et al. (1995) to conclude that implicit arguments are inferred in a discourse model but not at the level of LF that underlies a discourse model.

EXPERIMENT 1

Experiment 1 had people read sentences like those in (7), reproduced below as (10):

10a. The secretary typed something / but I / don't know what.
10b. The secretary typed / but I / don't know what.
10c. The secretary typed somewhere / but I / don't know where.
10d. The secretary typed / but I / don't know where.

Each sentence had a version in which the sluiced material contained an implicit or explicit argument (10a,b) and a version in which it contained an adjunct (10c,d). The phrases we used as adjuncts varied in how "pure" their adjunct status was, since the line between argument and adjunct is a matter

of some uncertainty (cf. Schütze & Gibson, 1996). Some phrases (like the one in 10) meet all the standard syntactic tests for adjunct status (e.g. they come after all arguments, they can be iterated, etc.; cf. Schütze & Gibson, 1996); others, like instrument phrases pass some tests for arguments but fail others. As we will document below, all our "adjuncts" contrasted strongly with arguments in the properties that should permit them to be inferred during reading, but we do acknowledge that they were not all the purest form of adjunct.

If an implicit argument or adjunct (as in (10b,d)) must be inferred (sprouted) when the portion of the sentence containing the sluice is read, then that portion of the sentence (*but I don't know where* in (10d)) should be read more slowly than when the argument or adjunct was present in the string. We will interpret the absence of any such difference in time as evidence that the implicit argument or adjunct had been inferred previously, when the initial portion of the sentence was read.

Methods

Sixteen experimental sentences were constructed along the lines illustrated in (10), with slash marks indicating the division into presentation regions. All sentences are reproduced in the Appendix. Each sentence began with an antecedent clause which contained an explicit argument (*something*) as complement of the verb in version (a), an explicit adjunct (*somewhere*) in version (c), and neither complement nor adjunct in versions (b) and (d). The sentence continued with a conjunction and the subject of the next clause as one region, with a third region containing the matrix verb of this clause if the final VP was in an embedded clause.[1] The final presentation region of the sentence consisted of the full (elided) simple VP containing the sluiced constituent. In sentence versions (a) and (b) the sluiced constituent required an argument, whereas in sentence versions (c) and (d) it required an adjunct. Orthogonal to this manipulation, in versions (a) and (c) the inner antecedent of the sluice (the source of the variable that must be bound by the wh-word of the sluice) was overt, whereas in versions (b) and (d) it was only implicit.

To assess how strongly the initial portion of our sentences supported the possibility of an implicit argument, we searched text from the Linguistic Data Consortium's *Wall Street Journal* corpus (Marcus, Santorini, & Marcinkiewicz, 1993) for instances of each verb (both in the tensed forms used in our materials and in their untensed forms). We identified the first occurrence (if any) of each form in each of 145 weekly *Wall Street Journal* files and scored it as being a transitive usage or an intransitive usage. Our verbs averaged 72% transitive usage. We also conducted a simple normative

---

[1]In one sentence, the argument or adjunct occurred in the second region.

study in which 15 undergraduate students at the University of Massachusetts completed a sentence-completion questionnaire. All subjects were presented with a single page listing all 16 subject + verb beginnings of our 16 experimental sentences (e.g. *The secretary typed*) and were asked to complete the sentence fragment in a "simple, natural, easy way". Seventy percent of all completions used an argument of the syntactic form we used in our argument sentences (i.e. a noun phrase direct object of the verb). Only 4% of all completions contained an adjunct of the form used in our adjunct sentences (e.g. *about NP* following *The poor soul sang* or *on NP* following *Charles checked*). We are satisfied that the initial portions of our experimental sentences would support the inference of an argument rather than our targeted adjunct, if any inference were made.

The 16 experimental sentences were supplemented by a total of 76 other filler sentences of a wide variety of constructions. Forty-four of the sentences (including half the experimental sentences) were followed by simple two-choice questions, as illustrated in the Appendix, to ensure comprehension accuracy.

Seventy-two University of Massachusetts undergraduates were tested in individual self-paced reading sessions lasting approximately 30 min. Each subject first saw a short practice list and then read all 92 sentences in the main experiment in an individually randomised order. Counterbalancing procedures were used to ensure that each sentence was tested equally often in each of its four versions and that each subject received an equal number of sentences in each version. Each sentence was presented in a region-by-region manner on a microcomputer-controlled screen using a moving window procedure (Just, Carpenter, & Woolley, 1982), which presented underscore marks to indicate where each successive region would appear. When the subject pulled a trigger with the right index finger, each region disappeared and the next appeared immediately. Reading times for each region were recorded, as were question-answering times and accuracy.

