

Annotating the Implicit Content of Sluices

Pranav Anand & Jim McCloskey

Department of Linguistics
University of California, Santa Cruz
1156 High St., Stevenson Academic
Santa Cruz, CA 95064
{panand, mcclosk}@ucsc.edu

Abstract

This paper reports on an effort to develop a linguistically-informed annotation scheme for *sluicing* (Ross, 1969), ellipsis that leaves behind a *wh*-phrase. We describe a scheme for annotating the elided content, both in terms of a free text representation and its degree of correspondence with its antecedent. We demonstrate that we can achieve reasonable IAA (α between .78 and .88 across eight annotation types) and describe some of the novel patterns that have arisen from this effort.

1 Introduction

Ellipsis is one of the central concerns of modern linguistic theory. Despite its importance, as noted by Bos & Spender (2011), large-scale annotated corpora of elliptical phenomena are rare. Bos & Spender’s own work is part of small group of papers attempting to annotate elliptical phenomena systematically. Much of this work has focused on studying Verb Phrase Ellipsis (VPE), which occurs when a verb phrase is replaced by an auxiliary, as in *I avoided meat, although I didn’t have to <avoid meat>*.¹ Here, we consider sluicing (Ross, 1969), a distinct variety of ellipsis in which all but the interrogative phrase of a content question is elided, leaving behind the **SLUICE**, or *wh*-remnant, subject to an available **ANTECEDENT**:

- (1) It’s clear that
the University has to change, but
in what ways <the University has to
change> is less clear.

¹We follow the convention of indicating the implicit content of ellipsis inside angle brackets.

One of the central debates in the study of ellipsis concerns the various syntactic and semantic mismatches between antecedents and elliptical content, and an animating goal in our research is uncovering a theory-neutral representation of elliptical content that can help sort out the ranges of mismatches. We choose sluicing as our initial target for annotating implicit content for several reasons: it is cross-linguistically common (unlike VPE), it is well studied (which means that we have the makings of a rich annotation system), and it interacts with many other linguistic areas (e.g., the syntax and semantics of questions, discourse dynamics, lexical argument structure).

We describe an effort to extract 4100 sluicing examples from the New York Times subset of the Gigaword Corpus (Graff et al., 2005). We have currently annotated 417 instances in our corpus, and have achieved interannotator α values between .75 and .86 across eight annotators and eight annotation types. We begin in Section 2 with an overview of the theoretical landscape of sluicing and some discussion of previous corpus work. Section 3 lays out our annotation scheme and section 4 provides evaluation of the procedure that led to this scheme. In section 5 we discuss some qualitative observations on the licensing of sluicing that have arisen so far from our annotation. Finally, in section 6 we conclude with areas for future development.

2 Background

2.1 Theoretical Landscape

Following Chung et al. (1995), the literature recognizes two central kinds of sluices. In *merger* sluices (as in (2a)), the antecedent contains a **correlate**

phrase which corresponds to the *wh*-phrase of the sluice. There are also *sprouting* sluices, in which the context contains no correlate, as in (2b).

- (2) a. They've made an offer to **one of the candidates**, but I'm not sure **which one**.
b. They were firing, but **at what** was unclear.

Whether or not the distinction between merger cases and sprouting cases is more than terminological has been a major point of contention: Chung et al. (1995) argue that merger sluices (but not sprouting) are not subject to syntactic island restrictions, a claim Merchant (2001) disputes but which Yoshida et al. (2013) provide experimental evidence for.

At a more basic level, though, the central question in research on sluicing is what, if anything, is the content of the ellipsis site. At one pole, anaphoric theories argue that ellipsis sites have no internal structure, and that resolving elliptical content is a species of anaphora resolution (Hardt, 1993; Darymple et al., 1991; Schieber et al., 1999; Ginzburg and Sag, 2000; Culicover and Jackendoff, 2005; Barker, 2013). At the other, parallelism theories assume that there is syntactic content to ellipsis sites that is somehow parallel to (or recycled from) the linguistic structure of the antecedent (Williams, 1977; Fiengo and May, 1994; Chung et al., 1995; Ross, 1969; Merchant, 2001; Craenenbroeck, 2010). While originally it was thought that parallelism should be defined in purely semantic terms, evidence has steadily accumulated that the availability of sluicing is sensitive to the morphosyntactic structure of the antecedent. First, unlike VPE (Kehler, 2002), sluicing does not tolerate voice mismatches (Merchant, 2001; Chung, 2005; Chung et al., 2011; Anderbois, 2010; Chung, 2013; Merchant, 2007):

- (3) a. The candidate was abducted but we don't know who by/by who.
b. Somebody abducted the candidate, but we don't know by who *(he was abducted).

