

Sluicing Annotation Guide

September 14, 2014

Contents

1	General outline	4
1.1	Preliminaries	4
1.2	The tagset in brief	4
1.3	General annotation workflow	6
1.4	What Follows	7
2	NOT A SLUICE : Things to Set Aside	9

3	SLUICE	10
3.1	The bounds of the sluice	10
3.2	Swiping	11
3.3	Type of the Sluice	11
3.3.1	Plain Entity Sluices	11
3.3.2	<i>Which</i> Sluices	11
3.3.3	Degree Sluices	12
3.3.4	Manner Sluices	12
3.3.5	Reason Sluices	12
3.3.6	Temporal Sluices	13
3.3.7	Locative Sluices	13
3.3.8	Classificatory Sluices	13
3.3.9	Possessor Sluices	13
3.3.10	Passive Subject Sluices	13
3.3.11	PP Sluices	14
3.3.12	Focus Sluices	14
3.3.13	None of the Above	14
3.4	Additional Annotations at the Sluice	14
3.4.1	Root or Non-root	14
3.4.2	Sluices Without Antecedents	15
3.4.3	Problems	15
4	ANTECEDENT	16
4.1	The bounds of the ANTECEDENT	16
4.2	Multiple Antecedents	18
4.3	Elliptical Antecedents	18
4.4	Ambiguity	19
4.5	Ignoring Material	20
4.6	Discontinuous Antecedents	20
5	CORRELATE	22
5.1	Typology of Correlates	24
5.2	Extension of the Correlate	25
6	The ELLIPSIS SITE	27
6.1	General Principles	27
6.2	The Form of the ELLIPSIS SITE	29
6.2.1	Omit As Much of the Correlate As Possible	30
6.2.2	Disjunctive Correlates	31
6.2.3	Avoid Pronominalization	31
6.2.4	Islandhood	32

6.3	Dealing with Differences in Form	32
6.3.1	Changes_fin	33
6.3.2	Changes_tense	33
6.3.3	Changes_indexical	33
6.3.4	Changes_pro	34
6.3.5	Changes_case	34
6.3.6	Changes_other	34
6.3.7	NewWords	35
6.3.8	Modal changes	35
6.3.9	E-Type interpretations	36
6.4	Punctuation	36
7	PREDICATE	37
7.1	Identifying the Predicate	37
7.1.1	Copular Constructions	37
7.1.2	Larger Predicates	37
7.1.3	Functional predicates	37
7.1.4	Existential constructions	38
8	E-TYPE	39
8.1	Some exemplary cases	39
8.1.1	WH-Phrases	39
8.1.2	Quantifiers	40
8.1.3	Indefinites	40
8.2	Type	40
8.3	Form in the Ellipsis Site	41
8.4	Ambiguity and necessity	41
9	NOTE	42

1 General outline

1.1 Preliminaries

In what follows, the term `SLUICE` refers to the orphaned `WH`-phrase; the term `ANTECEDENT` refers to the clause or clause-like element which provides the basis for the semantic interpretation of the sluice, and the term `CORRELATE` refers to a phrase contained within the Antecedent which corresponds to the sluice. For example in (1):

(1) I know they hired a phonologist, but I'm not sure which one <they hired>.

The phrase *which one* is the `SLUICE`; the phrase *they hired a phonologist* is the `ANTECEDENT`, and the phrase *a phonologist* is the `CORRELATE`. (The term 'inner antecedent' is also used of the element *a phonologist* in (1); we will try to be consistent in using `CORRELATE` here.)

Note that we have expressed the intuitive paraphrase of the Antecedent after the sluice inside angled brackets. This we will call the `ELLIPSIS SITE`. Properly identifying and tagging these four elements (`SLUICE`, `ANTECEDENT`, `CORRELATE`, and `ELLIPSIS SITE`) is the core of what we have to do.

1.2 The tagset in brief

In annotating sluices, you will be asked to fill out the following elements, some of them mandatory, some of them optional.

- `ANTECEDENT`: the material that serves as the intuitive fill in for the meaning of the ellipsis
- `CORRELATE`: the material in the `ANTECEDENT` that is replaced or elaborated on by the sluice itself. The Correlate has one attribute:
 - `CorrType`: what kind of Correlate it is
- `PREDICATE`: the material that functions as the main predicate for the clause in the `ANTECEDENT`
- `SLUICE`: the actual sluice site. You can specify the following attributes of sluices:
 - `root`: whether the sluice is a root or non-root sluice
 - `island`: whether the sluice is out of an island context
 - `grammaticality`: mark whether the sluice is of medium or low grammaticality
 - `missingPred/missingAnte`: whether there is or is not a corresponding Predicate or Antecedent

- **problematic** : Mark if there is something seriously wrong about this example. Add comments in the notes.

For every example, you will also be required to paraphrase the **ELLIPSIS SITE** in the notes section. It will show up between angled brackets. In addition at this point you will be asked to mark certain differences of form that there may be between the Antecedent and the ellipsis site. We annotate for the following six differences in particular:

- **changes_fin** : when the **ANTECEDENT** and ellipsis site have different specifications for finiteness
- **changes_tense** : disagreement in tense
- **changes_pro** : disagreement in whether the subject is overt or covert
- **changes_indexical** : when a pronoun in the **ANTECEDENT** and a corresponding pronoun in the **ELLIPSIS SITE** have different person or number values
- **changes_case** : when a noun in the **ANTECEDENT** and a corresponding noun in the **SLUICE** have different cases
- **changes_other** : any other noticeable differences
- **newWords** : insertion of words needed to render the material of the **ELLIPSIS SITE** grammatical

Beyond these, you may find yourself marking several more specialized forms.

- **NOT A SLUICE** : Mark this and NOTHING else if the example involves something that is not a sluice. In particular cases of interest, mark the **TypeNot** as one of interest (**coordinate**, **amalgam**, **reverse**).
- **ELLIPSIS ANTECEDENT** / **ALTERNATIVE ANTECEDENT** : Use these in cases where there are multiple propositionally identical Antecedents. **ELLIPSIS ANTECEDENT** is for cases where the closest Antecedent is itself an elided form (often because of VP ellipsis) and **ALTERNATIVE ANTECEDENT** is for cases involving multiple very close (often “sandwiching”) Antecedents.
- **E-TYPE** : Material in the **ANTECEDENT** that: a) is not definite but b) when paraphrased, seems to require a definite expression. Please mark the EType’s **Variety** and paraphrase it with an appropriate pronoun (like *he*, *it*, *they* etc).

- **IGNORE** : Use this to mark material which is present in the **ANTECEDENT** but missing from (the interpretation of) the **ELLIPSIS SITE** (and which is therefore not copied over).
- **NOTE** : Use this to make notes of interesting material.

All of this will be clarified and spelled out in detail in what follows.

1.3 General annotation workflow

We suggest the following workflow for tagging each instance.

1. Is the example a sluice or not? First determine whether this is a sluice as we define it or something else. If it something else, mark as **NOT A SLUICE**, determine the **TypeNot**, and move on.
2. If it is a sluice, highlight & select the **SLUICE**. Also mark the **SluiceType**.
3. Is this a root sluice or an embedded sluice? Mark **root** accordingly.
4. Is there an **ANTECEDENT**? If so, mark it and continue. At this point, you may have to contend with **ALTERNATIVE ANTECEDENT** and **ELLIPSIS ANTECEDENT** cases, so be careful. If there is no Antecedent, mark **MissingAnte** and you are done with that instance.
5. Is there a **CORRELATE**? If so, mark it and its **CorrType**.
6. Is there a main **PREDICATE**? If not, mark **MissingPred** on the **SLUICE**. If there is, mark it.
7. Now go back to the **SLUICE** and fill in the **ELLIPSIS SITE** in the **Notes** section. There are very specific conventions for this, which are indicated there. To help you, there are buttons that will copy in a previously highlighted **ANTECEDENT**. Finally, fill in the options about changes in form, about island-hood, and so on.
8. In filling in the **ELLIPSIS SITE**, you will encounter **E-TYPE** and **IGNORE** material. Tag all of that within the scope of the **ANTECEDENT** text.
9. If anything strikes you about an example, write a **NOTE**.

After this procedure, you should have (ideally):

- A **SLUICE** with free-text **ELLIPSIS SITE** and several options marked as well as a **SluiceType**.
- An **ANTECEDENT**.

- Possibly a **CORRELATE**, with **CorrType**.
- A **PREDICATE**.
- Possibly some **E-TYPE** or **IGNORE** tags.

