Proceedings of SALT 25: 18–37, 2015

On double access, cessation and parentheticality *

Daniel Altshuler
Heinrich-Heine-Universität Düsseldorf

Valentine Hacquard
University of Maryland

Thomas Roberts
University of Maryland

Aaron Steven White
University of Maryland

Abstract Arguably the biggest challenge in analyzing English tense is to account for the *double access* interpretation, which arises when a present tensed verb is embedded under a past attitude—e.g., *John said that Mary is pregnant*. Present-under-past does not always result in a felicitous utterance, however—cf. #*John believed that Mary is pregnant*. While such oddity has been noted, the contrast has never been explained. In fact, English grammars and manuals generally prohibit present-under-past. Work on double access, on the other hand, has either disregarded the oddity (e.g., Abusch 1997: 39) or treated it as a reflex of a particular dialect (e.g., Kratzer 1998: 14). The goal of the paper is to argue—based on a corpus study—that a present-under-past sentence is grammatical, but modulated by two, interacting pragmatic phenomena: *cessation* and *parentheticality*.

Keywords: tense, parentheticals, events, states, attitude reports, indirect speech, corpus

1 Introduction

Imagine the scenario below:

(1) a. *Conversation at 10am in the mall*
Sue to John: How is Mary today?
John to Sue: Mary is sick.
b. *Conversation at 3pm on the same day at the beach*
Bill to Sue: How is Mary today?
Sue to Bill: **John said that she is sick**.

The bolded sentence in (1b) is a prototypical case of a speech report, whereby Sue reports on a previous utterance by John. Although it seems quite easy to describe the

* We thank the SALT reviewers and participants for insightful questions and discussion. Thanks also to Zsófia Gyarmathy for data judgments and helpful discussion about the event/state distinction. This work was supported in part by NSF grant BCS-1124338. Corpus data were originally collected for work supported by NSF grant DDRIG-1456013.

context which warrants the use of such a report, it turns out to be incredibly difficult to state its truth-conditions.

The first complication concerns the intuition that the truth of the report is correlated with the time of the matrix verb (in this case, John's saying time). The problem is that—unlike what our intuitions lead us to believe—there is, in fact, no direct link between the time of the state described in the complement clause and the time of the matrix verb (von Stechow 1982, 1995). We see this in situations where the subject of the matrix class is mistaken about the time. While we don't discuss such examples here, we note that to account for such cases it's important that the embedded tense refer to the time that the attitude holder locates himself (at the time his attitude was expressed), rather than to the actual time at which his attitude was expressed. In other words, we interpret the present tense in the complement of a verb like *say* to be related to the attitude holder's now, rather than to the time of saying.

The second complication concerns the utterance time. As observed by Abusch (1997: 40), (2) below can be true even if Mary was never pregnant in the real world. For this reason (3) is not contradictory:

(2) John said that Mary is pregnant.

(3) John said that Mary is pregnant but she is simply overeating.

In light of these data, Abusch concludes that whatever the correct semantic analysis of (2) is, it should not entail actual pregnancy of Mary (in the past, present or future). Moreover, note that (2) is not entirely a description of the world *according to John*. When he made a claim about Mary, John was not making a prediction about the time that (2) would be uttered—i.e. a time that would be future from his point of view. He was just making a claim about how things were at his time. This means that although the sentence does describe John's attitude and although the sentence is about the utterance time (and therefore that time does, in fact, play a role in the semantics of the sentence), that time need not necessarily have played a role in John's mind.