Similarly, bare nominal *wh*-phrases cannot be sluiced in certain cases in which the antecedent clause lacks a crucial preposition (Chung, 2005):

- (4) a. They're jealous but it's unclear who *(of).

- b. Last night he was very afraid, but he couldn't tell us what *(of).

Nevertheless, the morphosyntactic requirements for parallelism are not absolute, allowing at least for mismatches in finiteness or syntactic category like those below (Merchant, 2001):

- (5) a. I can't play quarterback; I don't know how.
b. I remember meeting him but I don't remember when.

This conundrum — the simultaneous sensitivity of parallelism to fine-grained lexical and syntactic structure, alongside its blindness to finiteness or lexical category — highlights how little we still know about the range of potential mismatches. In our research, we aimed to create an annotation scheme that would allow us to bring to light the full variation permitted.

2.2 Related Work

As far as we know, there are precisely seven systematic corpus annotations of ellipsis, four focusing on verb phrase ellipsis (essentially, VPE and a handful of similar verbal processes, like pseudogapping and comparative deletion) (Hardt, 1997; Nielsen, 2005; Bos and Spénader, 2011; Shahabi and Baptista, 2012) and three on sluicing (Fernández et al., 2005; Beecher, 2008; Nykiel, 2010).

The first large-scale study of verbal ellipsis is due to Hardt (1997). 644 cases of VPE were extracted from the Penn Treebank, whose antecedents were then annotated by two coders. Hardt estimates that the tree patterns he looks for have a recall of less than 50%. As a result, two subsequent corpus-driven efforts have involved significant manual examination. Nielsen (2005) read through one million words across two corpora (444K words from the BNC, 680K words from the Penn Treebank), and uncovered 1510 instances of VPE. In addition to coding VPE antecedents, he provides text corresponding to an intuitive paraphrase of the ellipsis site and classifies the kind of mismatch between the antecedent and paraphrase according to thirteen criteria (e.g., tense mismatch, comparatives, inversion, split antecedents, inferred antecedent). In a similar effort, Bos & Spénader (2011) examined the entire WSJ portion of the Penn Treebank, focusing on modals

and auxiliaries that “trigger” VPE. They find 580 instances of VPE and related phenomena, which they code for antecedent as well as: the morphosyntactic category of the antecedent, the trigger, and 34 strings connecting the antecedent and elision site. The bilingual VPE corpus of Shahabi & Baptista (2012) is markedly different from the three efforts already mentioned. They examine the Tehran English Persian Parallel Corpus (Pilevar-Taher et al., 2011), an automatically aligned English-to-Persian parallel corpus drawn from Open-subtitles that comprises 3.7 million words in each language. Using a trigger-based search like Bos & Spenader, they find 10,515 instances of VPE in English; they then show that one can straightforwardly quantify the relative poverty of verbal elliptical processes in Persian by determining how many VPE cases are resolved in Persian.

In the case of sluicing, there are three principal efforts, all with very particular and divergent aims. Nykiel (2010), for example, is interested in tracing the relative rates of sprouting and merger in 1689 sluices across five eras of English, from Old English to Present Day English. Beecher (2008) focuses on the particular question of which prepositions support swiping (sluicing in which the *wh*-expression and a preposition undergo inversion, e.g., *by who*). Using a list of ten question embedding predicates and 38 prepositions from the OED, he uses the Google Search API to extract expressions of the form “predicate *wh/what* P”, which he then culls to 3000 sluices. Finally, Fernandez et al. (2005) focuses on ‘root’ sluices that are isolated sentences (e.g. *Who? Why?*). Using regular expressions, they extract 5343 root sluices from the BNC, which the authors then annotate for antecedent and sluice type, inspired by Ginzburg and Sag (2000): those asking about an indefinite correlate, those requesting clarification on a presupposition, and statements of general confusion.