1.4 What Follows

The remainder of this document specifies the tagset and the various options you must annotate in a more detailed and systematic way. We will proceed through the tags roughly in terms of the workflow: **NOT A SLUICE**, **SLUICE**, **ELLIPSIS SITE**, **ANTECEDENT**, **CORRELATE**, **PREDICATE**, and **E-TYPE**. Figure 1 lays out the complete tagset in tabular form.

MANDATORY TAGS

SLUICE : sluice site.

- **SluiceType** [Degree, Manner, Reason, Temporal, Locative, Classificatory, Possessor, Passive, PP Focus, None of the Above]
- **island** : whether sluice ‘crosses’ an island
- Mismatches [Finiteness, Tense, Person, Case, Subject Overtness, Additional Words, Other]

ANTECEDENT : intuitive fill for **ELLIPSIS SITE**

PREDICATE : main predicate for clause in **Antecedent** .

CORRELATE : material in **ANTECEDENT** replaced or elaborated on by WH-phrase.

- **CorrType** [Indefinite, Definite, Pronoun, Strong Quantifier, WH-phrase, Name, Disjunction, Temporal/Locative, Degree/Extent]

OPTIONAL TAGS

ELLIPSIS ANTECEDENT : **ANTECEDENT** is elided

ALTERNATIVE ANTECEDENT : Secondary **ANTECEDENT**

E-TYPE : Indefinite in **ANTECEDENT** that is anaphoric in **ELLIPSIS SITE**

IGNORE : Material not retained in **ELLIPSIS SITE**

NOTE : Note about something interesting.

Figure 1: Sluicing 1.5 Tagset

2 NOT A SLUICE: Things to Set Aside

There are several constructions involving WH-phrases that look like sluices, but are not clearly sluices (this is a matter of theoretical debate). For all of these, mark them as NOT A SLUICE, and for the specific types below, mark the TypeNot of non-sluice in the relevant attribute.

- REVERSE SLUICES **reverse**: These are cases where two or more WH-expressions are coordinated and followed by a TP that intuitively links to both. (It is unclear if the first WH-phrases are sluices or if the WH-phrases form a coordinated specifier.) Mark the WH-part as NOT A SLUICE.

(2) Nussey says he 's already accepted another position but would n't say what or where the job was. [5290]

- AMALGAMS **amalgam**: These are instances where you have something that looks like a sluiced clause syntactically substituting for an actual expression in the matrix sentence. The potentially sluicing clause ("interrupting clause") conveys the knowledge or surprise of the speaker or addressee, and the potential Antecedent (the "host clause") contains it. Mark the WH-phrase in the interrupting clause as NOT A SLUICE.

(3) He put the mail I don't know/you know/you'll never guess where.

(4) Nubheds : Grateful for trade Had Stars center Scott Thornton remained with the Montreal Canadiens , he would be doing ...
" I do n't know what , " he said . [360]

- COORDINATIONS **coordinated**: These are cases in which there are several WH-phrases connected by conjunctions/disjunctions/commas, but in which there is no overt TP next to the WH-phrase (unlike reverse sluicing). We label these as NOT A SLUICE because the more sophisticated properties of the Antecedent are difficult to mark.

(5) Oh sure , the recently signed agreement mentions the reuniting of families separated during the war .

But it does n't mention how many and when . [137786]

(6) Asiana 777 Broke Apart on Landing: We Can See How But Not Why

3 SLUICE

There are three issues that one must contend with when annotating the SLUICE :

- (i) what the bounds of the SLUICE are and what kind of SLUICE it is,
- (ii) what the form of the ELLIPSIS SITE is, and
- (iii) how the ANTECEDENT clause and ELLIPSIS SITE differ from one another.

In this section we deal with (i)—which essentially has to do with the orphaned WH-phrase. The following section deals with properties of the ELLIPSIS SITE and its relation to the Antecedent ((ii) and (iii)).

In addition, there are several properties of sluicing examples that may have no natural home (e.g., if the example lacks an ANTECEDENT). SLUICE is the location where these are marked. We turn to these at the end of the current section.

3.1 The bounds of the sluice

It should be relatively easy, in general, to appropriately identify the (WH-phrase of the) SLUICE . In general, it will contain all and only the WH-phrase that is orphaned. To the extent that you will encounter issues here, they will probably have to do with what counts as being part of this phrase – do surrounding prepositions, adverbs, or other material count as part? In general, be conservative and choose the minimal span that can reasonably be called the WH-phrase. There are four pertinent exceptions (in the discussion below the ELLIPSIS SITE is also specified for clarity):

- PREPOSITIONS: If the WH-phrase is part of a PP, include the preposition inside the sluice.

(7) They talked to him, but I don't know about/after what <they talked to him> .

(8) He likes her, but I don't know to what degree <he likes her> .

- NEGATION: If negation follows the WH-phrase, include it in the sluice. Do not include it if it precedes the WH-phrase.

(9) an MGM executive asked if I thought I could write a sequel
I said , “ I do n't know why not <I could write a sequel > . ” [63198]

(10) Chocolate seduces, but science knows not why <chocolate seduces>

- ADVERBIAL MODIFIERS: If there is a modifier to the WH-phrase, include it. All relevant examples we have seen are cases with an adverbial modifier of amount or precision (e.g., *just/exactly/precisely what*)

3.2 Swiping

Under sluicing in English and in some other Germanic languages, prepositions are allowed to follow a WH-pronoun which they govern. This inversion is known as ‘swiping’. Some examples are given in (11).

- (11) a. He told me he had talked to somebody but he wouldn’t say who to.
b. He went to the movies , but I do n’t know who with.

There are two guidelines to follow about such cases—(i) include the preposition in the SLUICE, and (ii) do not ‘fix’ the word order. So the two examples in (11) will be handled as in (12):

- (12) a. He told me he had talked to somebody but he wouldn’t say who to <he had talked> .
b. He went to the movies, but I do n’t know who with <he went to the movies> .

despite the fact that (12a) and (12b) would be ungrammatical if pronounced.

3.3 Type of the Sluice

We also at this point want to annotate for the type of the SLUICE in SluiceType . The categories of the typology (we recognize twelve at present, along with a catch-all ‘none of the above’ category) have both a syntactic and a semantic basis, and it should be possible almost always to determine which type is appropriate by looking at the sluice alone. We go through the subtypes currently recognized one by one in the sections that follow.

3.3.1 Plain Entity Sluices

Entity sluices are defined by WH-expressions like *who* or *what* which, in the general case at least, ask about the set of individuals which make some proposition true or false. Typical instances will be like those in (13):

- (13) a. They interviewed somebody , but they won’t say who.
b. They’re going to get me some expensive thing as a graduation gift, but I’ve no idea what.
(14) a. He’s very angry but he won’t say what about.
b. He’s in love, but he refuses to say who with.

In the examples of (13), the antecedent contains a CORRELATE ; in (14), it does not.

3.3.2 Which Sluices

We reserve a different annotation (Which sluice) for the type in (15):

- (15) a. They interviewed one of the candidates but they won’t say which (one) (of them).

- b. They're going to get me **one of those books** as a graduation gift, but I've no idea which (one) (of them).
- c. He's very angry about one of these policies, but he won't say which (one) (of them).
- d. He's in love with one of the neighbors, but he refuses to say (with) which (one) (of them).

In such cases, the immediate discourse context makes salient a set of entities or individuals of whom or of which some property or eventuality holds; the question projected by the sluice then asks for the identity of the specific member of that set of which the property of eventuality holds (while presupposing that there is a unique such individual or entity). Such sluices are characterized by the WH-determiner *which* and they will probably always involve a **CORRELATE**.

3.3.3 Degree Sluices

Degree sluices are defined by WH-expressions (like *how much*, *how many*, *how far* and so on) which ask about the degree or the extent to which some property or eventuality holds:

- (16) They advanced, but I'm not sure **how far**.
- (17) We've made **progress**, but I'm not sure **how much**.

Notice that such sluices come in two subtypes; some have a **CORRELATE** (which is *progress* in (17), for example) and some do not (as in (16)). The type of the WH-phrase is the same in the two cases, however, and that is all that we care about for this part of the annotation.

3.3.4 Manner Sluices

Manner sluices are introduced by interrogative manner adverbials:

- (18) He fixed this, but I'm not sure **how**.
- (19) He managed to solve this **somehow**, but I'm not sure **in what manner**.