Results

The results are presented in Table 1, both in terms of reading times (msec) uncorrected for differences in length of the region and in terms of the deviation (in msec) from the time predicted by a linear regression equation on the basis of the length of each region (Ferreira & Clifton, 1986; Trueswell, Tanenhaus & Kello, 1993). The only differences of interest among the experimental conditions occurred in the final region, which contained the sliced material. These differences were quite straightforward: The final frames of sentences with overt antecedents were read faster than those with implicit antecedents: 931 *vs* 1075 msec [$F_1(1,71) = 25.12$, $F_2(1,15) = 22.29$, $P < 0.001$]; in terms of deviations from predicted times, -103.5 *vs* $+45.7$

TABLE 1
Mean Reading Times in Experiment 1

Condition	Region 1	Region 2	Region 3	Final Region
Uncorrected times (msec)				
Explicit argument	1406	728	643	853
Implicit argument	1086	746	635	997
Explicit adjunct	1646	787	617	1008
Implicit adjunct	1056	761	692	1152
Corrected times (deviation in msec)				
Explicit argument	+152	+6	−62	−127
Implicit argument	+109	+35	−72	+28
Explicit adjunct	+290	+60	−80	−79
Implicit adjunct	+98	+61	−1	+62

Note: Region 3 existed only for sentences with four regions ($n = 4$).

msec, $[F_1(1,71) = 28.37, F_2(1,15) = 29.74, P < 0.001]$. Sluiced arguments (explicit or implicit) appeared to be read faster than sluiced adjuncts (926 *vs* 1081 msec), but this difference disappeared in terms of the deviations from predicted times: −49.2 *vs* −8.5 msec $[F_1(1,71) = 1.79, F_2(1,15) = 1.91, P > 0.15]$ (in any event, no observed difference between argument and adjunct sluices could be unambiguously interpreted, since the sluiced constituents differed in length and lexical content). Counter to the predictions of the view that implicit arguments are represented at LF, there was no interaction between these two factors ($F < 1$ in all analyses). Question-answering accuracy was above 92% correct in all conditions.

As indicated in the Introduction to Experiment 1, not all our adjunct sentences used phrases that were prototypical adjuncts. Three (items 1, 3 and 4 in the Appendix) contained instrumental phrases as adjuncts and seven (items 6, 8, 10, 11, 12, 14 and 16) contained prepositional phrases that could be viewed as oblique objects of verbs. Phrases like these were essentially never provided as sentence completions in the questionnaire study described in the Methods section and, unlike direct objects, they fail some of the conventional syntactic tests for argumenthood (Schütze & Gibson, 1996), but they do have some characteristics of arguments. The remaining six items, however, were pure adjuncts on all criteria. Therefore, a *post-hoc* analysis of just these six items was conducted, and the mean reading times for the critical final regions (containing the sluice) are shown in Table 2. It can readily be seen that these items exhibit the same effects seen in all 16 items. Statistical analysis bore this out. An analysis of variance[2] of

[2]Because eliminating items resulted in each subject having no data in one of the four conditions, squads of four subjects, one in each counterbalancing condition, were treated as the sampling unit in the by-subjects analysis, resulting in 11 "hyper-subjects" in all.

TABLE 2
Mean Reading Times in Experiment 1: Final Region, Pure Adjunct
Items Only

Condition	Uncorrected Times (msec)	Corrected Times (dev. msec)
Explicit argument	856	−161
Implicit argument	1085	53
Explicit adjunct	979	−107
Implicit adjunct	1275	158

Note: Data from items 2, 5, 7, 9, 13 and 15 only in the Appendix.

unadjusted reading times indicated that items with overt antecedents were read faster than items with implicit antecedents: 918 *vs* 1180 msec [$F_1(1,10) = 21.10$, $P < 0.01$; $F_2(1,5) = 25.59$, $P < 0.01$]. Neither the effect of argument versus adjunct, nor the interaction of the two factors, reached significance ($F < 1$ for the interaction).