These two complications have led to the name *double access*, used to describe the interpretation that often arises from a present tensed complement embedded under a past tensed attitude or speech verb (henceforth: *present-under-past*). The basic idea behind this name—as we have already seen—is that for, e.g., (2) to be true, there must be some description of the pregnancy state that holds throughout two points, namely the time that the attitude holder self locates himself (his now) and the utterance time. Exactly how to derive these truth-conditions—or even spell them out—is quite tricky and there have been many attempts in the last 37 years, starting with Smith 1978.[1] In what follows, we will not attempt to take further strides in

1 See Ogihara 1989, Heim 1994, Ogihara 1995, Abusch 1997 for more discussion and Bar-Lev 2014 for a recent overview.

this respect (at least not directly). Rather, we are interested in the following two questions, which are rarely asked:

(4) *Two key questions*
   a. Does double access arise in naturalistic settings?
   b. What conditions its appearance?

These questions are motivated by the observation that it is quite easy to construct infelicitous present-under-past reports. Compare, for example, (1) with (5) below:

(5) a. *Conversation at 10am in the mall.*
   Sue to John: How is Mary today?
   John to Sue: Mary is sick.
   b. *Conversation at 3pm on the same day at the beach.*
   Bill to Sue: How is Mary today?
   Sue to Bill: **#John believed that she is sick**.

Note that the bolded report has *believed* rather than *said* and that this difference alone makes the report odd. Why should this be? After all, John surely had a belief of the relevant kind at the mall, which prompted him to tell Sue that Mary is sick.

While such oddity has been noted, the contrast between (1) and (5) has never been explained. Work on double access has either disregarded the oddity (e.g., Abusch 1997: 39) or treated it as a reflex of a particular dialect. For example, Kratzer (1998: 14) writes that present-under-past reports "are in fact ungrammatical or marginal for many speakers, including some of my linguist colleagues. But there are enough speakers who like them, and this has to be explained."

Similarly, English grammars and manuals generally prohibit present-under-past reports. For example, the website English Practice provides the following rule, sometimes referred to as *sequence of tense*: "[i]f the tense in the principal clause is in the past tense, the tense in the subordinate clause will be in the corresponding past tense."[2] Interestingly, however, the website notes that "there are a few exceptions to this rule: A past tense in the main clause may be followed by a present tense in the subordinate clause when the subordinate clause expresses some universal truth." The following examples are provided:

(6) Copernicus proved that the earth moves round the sun.

(7) The teacher told us that honesty is the best policy.

Notice, however, that neither (1) nor (2) express a universal truth in the way that the examples above do. Moreover, this "universal truth" intuition does not explain the

2 http://www.englishpractice.com

contrast between (1) and (5).

The goal of the paper is to argue—based on a corpus study—that a present-under-past sentence is grammatical, modulated by two interacting pragmatic phenomena: *cessation* and *parentheticality*. The paper proceeds as follows. In the next section, we briefly introduce these two notions independently of double access before showing how they interact to yield a *pragmatic clash* in discourses like (5). Subsequently, we outline our corpus study which supports our analysis and raises some interesting new questions, which we summarize in the conclusion.

2 Deriving a pragmatic clash

2.1 Cessation

Imagine that Susie is at the doctor's office and sees a contorted look on Bob's face. Imagine further that Susie asks what the problem is and Bob replies:

(8) My heart was racing.

Susie will likely understand from (8) that Bob's heart was racing but no longer is. Notice that this inference can be canceled:

(9) My heart was racing and it still is.

This suggests that the inference in (8) is an implicature, which Altshuler & Schwarzschild (2012) call *cessation*. They define it as follows:

(10) *Cessation Implicature*
The utterance of a past tensed sentence implicates that no state of the kind described holds at the time of utterance.

While cessation may arise with a stative verb in the past tense, it need not. Consider what happens in the discourse below from Klein 1994 in which a judge poses the question (11a) to a witness, who then replies with (11b)-(11c):

(11) a. What did you notice when you looked in the room?
b. The light was on. There was a book on the table.
c. It was in Russian.

Here we infer that the book, if it still exists, is most likely still in Russian and nothing the witness says contravenes this. In other words, no cessation implicature is triggered by the past tensed (11c).