What should emerge from this overview is that while there is clearly important antecedent work in this area, the kind of systematic, exhaustive corpus we intend here is novel. Consider the issue of representation. All of the corpora above mark the antecedent and ellipsis site, but the ways they relate the two, if at all, are idiosyncratic. Both Nykiel and Fernandez et al. classify how the sluice *wh*-

expression integrates with the antecedent, but neither of them provides a way of locating other potential (mis)matches. Nielsen additionally provides a text-based resolution and a category for the kind of mismatch, but the categories are quite broad and designed to be mutually exclusive. In addition, as Nielsen alone annotated these sluices, it is unclear whether resolving ellipsis sites in plain text can be done reliably across several annotators. Our goal, in some sense, is to unify all of these efforts.

3 Annotation Scheme Development

3.1 Introduction

The central research questions of this project are the representation schema we will use for resolving sluices and how we will notate mismatch. The representation schema is a tricky eye to thread. On the one hand, as we have seen, the range of representation assumptions is fairly broad. Bos & Spenader notably refrain from following Nielsen in resolving the ellipsis site, precisely because of the theoretical commitments that any choice brings. However, choosing not to resolve in turn means that one cannot catalog mismatches. Instead, our aim is to adopt the minimal representational commitments we must in order to document mismatches.

3.2 Data Selection

Our data comes from the New York Times subset of the English Gigaword Second Edition corpus (Graff et al., 2005). We first parsed the subset with the Stanford parser and then extracted all verb phrases whose final child was a *wh*-phrase. This yielded 5100 verb phrases. One author manually culled this to 4100 sluices (eliminated expressions were 40% idioms, 40% parsing errors, 15% repetitions we could not remove automatically, and 5% sluicing-like constructions we put aside for the moment). As a final quality check, the other author manually examined all 52,000 *wh*-phrases in a random 80th of the NYT subcorpus and discovered only one additional sluice.

Table 1 shows the distribution of the extracted sluices by embedding predicate and *wh*-remnant; for clarity, we only break out the top 7 remnants (95% of data) and top 8 predicates (80% of data). While *why* sluices are 44% of the data, somewhat surprisingly, 20% of the data came from degree sluices (*I know*

	<i>oth.</i>	which	where	what	when	how	how much	why	
<i>oth.</i>	58	40	50	67	70	75	132	250	742
figure	1		1			14	1	73	90
ask	4		3	1	1	6	9	79	103
specify	7	21	1	1	13	16	54	5	118
explain	5			1		10	1	189	206
understand	4					5	2	211	222
see	18		2	2		37	3	181	243
say	84	44	49	15	123	47	387	116	865
know	102	33	45	115	146	161	218	728	1548
	283	138	151	202	353	371	807	1832	4137

Table 1: Distribution of Sluices by Embedding Predicate and *wh*-remnant. *Oth.* designates all predicate or remnant types not listed.

he’s hurt, but I don’t know how bad.). As we discuss in section 5.3, these proved particularly challenging to annotate.

3.3 Scheme Development Procedure

Our annotation scheme was developed on 417 sluice instances over seven rounds of annotation and discussion. Sampling was biased to encourage diversity in *wh*-remnant type: we chose 50 examples randomly from each of the top seven remnant categories (*why*, *how much*, *how*, *when*, *what*, *where*, and *which*; see Table 1 for frequency breakdowns) and 67 randomly from the remaining data. In the first round, the authors first collaboratively annotated 4 sluices chosen for diversity of *wh*-remnant (*why*, *what kind*, *how much*, *what color*) and constructed an initial scheme. In addition to identifying the antecedent, like Nielsen, we resolved the ellipsis site with plain text. We also constructed taxonomies for the types of mismatch, the kind of implicit argument in cases of sprouting, and, in the case of merger, the varieties of correlates. We found that a context window radius of five sentences was sufficient to perform these tasks; crucially, even when the antecedent was nearby, determining the proper antecedent scope and ellipsis resolution often involved understanding the larger questions under discussion in the text. We then each annotated 33 sluices, and adjusted the taxonomies. For the remaining rounds, we recruited six annotators: five advanced undergraduate linguistics students (all with at least two courses in syntax and semantics) and one graduate linguistics student. All eight of us then annotated, in sequence, 40 sluices, followed by two additional rounds of 100 sluices, and

one round of 140 sluices. We met weekly to compare and discuss problematic cases, revising the annotation scheme and reannotating all previous material. By round 5, annotators reported being able to annotate 15-20 annotations per hour. Although we considered using the automatic parses in annotation, we found the parsetrees too error-prone to adequately help with the fine-grained constituency analysis we required and elected to use text spans alone.