Discussion

The linguistic analysis of sluicing provided by Chung et al. (1995) claims that the structure of the LF of the initial clause of sentences like those we used must be "recycled" as the LF of the sluiced constituent. To be a well-formed LF structure, it must contain a variable that the *wh*-operator can bind. In the sentences we used, this variable must correspond to a free indefinite argument or adjunct of the initial clause (the inner antecedent). Since the sluiced constituent was read more slowly when the inner antecedent was implicit than when it was explicit, we conclude that the operation of inferring the existence of an implicit inner antecedent was costly in processing time. We propose that this operation corresponds to what Chung et al. identified as sprouting of an inner antecedent, claiming that such sprouting was a prerequisite for constructing a grammatically well-formed structure/ interpretation for sluiced sentences that lack an overt inner antecedent.

The processing time penalty for having to sprout an implicit inner antecedent was similar when the implicit constituent was an argument and when it was an adjunct. This failure to observe an interaction suggests that neither implicit arguments nor implicit adjuncts are inferred at LF, which is presumably the level of linguistic representation relevant to processing a sluiced constituent. If implicit arguments (but not adjuncts) had in fact been inferred during reading of the initial clause, so that they were present at LF, we should have observed a smaller processing cost for sentences like (10b) than like (10d). As discussed earlier, Tanenhaus and his colleagues have presented evidence that implicit arguments are inferred in a discourse representation. Our data suggest that they are not inferred at LF.

It is perhaps logically possible that we observed no difference between implicit arguments and implicit adjuncts because *both* are inferred at the initial verb. In "John ate", one may infer not only that he ate something but also that he probably ate WITH something (and ate somewhere, in some manner, at some time, for some purpose, etc.). Apart from arguments that such extensive inferencing is not likely to take place (McKoon & Ratcliff, 1992), we note that this claim would require us to invoke some other account of the processing cost we observed in the sluiced constituent when the inner antecedent was implicit. We conclude that there is no reason to think that a *grammatical* (specifically, LF) representation of either the implied object or adjunct exists to facilitate the processing of the sluiced constituent with an implicit antecedent.

One might suggest (M. Tanenhaus, personal communication) that the verbs in our materials did not adequately support the inference of implicit arguments. We acknowledge that we do not know with certainty that tests like the ones Carlson and Tanenhaus (1988) and Mauner et al. (1995) used would have indicated the existence of inferred arguments in the discourse structure representations of our materials. However, the frequency counts and sentence completion questionnaire described earlier encourage our belief that our particular implicit argument items involved arguments that are frequent enough that they should have encouraged the inference of an object at some level of representation. We also note that short discourses like *John kept eating last night. The pasta must have tasted good*, which contain implicit arguments very similar to those in the sentences we used, seem to be as felicitous as those studied by Carlson and Tanenhaus. One could reasonably conclude that an implicit argument was inferred as the antecedent of the later definite description *the pasta* (although it remains to be determined experimentally whether such inference is automatic and cost-free, as inferring the agent of a passive seemed to be in the study of Mauner et al., 1995).

EXPERIMENT 2

The second experiment turns from asking how the processor infers a missing argument or adjunct to asking how it chooses from among multiple explicit possibilities. We investigated sentences like those in (11) which contain two indefinite NPs that could be copied into the sluice to be bound by the *wh*-operator (11a) versus only one (11b). (The / marks indicate the division into analysis regions used in Experiment 2.)

11a. Somebody claimed / that the president fired someone, / but nobody knows who. /

11b. Somebody claimed / that the president fired Fred, / but nobody knows who. /

Sentence (11a) contains two indefinite NPs that could serve as the inner antecedent of a legal sluice—*somebody* and *someone*. Our intuition about (11a) suggests that the lower indefinite, *someone*, is the preferred antecedent. Below we report a rating study to test the generality of this preference using the experimental materials. We suggest that, in general, focused antecedents are preferred as the inner antecedent of a sluiced constituent.

We use the term "focus" to designate the role a sentence constituent plays in the information structure of a sentence, most commonly the role of conveying novel information (Halliday, 1967; Vallduvi, 1992). Contrastive phrases, and phrases presenting new information or information not presumed already to be established, typically receive focus. In spoken English, a focused phrase receives a pitch accent (see Selkirk, 1984, 1995), but in written text, the placement of focus is typically not indicated overtly in the input. However, a reader presumably uses general principles of information structure to assign a focus structure to a sentence during reading. In English, the matrix subject generally expresses the topic, which is by default old, non-focused information. The default position for focus is a constituent in the VP, sentence-final in English. Thus, in sentences like (11), the embedded object will probably be focused. This could account for the intuitive preference for the lower antecedent in (11) above. We will later present a test of this account, using auditory material in which focus is marked by a pitch accent, as Ancillary Experiment 2, but first we will establish that the phrase in the position of the embedded object is the preferred source of the inner antecedent of a sluice.