Once again the antecedent may or may not contain a **CORRELATE** (no in (18), yes in (19)).

3.3.5 Reason Sluices

Reason sluices are introduced by interrogative adverbial phrases expressing reasons or purposes:

- (20) He fixed this, but I'm not sure **why**.
- (21) He contacted me, but I'm not sure **for what reason**.

3.3.6 Temporal Sluices

Temporal sluices are introduced by temporal interrogative phrases:

- (22) He fixed this, but I'm not sure **when** .
(23) This has been fixed, but I'm not sure **for how long** .
(24) We'll finish this manual **someday** , but I'm not sure **in what year** .

(24) includes a **CORRELATE** (*someday*); (22) and (23) do not.

3.3.7 Locative Sluices

Locative sluices are introduced by locative interrogative phrases:

- (25) She's planted kale, but I'm not sure **in which bed** .
(26) She's planted kale **somewhere** in the garden, but I'm not sure **exactly where** .

As before, (26) involves a **CORRELATE** (*somewhere*), but (25) does not.

3.3.8 Classificatory Sluices

Classificatory Sluices are introduced by interrogative phrases denoting kinds or types or classes:

- (27) She bought **a car** , but I don't know **what kind** .

The **ANTECEDENT** of (27), of course, includes a **CORRELATE** (*a car*). As far as we know at present, this is true of all Classificatory Sluices.

3.3.9 Possessor Sluices

Possessor sluices are introduced by an interrogative possessive:

- (28) She stole **a car** , but I'm not sure **which resident's** .
(29) She stole **someone**'s car, but I'm not sure **whose** .

3.3.10 Passive Subject Sluices

Passive Subject sluices are introduced by interrogative phrases which express the demoted agent of a passive clause. Such phrases may be Prepositional Phrases headed by passive *by*, as in (30) and (31):

- (30) She was attacked, but I'm not sure **by who(m)/who by** .
(31) She was attacked by **someone** , but I'm not sure by **who(m)/who by** .

Alternatively, they may be introduced by a bare interrogative DP, as in (32):

- (32) She was attacked by **one of the reviewers** , but I'm not sure **which one** .

3.3.11 PP Sluices

PP sluices are introduced by interrogative phrases governed by prepositions:

- (33) She took **a picture** but I do n't know **what of** .

Use **PP sluice** only where there is not already a more specific descriptor that can be used. Notice that many of the examples above involve PPs: (19), (21), (23), (24), (25), (30) and (31). For such cases, however, we will always prefer the more specific annotations provided above.

3.3.12 Focus Sluices

In **Focus** sluices, a **CORRELATE** (often definite) is extended with a focus particle like *else*:

- (34) **Sally** was there but I do n't know **who else** .

3.3.13 None of the Above

The catch-all category. Use **NoneoftheAbove** if the interrogative phrase does not naturally fit into any of the ten sub-types just illustrated.

3.4 Additional Annotations at the Sluice

3.4.1 Root or Non-root

Mark if this a **root** sluice. Root sluices are unembedded, as in (35); non-root sluices are sub-parts of larger structures, as in the examples of (36).

- (35) a. A: We should buy a painting. B: **Which one?**
b. A: We should go home. B: **Why/when/what for/how?**
- (36) a. The university has to change, but **in what ways** is not clear.
b. The university has to change, but it's not clear **in what ways** .

One good diagnostic for a root sluice is that the **ELLIPSIS SITE** will require subject-auxiliary inversion, as in the cases below:

- (37) To which Tom Minnery , vice president of public policy for Focus on the Family , a conservative religious group , said : “ Of course government legislates morality . The only question is : **whose morality <does government legislate> ?** ” [17435]

(38) If it 's not an on-going pattern of gender discrimination , then **what <is it>** ? [92602]

(See section 6.3.6 below for additional discussion of inversion and how to annotate such cases.)

3.4.2 Sluices Without Antecedents

This is also the point at which you should mark those sluices for which there is no Antecedent with **MissingAnte** . It will of course be important for you to check the surrounding context carefully when you make this determination.

Cases of this type that have been noted in the literature are like those in (39), where the italics provides some non-linguistic situational context:

- (39) a. *Woman gets in to taxi.*
DRIVER: Where to, lady?
b. *Distraught observer surveys a scene of mass destruction.*
OBSERVER: Why?

It might also occur that the **ANTECEDENT** is missing a **PREDICATE** . If so, please mark **MissingPred** .

3.4.3 Problems

Occasionally, an example is difficult for reasons beyond the usual. We have two notations on the **SLUICE** to help identify these.

First, you may find the example to be of uncertain grammaticality. If so, mark the **Grammaticality** as **Low** or **Medium** , depending on your judgment. We will assume anything unmarked is **High** , so you can leave grammatical cases unmarked.

In addition, there may be some issue we haven't dealt with yet. If so, please mark the sluice as **Problematic** and make a note for our discussions. Forward the example to us as well.

4 ANTECEDENT

The ANTECEDENT is the proposition-denoting span of text that most closely provides the content of the ELLIPSIS SITE. If there is an identifiable Antecedent, mark it. In general, anything which is semantically propositional can function as an Antecedent. In syntactic terms, this can be a finite clause (40), a non-finite clause (41), or a gerund (42).

(40) We should meet in some coffee-shop ; you decide which one.

(41) Sally seems to have left , but I'm not sure why.

(42) I remember meeting him , but I don't remember when.

Note that in the latter two cases, the marked ANTECEDENT is discontinuous to indicate the semantic value of the non-overt gap/trace/null argument in the non-finite clause and gerund; for more discussion, see the section on discontinuous Antecedents below. In addition, both of these cases involve morphosyntactic mismatches in tense between the Antecedent and sluice, and should be marked as `changes_tense`.

4.1 The bounds of the ANTECEDENT

We aim to choose an Antecedent that most intuitively could provide the appropriate content for the ELLIPSIS SITE. There are three general rules:

- Make the Antecedent as small as possible.
- Make the Antecedent as close to the sluice as possible.
- Prefer preceding Antecedents to following ones.

The first rule is the one that will apply in most cases. It means that you should find the smallest proposition denoting constituent, e.g., the lower clause in a control or raising construction, the embedded proposition of proposition selecting predicates, small clauses, etc. Often, this will be a TP, and so the Antecedent should include TP level adverbs (e.g., temporal adverbials). In the general case, CP level information (C and higher: *if*, *that*, WH-phrases) should not be included in the ANTECEDENT. Nor should rhetorical particles like *so*, *but*, *however*, *still* and so on.

Such elements should be included only if there is good reason to believe that they form part of the elliptical content. Two cases of this type come up with some frequency—one is the case of verbs that have undergone T-to-C movement (“Subject Aux Inversion”) in root questions.

Another case in which such choices arise is when the ANTECEDENT is a WH-question. Here the question arises: Is the WH-phrase of the Antecedent material to be included in the ELLIPSIS SITE or not? In answering that question, we ask if it is important for the meaning of the sluice that the

WH-phrase be included in the **ELLIPSIS SITE** . If it is, we will want to include the WH-phrase in the Antecedent. This possibility arises principally when the WH-phrase has a so-called **E-TYPE** interpretation in the **ELLIPSIS SITE** . These matters are discussed in detail in [section 8](#), but for now consider examples like (43):

(43) **Who** did the suspect call and **when** ?

An important facet of the interpretation of (43) is that the sluice is interpreted along the lines of (44):

(44) Who did the suspect call and when did he call them?

in which the reference of the pronoun *them* derives ultimately from the WH-phrase of the Antecedent. In such cases, we must include the WH-phrase in the **ANTECEDENT** . For (43), matters are clear. For other, apparently similar cases, however, it is less clear how to proceed. Consider (45), for instance:

(45) I know when he left, but I don't know why.

For such a case, the crucial question for the annotator is whether the speaker is wondering why the individual left or why he left at a particular time. The first possibility requires an annotation like (46a); the second requires an annotation like (46b). Notice that they differ exactly in whether the Antecedent is TP (as in (46a)) or CP (as in (46b)).

- (46) a. I know when **he left** , but I do n't know **why <he left>** .
b. I know **when** **he left** , but I do n't know **why <he left then>** .

This is not easy to resolve. To avoid needless complexity, if the context is unclear, we will in general disfavor the larger candidate (46b) and favor the smaller (46a); this aligns with what seems to be the natural interpretation of (45), and also with a general dispreference for E-Type interpretations in annotation (see [section 8](#) for a fuller discussion of these issues).