Annotation was conducted on a modified version of the brat web-based annotation tool (Stenetorp et al., 2012). Existing tools render the annotation of elided content difficult, since those that allow insertion of new markables (e.g., MMAX2 (Mueller and Strube, 2006)) completely alter the document, making inter-annotator comparison difficult. We have minimally modified brat to accept and display a free text paraphrase, but we aim in subsequent versions of this project to allow it to accept new content that can be further annotated as well (i.e. for mismatches with the antecedent).

3.4 Final Annotation Scheme

Our current annotation scheme codebook and a sample of our gold standard annotations in stand-off annotation format are available at <http://ohlone.ucsc.edu/SCEC> for browsing. Each sluice example is annotated with four obligatory tags: the **antecedent**, the **sluiced expression** – including a plain-text paraphrase of the elided content – the **main predicate** of the antecedent clause, and the **correlate**, if there is one. The correlate and sluice are also tagged with the taxonomic

MANDATORY TAGS

SLUICE: sluice site.

- **TEXT**: Free text paraphrase of elided material
- **TYPE** [Degree, Manner, Reason, Temporal, Locative, Classificatory, Possessive, Passive, PP, Focus, Other]
- **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 **A**.

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

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

OPTIONAL TAGS

ELLIPSIS ANTECEDENT: **A** is elided

E-TYPE: Indefinite in **A** that is anaphoric in **ES**

ALTERNATIVE ANTECEDENT: Secondary **A**

IGNORE: Material not retained in **ES**

Figure 1: Abridged Sluicing 1.5 Tagset

features mentioned above (type of sluice, type of correlate, and morphosyntactic mismatches). Figure 1 summarizes these features.

In addition, each sluice example may additionally bear six optional tags. Two correspond to cases where there are several possible antecedents. In the case of **Alternative Antecedent** we observed several cases of antecedent “sandwiching”, in which the sluice is buttressed by roughly synonymous potential antecedents, as in (6). **Ellipsis Antecedent** is used in cases where the antecedent for a sluice is itself elliptical (in all cases we have encountered, VPE).

(6) We lost our focus a little bit somewhere. I don’t know where. But **we lost it**. [27861]

Two additional tags deal with interpretive differences between Antecedent and elided content. **EType** marks indefinite material in the Antecedent that is interpreted anaphorically in the ellipsis site, as in (7). **Ignore** marks material that is semantically active in the Antecedent but does not seem to be carried over to the elided content at all, such as parenthetical material (8a) or additive particles (8b).

(7) She said that she would issue **a written ruling** as soon as possible, but did not say when. [35291]

- (8) a. **First, though**, they must teach. And, before that, figure out how. [36311]
 b. He said McDonald **also** owed federal taxes, but he would not say how much. [5912]

4 Analysis of Annotation Scheme Development

Table 2 provides a condensed measure of interannotator agreement over the tags across the rounds.² Because all of the tags are text spans, we use Krippendorff’s continuum metric (Krippendorff, 1995) (a special case of Krippendorff’s α (Krippendorff, 2014) for spans). In general, IAA rates drop in Round 3, as the additional annotators were introduced, and then rises.

Most of the agreement gains come from conventions about boundaries (e.g., when ignored material at clause-edge should be marked Ignore vs. excluded from the Antecedent, what the predicates of copula and existential sentences are). In addition, the gains for Antecedent in Round 5 are largely due to the introduction of the Elided and Alternative Antecedent tags, which served to resolve a disagreement about what ‘the’ antecedent was in such unclear

²Note that the IAA rates have been computed for the novel instances in each round.

cases. EType’s rise involved actual instruction of the annotators about the pragmatics of EType interpretations. Finally, Correlate increases are due both to implicit learning (e.g., what counted as the “real” correlate in an expression), but also due to a growing insight on our part about the complexity of degree sluices (see section 5.3). Agreement on the taxonomic features on Sluice and Correlate, not shown here for reasons of space, were consistently above 95% accuracy.