In (11a), we expect readers to take the lower NP as the inner antecedent. Since, as an indefinite NP, it is appropriate, reading of the sluice should be reasonably fast. However, in (11b), if readers prefer to take the lower NP as the inner antecedent of the sluice, they will be blocked by the fact that this NP (*Fred*) is not indefinite, and thus not an acceptable variable to be bound by the *wh*-operator of the sluice. We predict that readers will have to devote processing time to overcoming this preference, and thus will read the sluiced constituent of (11b) more slowly than that of (11a).[3]

---

[3]One could alternatively predict that (11b) will be read faster than (11a) if it takes time to resolve the competition between the two legal alternatives in (11a). However, we do not think that sentence processing difficulty in this case is the result of competition between competing alternatives; instead, we propose that focus-preference dominates initial choice.

Methods

Sixteen pairs of sentences like those in (11) were constructed (see Appendix).[4] Each was followed by a simple short sentence to make the discourse more plausible. These short discourses were embedded in a list containing 96 other sentences and short discourses of a variety of constructions. Half the experimental sentences (and 50 of the other sentences) were followed by a simple two-choice question that did not require the identity of the antecedent of the sluiced constituent to have been determined. Thus the questions could not provide information about the interpretation of the ambiguity in the ambiguous sentences.

Thirty-six University of Massachusetts students were tested in an eye-tracking procedure (Rayner et al., 1989). An SRI dual-Purkinje eyetracker interfaced to a microcomputer that sampled eye position every 1 msec was used. The experimental session began with calibration (approximately 5 min) and a short practice list. Each trial began with a calibration check in which the subject fixated on a pattern of five boxes, following which a sentence was displayed in its entirety on a VGA screen controlled by the microcomputer. Order of sentences was individually randomised for each subject, and counterbalancing procedures were used to ensure that each sentence was tested equally often in both of its versions and that each subject received an equal number of sentences in each version. The microcomputer recorded the duration and position of each fixation and the accuracy with which questions following the sentences were answered.

Results

The sentences were divided into analysis regions as indicated in (11). Analysis region 3 contained the sluiced material, and analysis region 4 (not illustrated in (11); see Appendix) contained the final sentence. Table 3 presents the first pass times (the sum of all fixation durations from first entering to first leaving a region) in terms of uncorrected millisecond reading times, the percentages of trials on which a first-pass fixation in a region was followed by a regressive eye movement to an earlier region, and the total reading times (the sum of all fixation durations in a region, including re-reading fixations).

The interesting effects began in the sluiced region, region 3, and appeared to spill over into the next region, region 4. Reading was slowed in the

---

[4]In fact, we made up four versions of each sentence, two with a factive verb like *realise* and two with a non-factive verb like *claim*. A factive verb entails the truth of its complement, which might encourage the reader to accept the existence of an antecedent for the free indefinite NP in the complement sentence. This might in turn make this NP more available as an antecedent for the sluiced constituent, speeding reading time for factives. However, factivity of the verb had no effect whatsoever, and will not be discussed further.

TABLE 3
Measures of Eye Fixations in Experiment 2

Condition	Region 1	Region 2	Region 3 (Sluiced)	Region 4 (Next Sentence)
First-pass time (msec)				
Ambiguous	953	1593	876	857
Unambiguous	953	1578	968	930
Percentage first-pass regressions				
Ambiguous	0	19	10	31
Unambiguous	0	15	13	35
Total reading time (msec)				
Ambiguous	1136	2053	1040	935
Unambiguous	1162	2106	1148	1060