Similar kinds of questions arise for those adverbials that attach relatively high in the clause and a decision has to be made whether or not they should be included in the **ANTECEDENT** . A case in point is (47):

(47) About a year ago , **word got around the department that he was Officer Otto** , the informer , though no one seems to know **how <word got around the department that he was Officer Otto>** . [111865]

The convention we have adopted for such cases is that we will include in the **ANTECEDENT** bounding adverbials up to TP only if that is clearly warranted by the interpretation. Therefore in the case

of (47), the adverbial *about a year ago* will not be included in the Antecedent. (An adverbial is at the TP-level if it can appear between *that* and the subject.)

4.2 Multiple Antecedents

It happens surprisingly often that there is more than one Antecedent for a given sluice. In these cases, our convention is to mark the closest Antecedent, and if the two are equally close (one before and one after), to mark the preceding Antecedent.

One common case of this is a “sandwich”, where there is an Antecedent, and then a sluice connected to another possible Antecedent by a *but*-clause, like the examples below. In such cases, choose the closest Antecedent first, and if they are equally close, choose the one on the left. Mark the alternative Antecedents with **ALTERNATIVE ANTECEDENT**.

- (48) “ Mario ’s going to be OK , ”said Westy .
 “ I do n’t know when <Mario ’s going to be OK> , but he ’s going to be OK . ”
 [100559]
- (49) But Thursday the market for other California municipal bonds recovered a bit .
 “ It ’s difficult to say how much <the market for other California municipal bonds recovered> ,
 because liquidity is relatively low and trading is sporadic , ” said Ian MacKinnon , se-
 nior vice president of fixed-income for the Vanguard Group of mutual funds .
 “ But I ’d say the California market has bounced back . ” [35463]

4.3 Elliptical Antecedents

Given the convention above, it may be the case that the closest Antecedent has itself undergone some kind of ellipsis. While this is the “true” Antecedent, since the important material is non-overt, we will not be able to mark important information on it. In such cases, our convention is:

- mark the elided expression as **ELLIPSIS ANTECEDENT**
- fill in the elided content in the Notes of the **ELLIPSIS ANTECEDENT**
- mark the closest overt Antecedent for ellipsis as **ANTECEDENT**
- mark all the important material on the **ANTECEDENT** , except whatever clearly comes from **ELLIPSIS ANTECEDENT**

- (50) “ We spoke with everyone who had useful information ,
 <we spoke with> some people multiple times , ” Leadbetter said , declining to say
 how many <they spoke with> [144677]

4.4 Ambiguity

Some of the hardest problems in selecting Antecedents come up when there is the possibility of ambiguity; more often than not, the ambiguity in question depends on there being two potential Antecedents, one larger (more inclusive) and one smaller (less inclusive), for the sluice. In these cases, all other things being equal, we go with the smallest justifiable candidate, unless context makes it clear that the larger candidate is salient.

In some of these cases there will be a structural ambiguity with no obvious interpretive consequence. One complex case is 104959:

- (51) Hattaway said that Gore's plan included
details that would protect consumers from higher gasoline prices , but he was unable to explain
how <those details would protect consumers from higher gasoline prices > . [104959a]

The Antecedent could be the larger: *Gore's plan included ...* or the smaller: *those details would protect consumers from higher gasoline prices*. While this ambiguity is present, in principle, it is both subtle and not clearly justified by the surrounding context, so we opt for the smaller Antecedent.

In other cases, however, we must decide differently. One case in which the 'higher' interpretation is salient is 21666:

- (52) It's difficult to get the black church to deal with the issues," said Campbell.
It's not that we don't want to, it's that we don't know how . [21666]

Here, there is a genuine ambiguity between the 'larger' interpretation (how to get the black church to deal with the issues) and the 'smaller' interpretation (how to deal with the issues). This is really an ambiguous case. For such cases like these, click the **Ambiguous** button at the top and it will clone the example for you to re-annotate. Since it clones everything already present, try to annotate as much common information as you can before hitting the button. This issue is most vexing in *why*-sluices when there is modality present. As we **already saw**, many modal auxiliaries expressed as non-finite expressions in Antecedents have a tendency to become a vaguer modal in the sluice. This can make the higher and lower readings pragmatically equivalent. Again, in these cases we opt for the lower, smaller Antecedent.

- (53) they asked me in two or three games for four or five series to sit down on third-down plays ,
and I did n't understand why <I MODAL sit down ...>
and I did n't understand why <they asked me ... to sit down ...> [40784b]

Because *why they asked me* can reference either reasoning or some internal motivation, in this context it is pragmatically synonymous with *why I should (in their view) sit down*. Keeping

with our convention, we mark these cases with the lower clause. Note that we will choose the upstairs reading then only in cases where it is clear that we are talking about the managers own internal motivations for doing something (e.g., if the whole example were followed by *because they clearly weren't interested in helping the team, so I was curious how this would benefit them*).

4.5 Ignoring Material

In many cases, the syntactic Antecedent has superfluous material that intuitively doesn't belong inside the **ELLIPSIS SITE**. Such material typically includes markers of not at issue content, such as:

- commentary expressions: *I suppose*
- non-restrictive relative clauses/appositives
- parentheticals
- rhetorical particles: *but, however, still*
- additive particles: *even, also, too*

For these kinds of expressions, if the material is at the edge of the clause, simply don't include it in the **ANTECEDENT**. If it is clause internal, mark it with **IGNORE**. Remember that **IGNORE** is used only for material in the **ANTECEDENT** that is not part of the **ELLIPSIS SITE**; **IGNORE** tags should never be outside of the bounds of **ANTECEDENT**.

4.6 Discontinuous Antecedents

There are several cases in which the **ANTECEDENT** will be discontinuous (with multiple fragments).

- control, raising, relative clauses: The Antecedent for the empty category (trace, operator, or PRO) is marked as part of the **ANTECEDENT**, and *changes_PRO* is marked on the sluice. For the relative clause, do not make the relative pronoun part of the Antecedent. Raising clauses include the usual suspects (*seem, appear, be likely*, etc.) as well as the semi-modals (*have to, got to, need to, used to*).
- coordination: the subject (or other element, but it's often a subject) is part of the Antecedent, no changes are marked on the sluice
- interpolation: occasionally, the sluiced clause is inside the Antecedent (e.g., in an aside). In such cases the Antecedent is necessarily discontinuous:

- (54) She had been the youngest -LRB- 14 -RRB- and tiniest -LRB- 4 feet 5 inches , 75 pounds -RRB- of the Magnificent Seven who found gold in Atlanta .
Now , Moceanu is nine inches taller and significantly -LRB- she wo n't say
how much <Moceanu is significantly heavier> -RRB- heavier . [135548]

Recall that many cases of intervening material (parenthetical remarks, particles, etc.) that are not intuitively part of the ELLIPSIS SITE are marked with IGNORE .
(For discussion of why the ELLIPSIS SITE has the form that it does in (54), see the next section.)

5 CORRELATE

Not all sluices have overt Correlates. (55a) has one (*one of these terms*); (55b) does not.

- (55) a. We need to choose **one of these terms** ; it does n't matter which one .
b. He finished on time , but it 's not clear with whose help .

In this section we deal with the cases in which a **CORRELATE** can be identified. We develop guidelines about how to identify a **CORRELATE** if there is one and how to categorize it. We return to cases like (55b), in which there is no **CORRELATE**, in [section 7](#).

In the simplest cases a **CORRELATE** picks out the material in the **ANTECEDENT** that correlates with, or corresponds to, the remnant WH-phrase. To find the **CORRELATE** in such cases, locate the overt stretch of material in the **ANTECEDENT** that the sluice WH-phrase intuitively substitutes for in the **ELLIPSIS SITE**. The **CORRELATE** should be a phrasal constituent (e.g., a DP, PP, AP etc.) and in these cases (like (55a)) it corresponds to an expression of the same semantic type as the WH-phrase.

There is another set of cases, however, in which we recognize a **CORRELATE**, but in which the relation between it and the remnant WH-phrase is not one of simple substitution but is rather more complex. In these cases, the WH-phrase builds on, or extends, some overt portion of the **ANTECEDENT** in some way. For example, the possessive WH-phrase *whose* in example (56) elaborates on the indefinite *an abstract*.

- (56) I reviewed **an abstract** but I do n't know **whose** .