Tag	Round				
	2	3	4	5	6
Sluice	.83	.75	.78	.88	.86
Ante	.83	.67	.73	.78	.88
Pred	.92	.56	.85	.85	.85
Corr	.72	.58	.60	.74	.78
Elided				.94	.94
AltAnte				.66	.78
EType	.21	.32	.67	.80	.87
Ignore			.43	.74	.78
Text	62.4	48.2	50.4	84.2	84.2
Instances	33	40	100	100	140

Table 2: Inter-Annotator Agreement by Annotation Round. IAA for the first 8 span categories is calculated in Krippendorff’s continuum metric and IAA for the free text paraphrases is in BLEU:3. Numbers are computed for new instances annotated in each round, which is provided at the bottom of the table.

4.1 Minimal Tampering and Maximal Omission

A significant portion of our discussions focused on the procedure for resolving the elided content. We found that many of the mismatch types were only clearly apparent on comparison of the free text paraphrase with the antecedent. However, the fact that paraphrases were free text gave annotators a great deal of latitude to modify the form of the antecedent – e.g., introducing an embedding predicate to preserve finiteness or paraphrasing away material to circumvent an island violating structure.

Two best practices arose during the process that increased consistency. First, we adopted a principle of “Minimal Tampering”, where annotators were asked to modify the Antecedent text minimally; this

was most successful after Round 3, where annotators were given the ability to alter a copy of the Antecedent (as opposed to constructing a paraphrase *de novo*). However, these paraphrases were often unnatural and prolix, because letter of the law Minimal Tampering required an annotator to overtly express material that is more naturally dropped in a typical conversational setting. For example, consider the temporal adjunct *Thursday* in (9a) and the locative adjunct *in the region* in (9b). Should these be explicitly mentioned, and if so, how should the paraphrase be structured (e.g., where should *in the region* go? with the *wh*-remnant or in its original location in the Antecedent?). Similarly, in (9c), the DP *thousands upon thousands of people* is an EType expression. Should that be expressed in the free-text paraphrase as *them, those people, those thousands upon thousands of people*?

- (9) a. But Thursday the market for other California municipal bonds recovered a bit. “It’s difficult to say how much, because liquidity is relatively low and trading is sporadic,” said Ian MacKinnon, senior vice president of fixed-income for the Vanguard Group of mutual funds. [35463]
- b. Among the proposals are new power plants in the region, although the report does not specify where. [143606]
- c. There was always something new improved equipment, innovative means of transmission, original shows coming down the network line from New York and Chicago and above all, the knowledge that thousands upon thousands of people clustered around a box that sat like a shrine in their living rooms, listening. It didn’t really matter to what. [36225]

We adopted Minimal Tampering in part to make links between the Antecedent and ellipsis more automatically recoverable, but after several rounds of unsuccessful additional conventions, we realized by round 5 that a more anaphorically reasonable approach was easier for annotators to reliably implement. We thus introduced a principle of ‘Maximal Correlate Omission’, which instructed annotators to preserve as little of the Correlate as they could. In the

end, this meant that many of the stylistic differences in this kind of redundant content were removed. Correspondingly, there is a spike in agreement rates for Text in Table 2 after round 5 (IAA for paraphrases is provided in BLEU:3 score (Papineni et al., 2002)).

4.2 Unresolved issues

Two issues proved too difficult to annotate reliably. First, because there is controversy in the literature about whether sprouting occurs with ‘core’ arguments or only adjuncts, we attempted in Round 3 to mark cases of sprouting with their FrameNet roles. However, this task proved too costly for the annotators; fully 30% of the predicates we considered lacked a clear FrameNet entry, and for the remainder, it was often unclear which frame was best suited to the data.³ This led us to adopt the streamlined sluice TYPE shown in Table 1. In addition, *wh*-remnants that coordinated phrases with distinct types and/or grammatical functions proved too challenging for us to annotate with current tools, since they interacted with the Antecedent in different ways. For example, in (10), the phrases link to different Correlates: *how many* picks up on the amount introduced by the vague partitive *a bunch* and *whom* targets the quantificational DP itself.