Following Altshuler & Schwarzschild 2012, 2013 we assume that cessation (and the lack thereof) has to do with tense choice. Altshuler & Schwarzschild propose

that (12) asymmetrically entails the proposition expressed by (8) and this supports a Gricean quantity implicature.

(12) My heart is racing.

Bob chose to utter (8) when he could have used the stronger statement (12). He must have avoided the stronger statement because it is false, assuming he possessed all the relevant information, which is plausible in this case. So the use of (8) implicates that Bob's heart is no longer racing.

Returning to (11), notice that the present tense is not possible in (11c):

(13) a. What did you notice when you looked in the room?
b. The light was on. There was a book on the table.
c. **#It is in Russian**.

(13c) is infelicitous because the reference time set by the previous discourse is wholly in the past. This conflicts with the semantics of the present tense, which (minimally) requires the reference time to overlap the utterance time. But if the present tense is not possible in (11c), we don't advance to Gricean reasoning in evaluating its past tense counterpart. Hence, no cessation.

2.2 Parentheticality

Let's return now to the discourse that we began the paper with:

(14) a. *Conversation at 10am in the mall.*
Sue to John: How is Mary today?
John to Sue: Mary is sick.
b. *Conversation at 3pm on the same day at the beach.*
Bill to Sue: How is Mary today?
Sue to Bill: **John said that she is sick**.

Notice that the literal content of Sue's bolded response in (14b) is not a felicitous answer to Bill's question (about how Mary is doing). However, Sue's answer is felicitous with a *parenthetical reading* (Urmson 1952; Hooper 1975), which allows Sue to offer the content expressed by the complement clause as a possible answer to Bill's question. According to Simons (2007), with such parenthetical uses, the complement carries the *main point* of the utterance while the matrix clause gets demoted to parenthetical status and plays an evidential function, indicating the source of evidence for the proffered content. Although attitude reports can be used in this way, they, of course, need not be. In (15b) below, the matrix clause is what is at issue—i.e. it carries the main point—given Sue's question to Bill.

(15) a. *Bill to Sue:* What did John just do?
b. *Sue to Bill:* **He said that Mary is sick**.

We propose that a pragmatic conflict between parentheticality and cessation triggers infelicity in the aforementioned discourse, repeated below. This discourse differs from (14) solely in that (16b) has *believed* rather than *said*.

(16) a. *Conversation at 10am in the mall.*
Sue to John: How is Mary today?
John to Sue: Mary is sick.
b. *Conversation at 3pm on the same day at the beach.*
Bill to Sue: How is Mary today?
Sue to Bill: **#John believed that she is sick**.

To see this, notice once again that, given Bill's question, we want to interpret the bolded report parenthetically, with the complement clause carrying the main point of the utterance, and the matrix clause playing an evidential function. Notice further that *believe* is a past tensed stative verb, where a present tense would have been possible: given that the question under discussion—*how is Mary today*?—concerns the speech time, Sue could have responded *John* ***believes*** *that Mary is sick*, but chose not to. This triggers a cessation implicature, whereby John is understood to no longer hold the described belief at the time that the belief report is uttered. This results in a *pragmatic clash*: on the one hand, Sue uses John's past belief as evidence for the suggestion that Mary is sick; on the other, John's belief cannot be taken to be evidence since Sue implies that it currently doesn't hold.

In sum, we propose that what explains the oddity of sentences like the bolded report in (16b) is a pragmatic clash between cessation and parentheticality. The prediction is that present-under-past sentences should be acceptable, unless the attitude report is interpreted parenthetically *and* there is a cessation implicature. Given that cessation only applies to stative predicates, we might expect more double access sentences with embedding eventive attitudes, such as *said* in (14b). The bolded report in (14b) is good precisely because it is eventive.

In the next section, we present corpus results showing that double access does occur in naturalistic settings and that it is more frequent with eventive embedders. Subsequently, we look at some naturally occurring discourses, to see how parentheticality and cessation interact.