In cases like these, we identify the material within the antecedent which is elaborated on, or extended by, the WH-phrase as the **CORRELATE**. It is so marked in (56). (56) is a **Possessor** sluice and in this and in similar cases, the interrogative possessor *whose* applies to *an abstract* in the antecedent clause to fully specify the meaning of the sluiced question. In other, basically similar, cases, the WH-phrase which extends such a **CORRELATE** may correspond to an implicit or adjunct category, such as time, place, or manner. (57) below is such an example—a **Temporal** sluice in this case, in which the **CORRELATE** *as soon as possible* is extended in the sluice by the temporal interrogative phrase *when*.

- (57) U.S. District Judge Vaughn R. Walker said **he will issue his ruling as quickly as possible** , but did not say when. [8340]

We provide three specific guidelines to help you identify **CORRELATES** :

- The **CORRELATE** should be phrasal (in most cases a DP). For example, for partitives like *a lot of* and *a fair amount*, assume that the **CORRELATE** is the full partitive (*a lot of people in San Francisco*, below), not just the embedded nominal (*people in San Francisco*):

(58) There are **a lot of people in San Francisco** like that waiter . I do n't know
how many <there are > , but I do know they all can write . [146456]

- However, in cases of structural ambiguity where there are two candidate phrases for the role of **CORRELATE** (one larger and one smaller) and the two structures are largely synonymous, choose the structure which yields the smaller **CORRELATE** . In (59), for example:

(59) As for any other changes , he said : “ You will see **some movement** within the management teams .
 But it too early to say **what <you will see within the management teams>** . ”
 [65139]

we might take the PP *within the management teams* to be part of a relatively large Correlate (*some movement within the management teams*), or alternatively we might take it to be adjoined at the VP level as a locative PP modifier. For semantic purposes, it makes little difference. In such a case, we take the Correlate to be as small as possible (just *some movement*) and don't include in it the PP modifier.

- Try to find a **CORRELATE** . (60) below is an example where a potential structural ambiguity leads either to a **CORRELATE** or to an **Adjunction** off the **PREDICATE** (on this possibility, see [section 7](#) below).

(60) We will have to do business ; I don't know how much.

Here, it is clear that the Antecedent is *we will have to do business*. What is unclear is whether the WH-phrase of the sluice modifies the VP headed by *do* in the **ELLIPSIS SITE** , as in the structure indicated informally in (61):

(61) I don't know [we will have to [_{VP} [_{VP} do business] how much]]

or if it modifies just the Correlate, as indicated roughly in (62):

(62) I don't know [we will have to [_{VP} do [_{DP} how much business]]].

In such cases, where it is unclear that there is any semantic basis for the choice, we will opt for the structure that favors the existence of the **CORRELATE** , namely, the one where **how much** **extends** **business** , as in (63):

(63) We will have to do **business** ; I don't know how much <we will have to do>.

5.1 Typology of Correlates

In addition to finding the **CORRELATE**, mark what kind it is in **CorrType**, according to the criteria below. Note that some of these possibilities have been claimed not to exist, and so some of our illustrative examples are ill-formed.

- **Def** : A definite description:

(64) A: We should open **the door** . B: Which one?

- **Pro** : a pronominal:

(65) *Joan ate dinner with **her** , but I refused to reveal to the investigators who with.

- **StrgQuant** : A non-indefinite (e.g., *every*, *all*, *most*):

(66) *Joan worked with **every second-year student** , but the letter doesn't say who with.

- **WH** : a WH-phrase:

(67) We know **how many abstracts** each reviewer has read, but not which ones.

- **Name** : A proper name:

(68) *I know that Meg is attracted to **Harry** , but nobody else knows who .

(69) **Sally** was there but I ca n't remember who else .

- **Disj** : A disjoined phrase:

(70) She talked to **either John or Mary** , but I 'm not sure which.

- **Temporal** : A temporal expression:

(71) She left **on Tuesday** , but I 'm not sure when .

- **Locative** : A locative expression:

(72) She lives **in California** , but I 'm not sure where (exactly).

- **WeakQuant** : Phrases introduced by weak quantifiers such as *some*, *one-third*, *many*, *a few*, along with specific indefinites like *a certain N*

(73) They've selected **a candidate** , but they wo n't reveal which one .

- **Other** : None of the above. In one common set of cases for which this is the appropriate annotation the implicit extent or degree of an expression (e.g., a mass noun, or a bare plural, or an adjective) is extended in a *Degree Sluice* involving a WH-phrase like *how many* or *how much* ((74)–(76)). These should all be marked as OTHER **CORRELATE** .

(74) The journey is **quite short** , but we 're not actually sure how short .

(75) We should prepare **food** , but what 's not clear to me is how much .

(76) These things have caused **accidents** , but nobody really knows how many .

Another common case, similar to the one above, is where an overt degree expression bounds the implicit extent or degree of another expression. In such cases, the overt expression counts as the correlate. In the case below, **substantially** is the Other correlate vaguely bounding the implicit degree/extent of the predicate **reduce** .

(77) [75986]The court ordered him to **reduce** the size of the display **substantially** but did not specify to what extent

5.2 Extension of the Correlate

As we saw above, in an important subset of cases, the orphaned WH-phrase of the sluice does not substitute for the **CORRELATE** , but rather it 'extends' the **CORRELATE** . The hallmark of such cases is that the **CORRELATE** makes its contribution as a sub-part of the WH-phrase, restricting its interpretation. Typical examples are given in (78)–(84) below.

- **POSSESSOR**: the correlate is extended with a possessor:

(78) I reviewed **an abstract** but I do n't know whose .

- **FOCUS**: the correlate is extended with a focus particle like *else*:

(79) Sally was there but I do n't know who else .

- DEGREE: the correlate is extended by way of an expression of amount, degree, or extent (such as *how much*, *how far* and so on):

(80) We've made progress , but I 'm not sure how much .

- CATEGORIZING: the correlate is extended with an expression of type or kind:

(81) She bought a new car but I 'm not sure what kind .

- PP: the correlate is extended with a prepositional phrase:

(82) She took a picture but I do n't know what of .

- ADJUNCT: the correlate is extended with an adjunct:

(83) He was wearing a sweater but I do n't remember what color .

- the correlate is extended with a locative or temporal expression which further delimits its denotational range:

(84) She arrived on Tuesday , but I 'm not sure exactly when .

At this point, recognition of the phenomenon of 'extension' matters most for the choices that have to be made about how to identify Correlates and about how the content of the ELLIPSIS SITE is to be filled out (do we or do we not copy the CORRELATE into the ELLIPSIS SITE?). That matter is discussed in some detail in section 6.2.1 below.

Here, no further annotation is required—we do not explicitly annotate extensions of the types just described. We identify Correlates and we let the annotation of the Sluice itself (as described in section 3.3 above) do the work of identifying the extension-types described informally here.

6 The ELLIPSIS SITE

The task of supplying the content of the ELLIPSIS SITE is important and sometimes tricky. As before, we first set out some general guidelines and then examine some specific issues that may arise.

6.1 General Principles

The intuitive content of the ELLIPSIS SITE should be paraphrased inside the Notes section of the SLUICE tag. Since this is free text, it has the greatest potential for generating inter-annotator disagreement, and the discussion below summarizes the main principles to be followed to minimize or eliminate such disagreement. In general, the goal in representing the ellipsis should be to come as close as possible to constructing a TP which maximizes over these constraints:

- **SYNONYMY**: in combination with the WH-phrase of the sluice it conveys the meaning most accurately
- **PARALLELISM**: is maximally similar in structure, word order, and lexical content to the ANTECEDENT
- **BREVITY**: is as short as possible
- **GRAMMATICALITY**: is as close to being well-formed as possible

That is, the goal here is to make the combination of SLUICE and ELLIPSIS SITE together form a reasonable, natural and well-formed WH-question which accurately renders the meaning of the whole construction. That's the ideal; however it will not always be possible to satisfy all of the constraints in such a way that that goal can actually be reached.

To illustrate the issues to be faced, and to anticipate the detailed discussion below, consider five alternative ways of constructing the ELLIPSIS SITE for (85):

- (85) It's reported that they arrested, amazingly, at least three of the dissident students from Santa Barbara, but it's not been revealed which ones.