- (10) To those who have faulted him for not lobbying aggressively for permanent trade relations for China , he said he had called “a bunch” of members of Congress , but would not say how many or whom . [89868]

5 Qualitative Results

Even though our current set of annotated examples is 10% of our extracted data, we are encouraged by the fact we have already encountered phenomena of real theoretical interest, but which one might have feared would be relatively rare – amnestied island-violations, for instance, as in (11) (note that the

³An anonymous reviewer asks why we chose FrameNet over Propbank, which is considerably less articulated. As our initial intent was to characterize precisely what the role was, not simply whether it was core, we believed that FrameNet’s specificity would be a benefit. The reviewer is right that Propbank may be good enough for the core-distinction, and we plan on following up on this idea.

elided content is ungrammatical, as expected if this is an island amelioration):

- (11) The handover took place at a British embassy in one of the newly independent Baltic states. Which one <the handover took place at a British embassy in> has never been confirmed.

In particular, several kinds of mismatch between antecedent and ellipsis site have turned up which have gone undiscussed or underdiscussed in previous work. Here we offer some examples, as an illustration of the potential for discovery that we think our resource holds out.

5.1 Modal mismatches

Since Merchant (2001), it has been known that a finite clause can antecede a nonfinite sluice, triggering attendant realis differences, as in (5a) above. But we have also found many (40) examples of the reverse pattern, where a non-finite (or modal) antecedes a sluice. In 30 of these cases, the precise modality intended inside the sluice is difficult to pin down. In (12), for example, is the intended modal here a simple future, or a future-oriented modal (if so, of what flavor?)? For the moment, we are simply annotating these cases with the expression MODAL, but our eventual goal is to understand why this previously unnoticed kind of vagueness is tolerated in sluicing.

- (12) “I want to return (to Peru) some day , but I don’t know when < I MODAL return to Peru> ... ” [117524]
- (13) Texas A&M coach Tony Barone unabashedly predicted that ... the Aggies could be better than a year ago. He just forgot to say when <the Aggies MODAL be better than a year ago>. [88489]

5.2 Compound Correlates

Several of our novel phenomena emerged originally as cases of annotator confusion, including the following:

- (14) Despite my inclination toward procrastination, I am determined to send holiday cards this year. It doesn’t much matter which holiday. [106579]

This example emerged as a problem during annotation precisely because it is unclear what the shape of the analysis is—what the elided content is, what the Antecedent is, how they correspond—and yet all annotators agreed it is grammatical. Three analyses of the elided content are possible: that the *wh*-remnant is sprouted off *holiday cards*; that it is extracted from the compound nominal *holiday cards*, violating numerous constraints on extraction; or that it is extracted from an elided cleft ‘pseudo-sluiice’ (as in (15c)).

- (15) a. It doesn’t much matter which holiday <I send holiday cards for>
 b. It doesn’t much matter which holiday <I send [__ cards]>
 c. It doesn’t much matter which holiday <it is that I send holiday cards for>

Of these options, both the sprouting and compound nominal cases are empirically novel. If sprouting, it should be as ill-formed as **They’re jealous but it’s unclear who*. If the compound analysis is correct, there are issues for the analysis both of compounds and of correlates.

5.3 Degree Expressions

Among our most vexing (and interesting) cases for annotation were degree sluices, underdiscussed in the theoretical literature, but very common in our data. A degree *wh*-remnant (like *how much*) may have no overt Correlate, as in (16), or may have as correlate a vague indefinite extent, as in (17).

- (16) a. They said this would save the government money, though they could not yet say how much <this would save the government money>. [2753]
 b. The review, Gilligan acknowledged, delayed the issuance of the notice about Strandflex, but she said she could not estimate by how much <the review delayed the issuance of the notice about Strandflex>. [60122]
- (17) a. The Atlanta-based company said Thursday that operating profit would be “substantially below” analysts’ estimates but didn’t specify how much <operating profit

would be below analysts’ estimates>. [104088]

- b. 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 ... ” [35463]

For our annotators, the question was: what is the correlate in cases like (17)? The apparent answer is that the correlates are the vague indefinite extent expressions *substantially* and *a bit*. But these elements are optional and in their absence sluicing with *how much* remains possible, much as in (16b). But that in turn suggests that the ‘real’ correlates for such cases are not *substantially* or *a bit*, but rather implicit degree expressions which are further restricted by *substantially* or *a bit*. However, if all of that is reasonable, it suggests an account for cases like (16) in which there are also implicit degree correlates—over extents saved, or delayed by.