3 Corpus study

In this section, we establish two properties of the distribution of matrix and embedded tense in English as a means for investigating double access. First, we show that

present-under-past tense configurations, which we take as an index of double access, are not only attested but, controlling for matrix tense, more prevalent than past-under-present configurations. This is important because we take it that the grammaticality of past-under-present is undisputed. Second, we establish that a verb's showing up in present-under-past configurations is conditioned by its eventivity: present-under-past is more common with eventives than with statives.

3.1 Dataset

Data about the distribution of matrix-embedded tense configurations were extracted from the Parsed UK Web as Corpus (PukWaC) corpus (Baroni, Bernardini, Ferraresi & Zanchetta 2009). PukWaC is the part-of-speech (POS) tagged and dependency parsed version of ukWaC, which is an approximately two billion word web scrape of the `uk` domain. To create PukWaC, ukWaC was lemmatized and POS tagged using TreeTagger (Schmid 1994) and dependency parsed using MaltParser (Nivre, Hall, Nilsson, Chanev, Eryigit, Kübler, Marinov & Marsi 2007).

Besides the fact that it has annotations useful for extracting tense sequence information, this corpus was chosen because it is large and has wide coverage—i.e. many different genres of text are represented. This wide coverage is useful since, in contrast to purely newswire-based parsed corpora, which will tend to include many quotations masquerading as double access—e.g., (17)[3]—there is likely to be more instances of informal, non-quotative text, such as that found in forums and blogs.

(17) Trump said, "He is a war hero because he was captured."

To begin, all cases of clausal embedding were extracted. This was done by extracting sentences in which a word whose tag matched the regular expression `VV[ZDGP]?` had as an OBJ dependent a word whose tag matched the regular expression `MD|V[BHV][ZDGP]?`.[4] An example of the relevant dependency structure is exemplified by the arrow labeled OBJ in Figure 1.

If the embedding verb was a VC dependent of an auxiliary verb or modal, the auxiliary chain was traced back until a non-VC dependency was found. An example of such a chain can be seen in the matrix clause of Figure 1. Each member of this chain along with its tense (if any) was recorded. If the embedded verb was an auxiliary that was an immediate VC parent of *have*—as is true of the embedded clause in Figure 1—this was also recorded.

3 http://www.politifact.com/truth-o-meter/statements/2015/jul/19/donald-trump/trump-i-called-mccain-hero-four-times

4 `MD` refers to a modal auxiliary. Any tag beginning with `V` refers to a verb: `VB` (*be*), `VH` (*have*), `VV` (other). The tags final letter specifies tense/aspect: `Z` (third person present tense), `D` (past tense), `G` (gerund), `P` (past participle), or the empty string (root form).

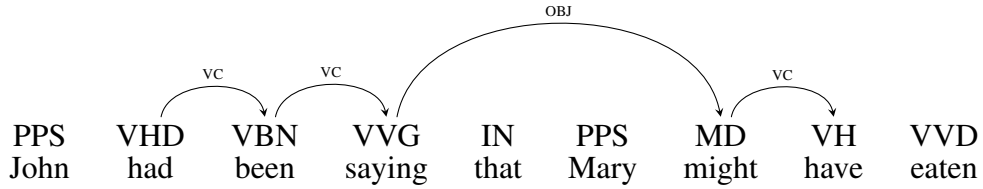

Figure 1 Example of dependency parse arrows relevant to dataset extraction.

Tense sequence data were then constructed in the following manner: if the highest element in a clause was a nonmodal, the tense encoded in its POS tag was mapped to the corresponding tense (*past* or *present*). Otherwise, it was mapped to *present* unless the modal was *could* or *would* or had a VC dependent with POS tag VH—e.g., *might have*, *may have*, etc. Two exceptions to the rule regarding *have* were *will (have)* and *shall (have)*, which were always mapped to *present*.