unambiguous conditions, in which the reader had to treat the subject NP as the inner antecedent of the sluice, compared to the ambiguous conditions, in which the presumably focused object was available. The analysis of variance of first pass times in region 3 indicated a significant effect of ambiguity only in the by-subjects analysis [$F_1(1,35) = 6.85$, $P < 0.02$; $F_2(1,15) = 2.42$, $P > 0.10$]. However, taking advantage of the spillover into region 4, an analysis of variance that treated regions 3 and 4 as a two-level factor did provide evidence for faster reading in the ambiguous condition (866 *vs* 949 msec) averaged over the two regions [$F_1(1,35) = 5.27$; $F_2(1,15) = 5.81$, $P < 0.05$]. In addition, an analysis of total reading times for region 3 resulted in a significant effect of ambiguity [$F_1(1,35) = 6.83$; $F_2(1,15) = 4.93$, $P < 0.05$]. No other effects approached significance in the analyses of either first-pass or total times, and no effects were significant in the analyses of percentage regressions in region 3. Analysis of reading times expressed as deviations from time expected on the basis of length (as in Experiment 1) yielded patterns of results and levels of significance closely comparable to the reported analyses of unadjusted times.

Question-answering accuracy ranged from 78 to 83% correct with no substantial variation among conditions (80.5% correct ambiguous, 79.1% correct unambiguous).

Discussion

The faster reading times for ambiguous (11a) than unambiguous (11b) sentences was predicted on the assumption that readers prefer an antecedent which may be interpreted as focused. In the unambiguous sentences, readers were forced to choose the matrix subject as antecedent—an NP which is not typically focused in sentences like those

tested here. This apparently slowed reading time compared to the ambiguous sentences, in which the direct object (the default focus) was available as an indefinite NP. This interpretation could be tested in additional ways, including (a) forcing the reader to choose the direct object as inner antecedent by making the subject a definite NP, or (b) providing only a single (subject) NP in the material before the sluiced constituent. In either case, we would predict that the sluice would be read quickly, since no more focused constituent would be available as a tempting but inappropriate antecedent for the sluice.

The reading time effect was fully significant only in the total time measure and in the first-pass time measure when the region after the sluice was included. This delayed appearance of the effects is not surprising, given that slowing presumably reflects the need to infer the sluiced material, and that this inference is not justified until the reader determined that no relevant material follows the sluice.

ANCILLARY STUDIES

To provide additional information concerning our interpretation of the reading time results of Experiment 2, we conducted two ancillary studies. Both studies obtained judgements of the preferred interpretations of ambiguous sluiced constituents. The first assessed our intuitive judgement that the NP in default focus position would be the preferred inner antecedent for a sluice. The second examined whether explicit manipulation of focus via pitch accent in auditory presentation would alter preferences.

Ancillary Study 1

To confirm that the lower antecedent is the preferred antecedent in cases of ambiguity, the ambiguous experimental sentences from Experiment 2 were presented in a written questionnaire study, randomly mixed in with 56 other sentences measuring the plausibility of adjective–noun modification and the interpretation of ambiguous pronouns. Following each sluicing sentence, two options were presented corresponding to the possible antecedents for the sluiced constituent, as illustrated in (12):

12. Some teacher says that the students will flunk an exam—guess which one.
    Which one = ⎯⎯ some teacher ⎯⎯ some exam.

Sixty University of Massachusetts students were tested individually. They were instructed to go through the questionnaire checking off the answer corresponding to their interpretation of the sentence (or, for some of the

other sentences in the questionnaire, rating its plausibility).[5] The remaining subjects saw a different form, testing the other half of the sentence forms. Over three quarters (77.0%; SE = 0.02) of the responses indicated an interpretation in which the lower NP (*an exam*) was taken as antecedent of the sluiced constituent.

This rating study thus confirms that the downstairs NP was the preferred antecedent in our experimental sentences. This supports our conclusion that the ambiguous sentences in Experiment 2 were read quickly because, unlike the unambiguous experimental sentences, they permitted a downstairs antecedent for the sluiced constituent.

Ancillary Study 2

A final auditory judgement study provided a further check of our focus-based account of antecedent choice. Sixteen students (mostly from undergraduate linguistics courses, and all native speakers of American English) heard 12 spoken sentences like those in (13). Half the sentences heard by any one subject had focal accent (indicated by upper-case letters) on the matrix subject (13a) and half on the embedded object (13b):

13a. SOME TOURIST suspected that the hotelkeeper was hiding someone. Guess who?

13b. Some tourist suspected that the hotelkeeper was hiding SOMEONE. Guess who?