There is a practical question of annotation here. But as is often the case, annotation dilemmas highlight theoretical puzzles. Cases like those in (16) would naturally be taken to be sprouting cases, while those in (17), because there is an overt indefinite, would naturally be taken to be cases of merger. But that bifurcation obscures important (semantic) commonalities between the two kinds of cases, and suggests once more how useful sluicing can be as a probe for implicit content. And since such cases suggest that at least some apparent cases of sprouting need to be analyzed in terms of implicit correlates, they force the question again of whether or not such interpretations are generally correct—a position which would in turn have important ramifications for theories of implicit content more generally. Vexation for annotators often signals phenomena of particular theoretical interest.

6 Conclusion

In this paper, we have presented a novel, linguistically-informed annotation scheme for tackling the elided content of sluices and have shown that the system can produce annotations with a high degree of reliability. We have also demonstrated that even in the small amount of data we have examined, patterns outside those traditionally

talked about are already cropping up. Looking ahead, one crucial question we are still considering is the representational schema for elided content. One key limitation of our present toolkit is the inability to mark correspondences between parts of the overt text and parts of the (annotator-generated) elided content. This has made the annotation of, for example, coordinated sluices, impossible and many other tasks cumbersome. In the future, we plan on adapting brat to allow us to relate parts of the Antecedent and elided content directly, building something akin to a word alignment corpus for ellipsis. Such a method could prove both powerful and reasonably theory-neutral across a range of elliptical constructions. We also are considering incorporating further syntactic and semantic annotation (e.g, lightweight syntactic or semantic dependencies) as an additional layer of representation that can be marshaled to (in)validate various theories of sluicing and ellipsis more generally.

Acknowledgments

We would like to first thank our three anonymous reviewers for helpful comments. We have benefited from conversations with: Sandy Chung, Amy Rose Deal, Dan Hardt, Bill Ladusaw, Jason Merchant, Craige Roberts, Bern Samko, and Matt Wagers. This work has been funded by a UCSC Institute for Humanities Research Grant to the Santa Cruz Ellipsis Consortium and a UCSC Committee on Research grant to Jim McCloskey. Finally, this work would not have been possible without the diligent, incisive labor of annotation-wrangler Bern Samko and our undergraduate annotators at various stages of this project: Philip King, Chelsea Miller, Emma Peoples, Nick Primrose, Michael Titone, Danny de la Vega, Tony Zavala, Jasmine Embry, Jack Haskins, Lily Ng and Rachelle Boyson.

References

Scott Anderbois. 2010. Sluicing as anaphora to issues. Presented at SALT 20, University of British Columbia and Simon Fraser University, April 29–May 1, 2010.

Chris Barker. 2013. Scopability and sluicing. *Linguistics and Philosophy*, 36:187–223.

Henry Beecher. 2008. Pramatic inference in the interpretation of sluiced Prepositional Phrases. In *San Diego Linguistic Papers*, volume 3, pages 2–10. Department of Linguistics, UCSD, La Jolla, California.

Johan Bos and Jennifer Spender. 2011. An annotated corpus for the analysis of vp ellipsis. *Language Resources and Evaluation*, 45(4):463–494.

Sandra Chung, William Ladusaw, and James McCloskey. 1995. Sluicing and logical form. *Natural Language Semantics*, 3:239–282.

Sandra Chung, William Ladusaw, and James McCloskey. 2011. Sluicing(:) between structure and inference. In Rodrigo Gutiérrez-Bravo, Line Mikkelsen, and Eric Potsdam, editors, *Representing Language: Essays in Honor of Judith Aissen*, pages 31–50. California Digital Library eScholarship Repository. Linguistic Research Center, University of California Santa Cruz, Santa Cruz, California.

Sandra Chung. 2005. Sluicing and the lexicon: The point of no return. In Rebecca Cover and Yuni Kim, editors, *BLS 31, Proceedings of the Thirty-First Annual Meeting of the Berkeley Linguistics Society*, pages 73–91. Department of Linguistics, UC Berkeley, Berkeley, Calif.

Sandra Chung. 2013. Syntactic identity in sluicing: How much and why. *Linguistic Inquiry*, 44:1–44.

Jeroen van Craenenbroeck. 2010. *The Syntax of Ellipsis: Evidence from Dutch Dialects*. Oxford University Press, Oxford.