After this initial extraction, a crude filtering was applied to remove sentences that involve quotation by filtering any sentence containing a " character. This is necessary to remove cases like (17) which occur frequently in the newswire portions of ukWaC and which would be labeled present-under-past despite not involving double access.

Given the above criteria, a total of 180,847 sentences were extracted. These sentences were then further filtered to remove cases of nonfinite embedding—e.g., small clauses (18a) and various infinitival complements (18b).[5] Such cases were frequent in our sample of sentences, and this filtering step reduced the total number of sentences under consideration to 62,178.

(18) a. John saw Mary go to the store.
b. John wanted (Mary) to go to the store.

Next, sentences with matrix verbs not in the 100 most common were filtered. Lower-frequency verbs, many of which were *hapax legomena*, were determined by manual analysis to be highly likely to be misparses, including non-embedding verbs, non-verbs, and non-words. The sample of 100 verbs was enough to capture a large portion of the corpus, while still being a manageable number of verbs to evaluate individually for traits like eventivity. After this step, 44,808 sentences remained.

Since discussions of double access tend to focus solely on declarative complements, we also removed all question complements, yielding 40,512 sentences.

5 This was carried out by searching for dependents of the highest embedded verb that were POS tagged with TO. Small clauses are harder to filter. One method we employed was to check whether the embedded subject was a pronoun with accusative case.

MATRIX TENSE	EMBEDDED TENSE	RELATIVE FREQ	95% CI
past	*past*	0.264	(0.260, 0.268)
past	***present***	**0.104**	**(0.101, 0.107)**
present	*past*	0.114	(0.111, 0.117)
present	*present*	0.518	(0.513, 0.523)

Table 1 Relative frequency of each tense sequence. Confidence intervals were calculated using a nonparametric bootstrap with 10000 replicates.

Finally, the 100 verbs were then checked for whether they all at least allow a finite nonquestion complement. Two—*wonder* and *examine*—were determined to allow only embedded question complements, and so any sentences (erroneously) marked as taking declarative complements were further removed. This brought the final number of datapoints to 40,382 and the final number of verbs to 98.

3.2 Results

3.2.1 Attesting double access

To establish that double access is attested, we begin by assessing how often each tense sequence occurs in our dataset. Table 1 gives the joint relative frequency of each tense sequence—i.e. the proportion of times each sequence occurs in the dataset with respect to all of the others—along with 95% confidence intervals of that relative frequency calculated using a simple nonparametric bootstrap with 10000 replicates. We see here that, far from being unattested, present-under-past (0.103) occurs almost as often as past-under-present (0.114). This is interesting for the fact that there is no dispute that past-under-present is grammatical.

Indeed, while mismatching tense sequences (present-under-past and past-under-present) are less frequent than matching tense sequences (past-under-past and present-under-present) overall, among matrix past tense sentences, embedded present tense is more frequent (0.300) than embedded past tense among matrix present tense sentences (0.181).[6] That is, we find more tense mismatching given that the matrix tense is past than given that it is present. This result is corroborated by a Fisher's exact test ($p < 0.001$).

This apparent preference for present-under-past could, however, have the uninteresting explanations (i) that a small number of high frequency verbs—maybe

6 This is a consequence of matrix present tense being more frequent overall. While matrix past tense constitutes about 37% of the data, matrix present tense constitutes about 63%. Thus, nearly equivalent joint relative frequencies for the tense mismatch cases are converted to quite different conditional relative frequencies given matrix tense.

a single verb like *say*—allows present-under-past exceptionally; or (ii) that a few large documents containing many present-under-past—e.g., newswire text that did not place quotes around quotations—inflate the present-under-past proportions. To address these possibilities, we employ a mixed effects logistic regression with TENSE MATCH (MATRIX=EMBEDDED, MATRIX≠EMBEDDED) as the dependent variable, MATRIX TENSE (*past*, *present*) as a fixed effect, and VERB and DOCUMENT as grouping factors for random intercepts and random slopes for MATRIX TENSE. The fixed effect term explains differing prevalence of TENSE MATCH given MATRIX TENSE, averaging across VERB and DOCUMENT, and the random effect terms explain verb-specific and document-specific effects. Thus, if the fixed effect term is significant even in the presence of the random effects, this suggests that the apparent difference between the prevalence of present-under-past and the prevalence of past-under-under-present cited above is not driven by a few highly frequent verbs or large documents.