Two counterbalanced tape-recorded lists of sentences were prepared, with each form of a sentence (subject or object focus) appearing on one list and each list containing six subject focus and six object focus sentences. Half the subjects listened to the sentences on one list, while the other half heard the other list. After hearing each sentence, the subject checked off one of two possible answers, as illustrated in (14), to indicate whether the subject (14a) or object (14b) was taken as the antecedent of the auditory sentence:

14a. The tourist who suspected that the hotelkeeper was hiding someone was Don Knotts.

14b. The person who the tourist suspected the hotelkeeper was hiding was Don Knotts.

When the embedded object was focused, 72% of all responses indicated that it was the antecedent of the sluiced constituent (as in (14b)). However, when the matrix subject was focused, only 48% of all responses indicated that the antecedent was the embedded object. The effect of focus was

---

[5]Half the subjects saw one counterbalanced and randomised form of the questionnaire, in which half the sentences appeared with a factive verb and the other half with a non-factive verb (see footnote 4).

significant [$t(15) = 3.62$, $P < 0.01$; SE = 6.6%]. Certainly an overall bias (60%) exists towards interpreting the embedded object, not the matrix subject, as the inner antecedent of the sluiced constituent. However, this bias disappeared when the matrix subject received focal accent, indicating that focus is one factor that plays an important role in determining the choice of antecedent of a sluiced constituent.

CONCLUSIONS

Experiment 1 revealed a penalty for sluiced sentences with implicit antecedents, compared to ones with overt antecedents. The penalty was equally large for argument and for adjunct antecedents. This finding contrasts with previous evidence suggesting that implicit arguments are immediately represented—that is, entered into the perceiver's discourse representation (Carlson & Tanenhaus, 1988; Mauner et al., 1995). Since the linguistic analysis of sluicing we have adopted (Chung et al., 1995) claims that the structure required to interpret a sluiced constituent exists at LF, this contrast suggests the interesting possibility that implicit arguments are represented in discourse representation but are not readily accessible for operations at LF. This speculation needs to be tested in future work by tracking implicit arguments in a variety of contexts, and determining for what purposes they are immediately and readily accessible and for what purposes they act like they are not represented. In principle, the answers to these questions might help identify the fractures in the language processing system, revealing which mental operations are performed on which type or level of mental representation.

If implicit arguments are immediately represented (as shown by Mauner et al., 1995), and if implicit arguments are not visible at LF (as shown by Experiment 1), then we may tentatively conclude that finding an antecedent for PRO in a rationale clause (as required for the materials used by Mauner et al., cf. our example (9) above) is not an LF operation. Plausibly, it involves either a thematic representation or perhaps a full discourse representation embellished with inferences available from world knowledge. The interpretation of sentences like (15) supports the latter possibility:

15. The lights on stage were low to achieve the effect of a smoky pool hall.

In (15), presumably it is the director or the lighting designer—whoever is in charge of lighting effects—that controls interpretation of the PRO in the rational clause. When a "controller" is available from the thematic role of an implicit argument, as in the study of Mauner et al., presumably the most available, most relevant and least costly (in terms of inference complexity) controller will be the one corresponding to the implicit agent. In sentences

like (15), where no thematically provided controller is available, the controller must be inferred "from scratch" using world knowledge together with the content of the sentence containing the rationale clause.

Experiment 2 and the first ancillary study indicated a preference for embedded clause antecedents in ambiguous sluicing sentences. The lower antecedent was chosen more often in the rating study and, in the eye movement study, the ambiguous sentences were processed faster than the unambiguous sentences, which forced the matrix subject to be chosen as antecedent. This was attributed to a preference to assign focus within the VP, with the consequence that only the lower antecedent is likely to have been interpreted as focused in the ambiguous experimental sentences. The final study that was reported indicated that placing a pitch accent on the matrix subject of an auditory sentence changes the preferred interpretation of ambiguous sentences, presumably via the influence of pitch accent on focus. Our intuitions suggest that placement of pitch accent on the matrix subject of auditory versions of the unambiguous sentences of Experiment 2 (11c,d) would have facilitated their comprehension.

At present, we may only speculate about the reason why focus matters in the processing of ellipsis. We suspect that ultimately the reason can be traced to pragmatic factors: Language producers presumably make informative contributions to a discourse, avoiding the repetition of redundant material, while simultaneously making their contributions relevant to prior material. If focus is placed on new and important material, then the remnant in a clause with elided constituents is likely to be novel material relevant to the important contribution of a preceding clause. Hence the remnant is likely to be focused because it is novel and the antecedent is likely to be focused because it highlights important information.