First, (86) provides the appropriate ELLIPSIS SITE:

- (86) It's reported that they arrested, amazingly, at least three of the dissident students from Santa Barbara, but it's not been revealed which ones <they arrested>

In contrast, (87) clearly makes the Antecedent too large (since *reported* and *revealed* reference the same arresting event):

- (87) It's reported that they arrested, amazingly, at least three of the dissident students from Santa Barbara, but it's not been revealed which ones <it's reported that they arrested>

But there are other ways of filling out the **ELLIPSIS SITE** whose inappropriateness is less obvious. Consider the three partial annotations below for example. In (88) a parenthetical element has been inappropriately copied; in (89) there is failure of parallelism between Ellipsis-Site and Antecedent; in (90) too much of the Correlate has been copied into the **ELLIPSIS SITE**:

- (88) It's reported that they arrested, amazingly, at least three of the dissident students from Santa Barbara, but it's not been revealed **which ones <they arrested, amazingly>**
- (89) It's reported that they arrested, amazingly, at least three of the dissident students from Santa Barbara, but it's not been revealed **which ones <they in fact were>**
- (90) It's reported that they arrested at least three of the dissident students from Santa Barbara, but it's not been revealed **which ones <of the dissident students from Santa Barbara they arrested>**

We will go into detail on all of this in what follows, but we begin with the general observation that the constraints laid out above lead to several simplifying generalizations. **SYNONYMY** and **BREVITY** together will dictate that:

- appositive and parenthetical material in the Antecedent is excluded from the **ELLIPSIS SITE** (see **IGNORE** above and example (88) above).
- smaller Antecedents are preferred to larger ones (see **ANTECEDENT** above and example (87) above)
- material from within the **CORRELATE** is not, in general, copied into the Ellipsis Site (see **SLUICE** above and see **section 6.2.1** below for some exceptions).

But in many interesting cases, the constraints will conflict:

- obeying **SYNONYMY** may create a conflict with the lexical items in the **ANTECEDENT** (thereby violating **PARALLELISM**). This happens, for instance, when the modality in the Ellipsis Site is less determined than in the Antecedent:
 - (91) We have an interest in hiring him; I just don't know **when <we MODAL hire him>** .
 - (92) He expects Foote to return during the series .
Hartley just does n't know **when <Foote MODAL return during the series>** [85457]
- an **ELLIPSIS SITE** that best satisfies **PARALLELISM** may violate **GRAMMATICALITY**
 - the trace of the sluiced WH-phrase may occur inside an island:

- (93) Up on the third floor where he writes , Dubus pulls on a set of untethered earphones to muffle the sound of his children playing below and is hard at work on two pieces , although he will not say
 about what <Dubus is hard at work on two pieces> . [75782]
- copying referential expressions may result in perceived disfluency (as in (94)) or even in ungrammaticality when principles of binding are violated (as in (95) or as in (93)):
- (94) I ’ll fix the car but I do n’t know when <I ’ll fix the car> .
- (95) That student of Mary’s has just won some big award, but she doesn’t know
 which one <that student of Mary’s has just won> .
- the SLUICE may show patterns of pied piping or absence of pied piping that would be impossible in an overt WH-question:
- (96) A: It’s a new medical building.
 B: A medical building owned by who <is it> ?

To resolve such conflicts, we prioritize the constraints in the following order (from most important to least important): SYNONYMY, PARALLELISM, BREVITY, and, last of all, GRAMMATICALITY. We discuss all of this in more detail in the sections that follow.

6.2 The Form of the ELLIPSIS SITE

The ELLIPSIS SITE should be entered as free text into the Notes section of SLUICE . In determining what to copy from the Antecedent, follow these general guidelines:

- In cases in which there is a CORRELATE , the sluice will in general substitute for it and so the phrase you have identified as the CORRELATE should not in general be copied into the ELLIPSIS SITE (there are some important exceptions to this general rule which will be considered in the next subsection).
- Otherwise, try to match the form and content of the ANTECEDENT as much as possible, including syntactic structure (that is, obey PARALLELISM).
- Obey the morpheme boundaries of the parser.
- For the most part you should leave in punctuation (e.g., commas, periods, etc.). See section 6.4 below for more detailed guidelines.

We illustrate these principles and give more detail in the following subsections.

6.2.1 Omit As Much of the Correlate As Possible

In cases like (97a), which involve a **CORRELATE**, do not include the Correlate in the spelling out of the elided material. (97a), that is, should be handled as in (97b):

- (97) a. He criticized someone , but I do n't know who .
b. He criticized someone , but I don't know **who <he criticized>** .

Absence of the Correlate is the principal way in which a typical **ELLIPSIS SITE** will differ systematically from its **ANTECEDENT**.¹

In some cases, however, it will be impossible to omit the **CORRELATE** altogether. This will come up particularly often when the WH-phrase of the sluice **extends** the Correlate; put differently, these are cases in which the content of the Correlate ends up functioning as a subpart of the WH-phrase of the sluice. In such cases, it may be impossible to completely excise the **CORRELATE** while preserving a coherent representation of what the sluice means. A case in point is the example in (98), where the Correlate *a fine* cannot be removed from the **ELLIPSIS SITE**, since it hosts the amount sluice (see also (93) above):

- (98) The company would also face **a fine** under the state charge , although prosecutors did not say **how large <the company MODAL face a fine>** . [61357]

In these cases, remove as much as you can. A good bet is to retain only the head of the **CORRELATE**. We will be adding to this section as problematic cases arise and get resolved, but the ways in which various kinds of extended Correlate will be handled under these guidelines are illustrated below:

- POSSESSOR:

- (99) I reviewed **an abstract** but I do n't know **whose <I reviewed>** .

- FOCUS:

- (100) **Sally** was there but I do n't know **who else <was there>** .

- EXTENT:

- (101) We've made **progress** , but I 'm not sure **how much <we've made>** .

¹We thus omit parts of the **CORRELATE** as well as temporal/locative adverbs (which may not be part of the **ANTECEDENT** to begin with), trusting that both of those are anaphorically resolved as restrictions. There are other kinds of material that could be eliminated in principle (e.g., null complement anaphora as in *he knows that she is here but I don't know why <he knows>*), but we will preserve that content in the ellipsis site.

- KIND:

(102) She bought **a new car** but I'm not sure **what kind <she bought>** .

- PP

(103) She took **a picture** but I do n't know **what of <she took a picture>** .

- ADJUNCT:

(104) He was driving **a minivan** but I do n't remember **what color <he was driving a minivan>** .

(105) The company would also face a fine under the state charge , although prosecutors did not say **how large <the company MODAL face a fine>** . [61357]

- TEMPORAL AND LOCATIVE:

(106) She arrived **on Tuesday** , but I'm not sure **exactly when <she arrived>** .

(107) She lives **in California** , but I do n't remember **exactly where <she lives>** .

We will annotate the various types of extension cataloged above on the **SLUICE** itself (not on the **CORRELATE**). For the details about how to do that, refer to [section 3.3](#) above. Here we are concerned only with the question of how much of the **CORRELATE** to include in the **ELLIPSIS SITE** .

6.2.2 Disjunctive Correlates

which-sluices can be anteceded by disjunctions. In these cases, we mark the disjunction itself as the **CORRELATE** , and the sluice replaces it directly.

(108) -LRB- He 's still registered in Georgia **either** as a Democrat or an independent ; he ca n't remember **which <he 's still registered in Georgia as>** -RBR- [65139]

6.2.3 Avoid Pronominalization

It will often be tempting to violate **PARALLELISM** by using a pronoun in the **ELLIPSIS SITE** rather than repeat a definite description or proper name from the **ANTECEDENT** :

(109) a. I 'll fix the car but I do n't know **when <I 'll fix it>** .
 b. John said Bill was happy and I can guess **why <he is happy>** .

Do not do this. Aim instead for the (more stilted) forms in (110);

- (110) a. I'll fix the car but I do n't know **when <I'll fix the car>** .
b. John said Bill was happy and I can guess **why <Bill is happy>** .

As we have already seen (see the discussion of (94) and (95) above), this is a case in which the goal of constructing a natural-sounding paraphrase of the **ELLIPSIS SITE** takes second place to other goals (in this case **PARALLELISM**).

6.2.4 Islandhood

There is another context (one that is of theoretical importance) in which the drive not to tamper with the **ANTECEDENT** when we copy it into the **ELLIPSIS SITE** leads to 'ungrammatical' or awkward results. The resultant paraphrases may involve violations of islandhood restrictions or similar syntactic conditions on movement. In such cases, we preserve the word order from the **ANTECEDENT** but use the tag **island** in **SLUICE**. Consider the following two examples (the first repeated from (93) above):

- (111) Up on the third floor where he writes , Dubus pulls on a set of untethered earphones to muffle the sound of his children playing below and is hard at work on two pieces , although he will not say **about what <Dubus is hard at work on two pieces>** .
[75782]

In this example, the PP *about what* is extracted out of the DP *two pieces*, violating a restriction on WH-movement. Rather than modifying the **ELLIPSIS SITE** to form a grammatical pre-sluiice, we mark this as an **island** sluiice.