To test the significance of MATRIX TENSE, we utilize a likelihood ratio test, comparing a model with MATRIX TENSE as a predictor to one without it. Under this test, the fixed effect of MATRIX TENSE is significant ($\chi^2(1) = 31.89$, $p < 0.001$) and goes in the expected direction: tense mismatch is more common with matrix past tense than with matrix present, even controlling for verb-specific effects. (The increase in the log-odds of a tense mismatch when the matrix tense is present is approximately 0.912.) The upshot of this is that, if one accepts that past-under-present is indisputably grammatical but would like to explain away the apparent existence of present-under-past, it will be difficult to do so as a purely verb-specific or document-specific effect.

This, however, should not be taken to mean that verbs show no variability in their ability to occur with present-under-past. Indeed, the above model puts the (marginal) predicted probability of embedded present given matrix past at about 0.271 (estimated log-odds: -0.987), but it furthermore predicts that 95% of verbs will fall in the quite large interval [0.045, 0.745] (estimated standard deviation of the verb random intercept: 1.050). Thus, there is actually quite a bit of variability among verbs that might be explained by particular properties of those verbs. In light of the preceding discussion, one such property that seems likely to be relevant is a verb's eventivity.

We investigate this possibility more fully in the next section, but suggestive evidence can be seen in Figure 2. This figure shows the Best Linear Unbiased Predictors (BLUPs) of the verb random effects—roughly, how much the model believes each particular verb deviates from the mean across all verbs.[7] On the *y*-axis is the log-odds of embedded present (v. embedded past) given matrix past,

7 In fact, these are not the BLUPs themselves, but rather linear combinations thereof. The 0s on the axes correspond to the estimated population means for each cell of the regression design.

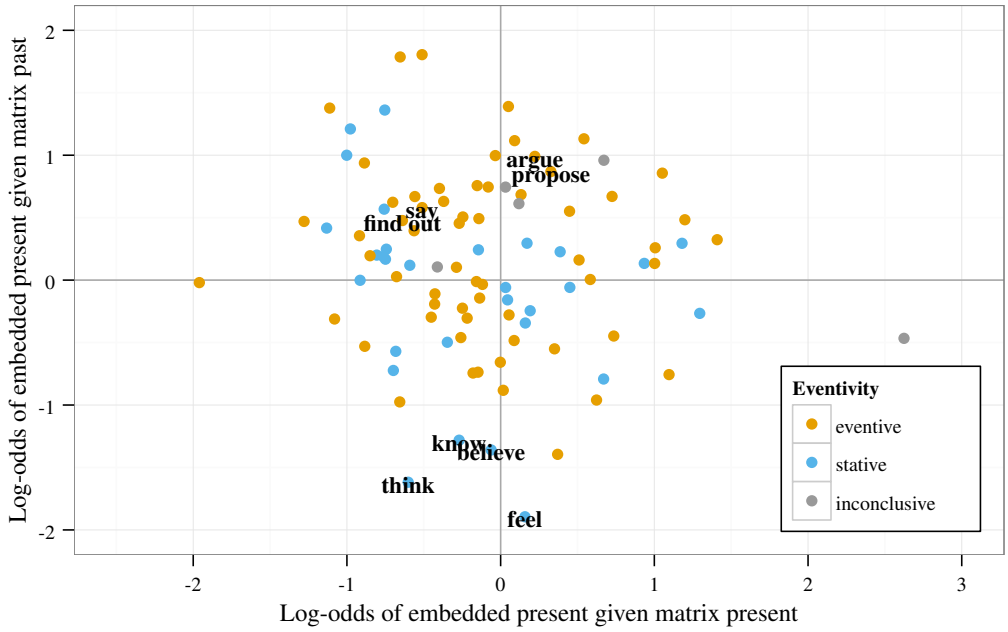

Figure 2 Best Linear Unbiased Predictors for verb random effects.