Studying LF.   We view the research we have presented here as an initial attempt to study the processing of LF. Few processing studies of LF exist. Kurtzman and MacDonald (1993), Pritchett and Whitman (1995) and Tunstall (1997) discuss the psycholinguistics of resolving quantifier scope (e.g. *A boy ate each apple*), a topic which is generally analysed as involving LF representations. Shapiro and Hestvik (1995) addressed the question of the logical form (LF) of the elided constituent in VP anaphors. They used cross-modal semantic priming to demonstrate immediate activation of an antecedent of an anaphor in the elided VP in sentences like *The policeman defended himself and the fireman did too*. They reported measurable activation of *policeman*, which is the antecedent on the "strict" (co-referential, *the fireman defended the policeman*) reading of VP-ellipsis. It would be interesting to compare this activation with the amount of activation of the "sloppy" (bound variable, *fireman*) reading, which intuition (at least ours) suggests is preferred.

What have we learned about LF from the present study? Assuming that Carlson, Tanenhaus and colleagues are correct about implicit arguments, we may conclude that LF is separate from a discourse representation and is an identifiable representation lacking representation of implicit arguments. Furthermore, Experiment 2 suggests that processing of deleted constituents may largely be regulated by assignment of focus, bringing sluicing in line with other types of ellipsis, which appear to be focus-sensitive (Rooth, 1992).

REFERENCES

Carlson, G.N., & Tanenhaus, M.K. (1988). Thematic roles and language comprehension. In W. Wilkins (Ed.), *Syntax and semantics: Thematic relations*, pp. 263–300. New York: Academic Press.

Chung, S., Ladusaw, W., & McCloskey, J. (1995). Sluicing and logical form. *Natural Language Semantics*, *3*, 239–282.

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

Halliday, M.A.K. (1967). Notes on transitivity and theme: Part 2. *Journal of Linguistics*, *3*, 177–244.

Hankamer, J., & Sag, I. (1976). Deep and surface anaphora. *Linguistic Inquiry*, *7*, 391–426.

Just, M.A., Carpenter, P.A., & Woolley, J.D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*, 228–238.

Kurtzman, H.S., & MacDonald, M.C. (1993). Resolution of quantifier scope ambiguities. *Cognition*, *48*, 243–279.

Malt, B. (1985). The role of discourse structure in understanding anaphora. *Journal of Memory and Language*, *24*, 271–289.

Marcus, M.P., Santorini, B., & Marcinkiewicz, M.A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, *19*, 313–330.

Mauner, G., Tanenhaus, M.K., & Carlson, G.N. (1995). Implicit arguments in sentence processing. *Journal of Memory and Language*, *34*, 337–382.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*, 440–466.

Pritchett, B., & Whitman, J. (1995). Syntactic representation and interpretive preference. In R. Mazuka & N. Nagai (Eds), *Japanese sentence processing*, pp. 65–76. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Rayner, K., Sereno, S., Morris, R., Schmauder, R., & Clifton, C.J. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, *4*, 21–50.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*, 75–116.

Ross, J.R. (1967). *Constraints on variables in syntax*. Unpublished PhD, dissertation, MIT, Cambridge, MA.

Schütze, C.T., & Gibson, E. (1996). Argument preference in English PP attachment. Paper presented at the *Meeting of the Linguistic Society of America*, San Diego, CA, January.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Selkirk, E.O. (1995). Sentence prosody: Intonation, stress, and phasing. In J. Goldsmith (Ed.), *Handbook of phonological theory*, pp. 550–569. Oxford: Blackwell.

Shapiro, L.P., & Hestvik, A. (1995). On-line comprehension of VP-ellipsis: Syntactic reconstruction and semantic influence. *Journal of Psycholinguistic Research*, *24*, 517–532.

Tanenhaus, M.K., & Carlson, G.N. (1990). Comprehension of deep and surface verb phrase anaphors. *Language and Cognitive Processes*, *5*, 257–280.

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

Tunstall, S. (1997). *The interpretation of quantifiers: Semantics and processing*. Unpublished doctoral dissertation, Department of Linguistics, University of Massachusetts, Amherst, MA.

Vallduvi, E. (1992). *The information component*. New York: Garland Press.

Ward, G., Sproat, R., & McKoon, G. (1991). A pragmatic analysis of so-called anaphoric islands. *Language*, *67*, 439–473.