- (112) The company would also face a fine under the state charge , although prosecutors did not say **how large <the company MODAL face a fine>** . [61357]

Similarly, in example (112) above, even though the pre-sluiice would require us to pied-pipe *a fine* to avoid a Left-Branch Extraction, we leave the DP in situ and mark the construction as an **island** sluiice. In general, if you feel the need to change the word order to render a grammatical pre-sluiice (or produce a pseudo-sluiice), you are likely to be dealing with an island-violating sluiice and you should consider marking it as such.

6.3 Dealing with Differences in Form

There are other cases, however, in which we want to allow violations of **PARALLELISM** by permitting differences between **ANTECEDENT** and **ELLIPSIS SITE**. These are cases in which it is not possible to simply transfer the **ANTECEDENT**, un-modified, into the **ELLIPSIS SITE** and get a result which accurately reflects meaning. We have a family of check boxes on **SLUICE** to mark these differences. They are discussed in the subsections below.

6.3.1 Changes_fin

This marks cases where the ANTECEDENT and ELLIPSIS SITE do not match in finiteness. Some examples are given below (all taken from Jason Merchant's 2001 book):

- (113) a. Decorating for the holidays is easy if you know how <to decorate for the holidays>
b. I 'll fix the car if you tell me how <to fix the car> .
c. I ca n't play quarterback . I do n't even know how <to play quarterback> .
d. I remember meeting him , but I do n't remember when <I met him> .
e. John seems to be happy and I can guess why <John is happy> .

6.3.2 Changes_tense

This marks cases where the ANTECEDENT and ELLIPSIS SITE do not match in tense. As far as we understand, these arise because of the precise embedding environments of the ANTECEDENT and the sluice. One case where this may occur is if the Antecedent mentions some future/uncertain event and the predicate embedding the sluice is interpreted as subsequent to that time. Hence, in the example below, the understanding time is taken to be after the endorsement time, and hence to accurately reflect meaning, the ellipsis-site should indicate past tense, despite the present tense of the ANTECEDENT .

- (114) “ If he does n't endorse -LRB- Tuesday -RRB- , I will not understand why <he did n't endorse> , ” veteran Republican consultant Lyn Nofziger said .
[23006]

Another potential case of this may occur across quotation, as below (note that quotations will involve many additional indexical changes; see the following section (6.3.3)):

- (115) He said, “ I am seeing someone now . ” but he would n't say who <he was seeing> .

6.3.3 Changes_indexical

This marks cases where the ANTECEDENT and ELLIPSIS SITE containing corresponding items which do not match in indexical features like person (e.g., shifts from 1st to 3rd person). Example (115), where 1st person changes to 3rd person, also illustrates this guideline. This is a case where SYNONYMY trumps PARALLELISM.

- (116) He said, “ I am seeing someone now . ” but he would n't say who <he was seeing> .

Other cases arise that involve bound variables, such as the following:

- (117) “I believe that I will get better ,” he said .
 “ The hard part is you just do n’t know when <you will get better> . ” [190577]

6.3.4 Changes_pro

This marks cases where the ANTECEDENT and ELLIPSIS SITE differ on whether the subject is overt or covert. Covert subjects include for us cases of control (118), raising (119), and relative clause subjects (120). Note that if the Antecedent is the one with the covert subject, it will also be discontinuous.

- (118) Oz Chairman Robert Kory vowed to push ahead but would not say
 how <Oz Chairman Robert Kory MODAL push ahead> . [30594]
- (119) Once the city and team complete a contract , the Firebirds will begin to raise \$ 9 million , team president Yount said , but he would not say
 how <the Firebirds MODAL raise \$ 9 million> . [88794]
- (120) It is the side that sneaks cigarettes in some secret corner of the ostensibly smoke-free criminal courts building but will not say where <that side sneaks cigarettes> . [96740]

6.3.5 Changes_case

This change involves instances where an expression in the ANTECEDENT differs in case from its correspondent in the ELLIPSIS SITE . The instances we have found involve, e.g., raising-to-object/exceptional case marking mechanisms like that of *him*/*he* below:

- (121) Mauro , who got married again last spring and is expecting twins , said he ’s enjoying private life for a change , but you can expect him to be back in politics .
 He would n’t predict when <he MODAL be back in politics> . [75036]

6.3.6 Changes_other

Use this tag to mark any changes you note between the ANTECEDENT and Ellipsis Site that aren’t otherwise marked (either by one of the check boxes or by a tag like IGNORE). The example below, for instance, is marked with CHANGES_OTHER because it’s not entirely clear that it is a sluice:

- (122) So why are so few women judges ?
 If it ’s not an on-going pattern of gender discrimination , then what <is it> ? [92602]

Another instance includes changes in word order or lexical category, as in the example below:

- (123) She had spraying to do and trees to prune . But first , she had to learn

how <to spray and to prune trees> . [58227]

We also use this to mark syntactic manipulations beyond those we already mark, such as the subject-auxiliary inversion that occurs for root sluices, like that below:

- (124) To which Tom Minnery , vice president of public policy for Focus on the Family , a conservative religious group , said : “ Of course government legislates morality . The only question is : whose morality <does government legislate> ? ” [17435]

6.3.7 NewWords

This is used to mark any linguistic material not in the ANTECEDENT that is inserted into the ELLIPSIS SITE . It is, in effect, the opposite of IGNORE . The need for this often arise in cases where there is some substantial deviation in form between the ANTECEDENT and the ELLIPSIS SITE , such as changes of syntactic category.

- (125) I am ready for a knockout at any time , but I can not predict which round <that knockout MODAL happen> . [109455]

NewWords is frequently found with Changes_other,

- (126) Our local priest – he’s brought an Italian over to paint them.
- What part of Italy <has he brought an Italian from over to paint them> [111]

6.3.8 Modal changes

In one particularly interesting set of cases, in which the Antecedent introduces a proposition with a future or modal component, the ELLIPSIS SITE involves a vague modal interpretation that is hard to resolve in context. In (127) and (128), for example, it is unclear if the best modal would be *will* or *might*.

- (127) We have an interest in hiring him; I just don’t know when <we MODAL hire him> . [25623]
- (128) he expects Foote to return during the series Hartley just does n’t know when <Foote MODAL return during the series> [85457]

In such cases, to make the meaning-shift clear, insert MODAL in the ELLIPSIS SITE in place of whatever modal is present in the ANTECEDENT . Our intention is that the term MODAL can pick out any modal or futurate meaning that seems appropriate, including deontic, teleological, or epistemic uses. Often it’s actually unclear what precise modal meaning is intended in context, and the use of MODAL is designed to be suitably agnostic on such details.

Use MODAL when the meaning is uncertain. In cases with overt modals in the ANTECEDENT, it is often reasonable to copy over that modal. Do this if what results is an appropriate interpretation (one that is not vague or under-specified with respect to modality).

- (129) he 's got to go up and in on him at his chin . I don't know
for what <he's got to go up and in on him at his chin> . [7212]

However, even with overt modals, things are often murky; in such cases, opt for MODAL:

- (130) “ I would n't be out there skating if I did n't think I could return , ” Langenbrunner said .
“ I do n't know when <I MODAL return> , but I 'm doing the things I need to do to try to get back . ” [156]

6.3.9 E-Type interpretations

There is another circumstance in which a difference between ANTECEDENT and ELLIPSIS SITE is warranted—cases which involve so-called ‘E-Type’ interpretations of certain expressions. That matter is important enough, and intricate enough, however, that we have devoted a separate section to it (section 8).

6.4 Punctuation

Sometimes the ANTECEDENT will involve punctuation—commas, periods, colons, semicolons and so on. Leave these intact in the ELLIPSIS SITE for the most part. The single exception is quotation marks, which are almost always best removed.

7 PREDICATE

One of the important issues in discussions of sluicing has been the question of how similar the argument structure of the antecedent clause and the argument structure of the ellipsis site must be. To facilitate investigation of these issues (and also to deal with cases which lack a CORRELATE) we identify and mark for each example the PREDICATE of the antecedent clause. The sections below provide guidance about how to do this. In most cases it will be straightforward.