Peter Culicover and Ray Jackendoff. 2005. *Simpler Syntax*. Oxford University Press, Oxford and New York.

Mary Darymple, Stuart M. Schieber, and Fernanda C. N. Pereira. 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy*, 14:399–452.

Raquel Fernández, Jonathan Ginzburg, and Shalom Lappin. 2005. Automatic bare sluice disambiguation in dialogue. In *Proceedings of the IWCS-6 (Sixth International Workshop on Computational Semantics)*, pages 115–127, Tilburg, the Netherlands, January. Available at: http://www.dcs.kcl.ac.uk/staff/lappin/recent_papers_index.html.

Robert Fiengo and Robert May. 1994. *Indices and Identity*. MIT Press, Cambridge, Mass.

Jonathan Ginzburg and Ivan Sag. 2000. *Interrogative Investigations: The Form, Meaning and Use of English Interrogatives*. CSLI Publications, Stanford, Calif.

David Graff, Junbo Kong, Ke Chen, and Kazuaki Maeda. 2005. English gigaword second edition ldc2007t07. Technical report, Linguistic Data Consortium, Philadelphia.

Daniel Hardt. 1993. *Verb Phrase Ellipsis: Form, Meaning and Processing*. Ph.D. thesis, University of Pennsylvania.

- Daniel Hardt. 1997. An empirical approach to vp ellipsis. *Computational Linguistics*, 23(4):525–541.
- Andrew Kehler. 2002. *Coherence in Discourse*. CSLI Publications, Stanford, Calif.
- Klaus Krippendorff. 1995. On the reliability of unitizing continuous data. In P. V. Marsden, editor, *Sociological Methodology 1995*, volume 25. Blackwell, Cambridge, Massachusetts.
- Klaus Krippendorff. 2014. *Content Analysis: An Introduction to its Methodology*. Sage Publications, Thousand Oaks, California, 3rd edition.
- Jason Merchant. 2001. *The syntax of silence: sluicing, islands, and the theory of ellipsis*. Oxford University Press, Oxford and New York.
- Jason Merchant. 2007. Voice and ellipsis. Manuscript, University of Chicago, available at <http://home.uchicago.edu/~merchant/publications.html>.
- Christoph Mueller and Michael Strube. 2006. Multi-level annotation of linguistic data with mmax2. In *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*.
- Leif Arda Nielsen. 2005. A corpus-based study of verb phrase ellipsis identification and resolution.
- Johanna Nykiel. 2010. Whatever happened to Old English sluicing. In Robert A. Cloutier, Anne Marie Hamilton-Brehm, and Jr. William A. Kretzschmar, editors, *Studies in the History of the English Language V: Variation and Change in English Grammar and Lexicon: Contemporary Approaches*, pages 37–59. Walter de Gruyter.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. pages 311–318.
- Mohammad Pilevar-Taher, Hesham Faili, and Abdol-Hamid Pilevar. 2011. TEP: Tehran English-Persian parallel corpus. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, volume 6609 of *Lecture Notes in Computer Science*, pages 68–79. Springer Berlin Heidelberg.
- John R. Ross. 1969. Guess who? In Robert Binnick, Alice Davison, Georgia Green, and Jerry Morgan, editors, *CLS 5: Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*, pages 252–286. Chicago Linguistic Society, Chicago, Illinois.
- Stuart Schieber, Fernanda Pereira, and Mary Dalrymple. 1999. Interaction of scope and ellipsis. In Shalom Lappin and Elabbas Benmamoun, editors, *Fragments: Studies in Ellipsis and Gapping*, pages 8–31. Oxford University Press, Oxford.
- Mitra Shahabi and Jorge Baptista. 2012. A corpus-based translation study on English-Persian verb phrase ellipsis. *ICAME Journal*, 36:95–112.
- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun’ichi Tsujii. 2012. brat: a web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations Session at EACL 2012*. Project Website: <http://brat.nlplab.org/about.html>.
- Edwin Williams. 1977. Discourse and logical form. *Linguistic Inquiry*, 8:101–139.
- Masaya Yoshida, Jiyeon Lee, and Michael Walsh Dickey. 2013. The island (in)sensitivity of sluicing and sprouting. In Jon Sprouse and Norbert Hornstein, editors, *Experimental Syntax and Island Effects*, pages 360–376. Cambridge University Press, Cambridge and New York.