APPENDIX

Experiment 1 materials: Options indicated by (^); presentation regions indicated by (/)

1. John ate (something ^ with something) / but nobody / seems to be able / to figure out (what ^ with what).
2. Marcie telephoned (someone ^ from somewhere) / but Sam / doesn't remember (who ^ from where).
3. Kathy wrote (something ^ with something) / but nobody / seems / to remember (what ^ with what).
4. Bill cooked (something ^ with something) / but I / don't recall (what ^ with what).
5. Tamara used our oven today / to bake (something ^ for some reason) / but we / never found out (what ^ why).
6. Jonathan screamed (something ^ for something) last night / but we / don't know (what ^ for what).
7. Alice dreamed (something ^ for some reason) last night / but she / wouldn't say (what ^ why).
8. Samuel moaned (something ^ about something) / and we / would like / to know (what ^ about what)
9. The kids cleaned (something ^ somewhere) / but we / can't figure out (what ^ where).
10. The man cursed (something ^ about something) / but the pedestrian / couldn't tell (what ^ about what).
11. The poor soul sang (something ^ about something) / but nobody / could figure out (what ^ about what).
12. Martha mumbled (something ^ about something) ^ but the hostess ^ couldn't hear (what ^ about what).
13. The strange man pushed (someone ^ toward someone) / but we / couldn't see (who ^ toward who).
14. Charles checked (something ^ about something) in his calendar / but Rachel / couldn't see (what ^ about what).
15. The secretary typed (something ^ somewhere) / but I / don't know (what ^ where).
16. The child was painting (something ^ on something) / but the adults / couldn't tell (what ^ on what).

Experiment 2 materials: Options indicated by (non-factive ^ factive) and (indefinite ^ definite); analysis regions indicated by (/). Illustrative question shown for item 1

1. Some tourist (suspected ^ discovered) / that the hotelkeeper was hiding (someone ^ Madonna) / —Guess who. / Someone you've seen on TV. /

   Who figured out that the hotelkeeper was hiding someone?   a tourist   a maid

2. Somebody (claimed ^ realized) / that the President fired (someone ^ Fred), / but nobody knows who. / And nobody will find out. /

3. Somebody (reported ^ discovered) / that the Dean expelled (two students ^ Margaret), / but we don't know who. / We can't even guess who it was. /

4. Two protesters (asserted ^ saw) / that the police clubbed (somebody ^ the mayor), / but I don't know who. / Probably there will be an investigation. /

5. Some janitor (thought ^ noticed) / that the intruder was (some faculty member ^ Mr. Williamson), / but I don't know which one. / The campus is petting dangerous. /

6. Some journalist (guessed ^ knew) / that the cult group imprisoned (some child ^ George) against his will, / but it isn't clear who. / Clearly the issue needs to be investigated. /

7. Some angry townperson (hopes ^ knows) / that the police will ticket (some storeowner ^ the town druggist) for negligence, / but I don't know which one. / The storeowners are responsible for taking care of the sidewalk. /

8. Three videos (suggest ^ show) / that the diplomat said (some words ^ the F word) under his breath, / but I don't know which ones. / Undoubtedly he would be embarrassed if someone tape recorded them. /

9. Some drivers (claim ^ revealed) / that the empty wheelchair hit (three cars ^ the TV van) / — guess which ones. / We aren't telling. /

10. Some teacher (says ^ knows) / that the students will flunk (an exam ^ the oral exam) / — guess which one. / I bet it is my favorite one. /

11. Someone (thinks ^ knows) / that the Governor will appoint (someone local ^ Miss Smith) / — guess who. / Probably it's a friend of his. /

12. Some mayor (assumes ^ realized) / that the citizens want (some new law ^ the tax increase) to pass / —guess which one. / It's fun to guess. /

13. Some spectator (said ^ knew) / that the player took (some drug ^ steroids), / but the investigators couldn't find out which one. / Who really wants to know? /

14. Two pieces of evidence (indicated ^ proved) / that the official took (several things ^ the Renoir) illegally / — guess which ones. / Probably the usual ones. /

15. Some customers (asserted ^ realized) / that the owner knew (some mafioso ^ Mr. Antonio) / — guess which ones. / Nobody I know. /

16. Some regulating procedures (indicate ^ prove) / that the industry tolerates (some types of disasters ^ disposal of low level waste) / — guess which. / It's discouraging to think about it. /