7.1 Identifying the Predicate

The main predicate should be a semantically meaningful predicate. Ordinarily, this will be the main verb of the clause or clause-like constituent which functions as ANTECEDENT (see section 4). However, there are cases where that verb is either semantically vacuous (e.g., the copula) or too small (in the case of verb particle constructions, say, or phrasal idioms). In general, try to find the overt material that corresponds to the appropriate predicate in a translation to Predicate Calculus. Here are some specific cases.

7.1.1 Copular Constructions

In copular constructions, the main verb is not the predicate, but the following predicate phrase (e.g., nominal, adjective, preposition):

(131) John is on the roof but I do n't know why . [PPpred]

Similarly, for small clauses, the predicate is the rightmost phrasal constituent:

(132) John considered Mary tall , but I do n't know how tall [SCpred]

7.1.2 Larger Predicates

Prepositional phrases should be included with a non-copular main verb in PREDICATE only if they are part of a particle construction (e.g., *look up*). With light verbs, include the nominal/adjectival associates of the light verb (e.g., *take a look*). More generally, idioms (*take a dim view of*) should be considered predicates.

Finally, there will be cases where the predicate is conjoined. If so, mark the conjunction as the predicate, perhaps discontinuously:

(133) They do n't feel like they can put the ball down and do their one-on-one moves.
They have to learn when and how . [24731]

7.1.3 Functional predicates

There will also be cases where the main predicate is actually a functional element of the verbal complex like a modal (*got to*, *below*):

7212 He 's got to go up and in on him at his chin .

I do n't know for what . [7212]

7.1.4 Existential constructions

For existential constructions (*There is/are ...*), we begin by assuming that existentials are of the general form in (134):

(134) There (Aux) be [pivot] [coda],

in which the *pivot* is a DP and the *coda* is a phrase predicated of the pivot—a PP, a VP, an AP, a relative clause, or a null contextual predicate. See the examples below:

- (135) a. There is [a ship] [on the dock] .
b. There is [a ship] [sailing into the harbor] .
c. There is [a ship] [adrift in the harbor]
d. There is [a ship] [that sits on the dock] .
e. There is [a ship] [<context>]

We will assume that the coda is the predicate in existential constructions.

Since in some cases, the coda is established by context alone (as in (135e)) it is sometimes hard to know whether to classify some phrase as a modifier of the pivot or as the pivot. This comes up, for example in cases like the one below, in which one might take *in San Francisco* to be either a modifier of the pivot or as the coda.

146456There are a lot of people in San Francisco like that waiter . I do n't know how many
, but I do know they all can write . [146456]

Knowing how to proceed in such cases isn't always easy If something seems unclear, mark the case as problematic and refer it to us.

8 E-TYPE

Indefinite descriptions (‘weak quantifiers’) like *some people* or *many politicians*, as well as some other kinds of quantificational elements, are often involved in anaphoric relations with pronouns which have what are known as ‘E-type’ interpretations. Typical examples are like those in (136):

- (136) a. A man broke into our apartment. Scott thinks that *he* got in through the bathroom window.
b. Every kid in my class who owns a VCR knows how to program *it*.

Pronouns of this type are neither referential nor bound in the traditional (logical) sense. Their interpretations can be best rendered in terms of paraphrases involving demonstratives or definite articles—like *that man* or *that man who broke into our apartment* in (136a), or *the VCR that she owns* in the case of (136b).

Elements in the ANTECEDENT of a sluicing construction often support E-type readings in the ELLIPSIS SITE. In the example below, for example, the indefinite *someone* is intuitively not copied, but replaced by something definite in the Ellipsis Site:

- (137) I punched a guy last night and he ’s calling me to figure out *why <I punched him>*. [etypeEx]

To determine if an expression is an EType, ask yourself if in the Ellipsis Site you would **have to** refer to it with a pronoun or something anaphoric for the sluice to capture the intended meaning. If so, it’s probably an EType. If not or if you aren’t sure, it’s probably not. When it doubt, assume it isn’t EType.

8.1 Some exemplary cases

8.1.1 WH-Phrases

Consider first examples like those in (138):

- (138) a. I know which students we hired but not why.
b. Who did the suspect call and when?

(138a), in its most salient reading cannot be paraphrased as in (139a), but rather as in (139b):

- (139) a. I don’t know why <we hired which students>.
b. I don’t know why <we hired them>.

Similarly for (138b), which must be interpreted as in (140b), not as in (140a):

- (140) a. Who did the suspect call and when <did he call who>?

- b. Who did the suspect call and when <did he call them>?

That is, the WH-phrase of the Antecedent (*which students we hired* in the case of (139a)) in the **ELLIPSIS SITE** supports an interpretation that can be expressed by way of a plural personal pronoun like *them* (referring to ‘the students that we hired’ in the case of (139b) and ‘the person or people that he called’ in the case of (140b)).

8.1.2 Quantifiers

The same interpretations are found with strong quantifiers, as seen in (141):

- (141) Most faculty members will be at the meeting, but it’s not clear why.

An available interpretation for (141) is that expressed by (142):

- (142) Most faculty members will be at the meeting but it’s not clear why <they will be at the meeting>.

8.1.3 Indefinites

The same phenomenon can be observed in the case of non-WH indefinites, as seen in (143):

- (143) We’re supposed to investigate if a certain priest took bribes, not why.

The salient interpretation of (143) can be paraphrased as in (144):

- (144) ... not why <he took bribes>

Therefore we need to annotate for three types of E-type interpretation within the **ELLIPSIS SITE**. In these cases, mark the E-type expression in the Antecedent with $_E$ and identify the kind of expression it is:

8.2 Type

In addition to marking EType expressions, also indicate the category of the expression:

- WH: Where the EType is a WH-phrase
- QUANT: The EType is a strong quantifier
- INDEF: The EType is a (non-WH) indefinite
- OTHER: The EType is some other category. We have not encountered cases of this category.

8.3 Form in the Ellipsis Site

In filling out the ellipsis, use a personal pronoun to suggest the appropriate interpretation:

- (145) I know **which students** we hired but not **why <we hired them>** . [ETypeWHa]
- (146) **Most faculty members** will be at the meeting but it's not clear **why <they will be at the meeting>** . [ETypeQuant]
- (147) We're supposed to investigate if **a certain official** took bribes, not **why <that official took bribes>** . [ETypeIndef]

8.4 Ambiguity and necessity

It's important that we not overuse the E-type category. That is, you should reinterpret an expression from within **ANTECEDENT** as a personal pronoun in the **ELLIPSIS SITE**, only if you have to do that to render a sensible and appropriate interpretation. In particular, don't convert to a demonstrative when the indefinite in the Antecedent is part of an idiom:

- (148) Word got around that he was to be charged but it's not clear how.

This example should not be rendered:

- (149) Word got around that he was to be charged but it's not clear how <it got around>

but rather as:

- (150) Word got around that he was to be charged but it's not clear how <word got around that he was to be charged>

Similarly in a case like (151):

- (151) Permanent residents have committed a number of crimes in the state, but we don't know how many.

the indefinite 'permanent residents' does not have an E-type interpretation in the **ELLIPSIS SITE** and can be appropriately rendered without such an adjustment; the ellipsis will be: be 'how many <permanent residents have committed>'.

The general guideline is: if the ellipsis has an interpretation which does not involve an E-type reading for an indefinite in the Antecedent, assume that interpretation and annotate accordingly. In other words, use **E-TYPE** only when you have to accurately represent the only, or a salient, interpretation of the sluice.

We are also not interested in things that are already definites in the **ANTECEDENT**: pronouns, definite descriptions, demonstratives and so on. Don't mark these as EType.

9 NOTE

Use this tag to make mention of examples that are interesting, either theoretically or empirically. These are rough and ready evaluative descriptors. Basically, we want a way of flagging things to return to for theoretical analysis. Note that these tags allow us to mark specific spans of text, which is useful for pinpointing issues precisely for review. In that way, they are a bit richer than some of the options on the tags themselves.

Remember to use the `Notes` of `NOTE` to explain why this span is worthy of note.

We have a growing stable of subtypes you can mark:

- `lexicon` : Cases which seem to violate Chung's (2006) generalization that the `ANTECEDENT` and `SLUICE` must be built from the same lexical items. Cases involving `NewWords` will mark this often, but there may be cases outside of that.
- `weird` : This is similar to `problematic` .
- `unexpected` : For material that would be surprising to some theoretical perspective.
- `cool` : Nifty example of some sort.