and on the *x*-axis is the log-odds of embedded present (v. embedded past) given matrix present. Being higher on the *y*-axis thus means showing higher preference for present-under-past compared to past-under-past. The color of each point shows the eventivity value of the corresponding verb, as determined by the annotation procedure described in the next section. As can be seen from the fact that there are more orange points toward the top of the graph and more blue toward the bottom, eventives tend to prefer present-under-past more than statives. The question we address in the next section is whether this trend is reliable.

3.2.2 Conditioning double access

Eventivity annotation Altshuler and Roberts annotated the 98 verbs in the final dataset for whether they were eventive or stative when used with declarative complements. To do this, both annotators individually applied each of the following five tests for stativity.

(19) a. The bare present form of the verb yields a nonhabitual interpretation
b. The verb may not be used in the imperative.

c. The verb may not be the complement of force.
d. The verb may not be in the complement of a pseudocleft.
e. The verb may not be used with progressive aspect.

For each test, the verb was marked as *stative*, *eventive*, or *inconclusive*.[8] Inter-annotator agreement was high (Cohen's κ=0.698). Cases where the annotators disagreed were resolved jointly by both annotators. After this resolution, verbs were mapped to *eventive* or *stative* based on plurality vote over the tests. When plurality vote did not resolve to *stative* or *eventive*, either because there was no plurality or because the plurality was *inconclusive*, the verb was mapped to *inconclusive*. Five verbs were mapped to this value: *show*, *ensure*, *demonstrate*, *prove*, and *establish* (grey dots in Figure 2). These *inconclusive* verbs are addressed in the analysis in two ways: (i) by removing them from consideration; and (ii) by retaining them and imputing their value.

Analysis As in the previous analysis, we begin by investigating the relative frequencies of the values of EVENTIVITY, MATRIX TENSE, and EMBEDDED TENSE. For this initial analysis, the five inconclusive verbs were removed from consideration. Figure 3 shows the conditional relative frequency of embedded present tense given matrix tense and eventivity. Fisher's exact tests suggest a reliable nonindependence between EVENTIVITY and EMBEDDED TENSE among matrix past sentences ($p < 0.001$) but not among matrix present sentences ($p = 0.117$). As such, we focus in on only the matrix past sentences for the remainder of this section.

As before, this effect could be a consequence of verb- or document-specific effects. To test the reliability of the EVENTIVITY effect among matrix past tense sentences, controlling for VERB and DOCUMENT, we again utilize a mixed effects logistic regression—this time with MATRIX TENSE as the dependent variable, a fixed effect of EVENTIVITY, and random intercepts for VERB and DOCUMENT. This model was compared against one without the fixed effect of EVENTIVITY in a likelihood ratio test. Consonant with the Fisher's exact test, EVENTIVITY comes out as a significant predictor of EMBEDDED TENSE ($\chi^2(1) = 5.66$, $p < 0.05$), and it goes in the expected direction: eventives are more likely (estimated increase in log-odds: 1.043) than statives to take embedded present tense given that the matrix tense is past.

8 Certain verbs, particularly verbs of communication, were noted to differ in eventivity based upon the animacy of their subject. Because such verbs occur in the corpus more frequently with animate subjects, these verbs' eventivity values were resolved to whatever value was associated with their behavior with animate subjects. This judgment is based off an evaluation carried out by Hacquard and Roberts of all subjects of matrix *say*, the most frequent attitude verb. These subjects were overwhelmingly animate.

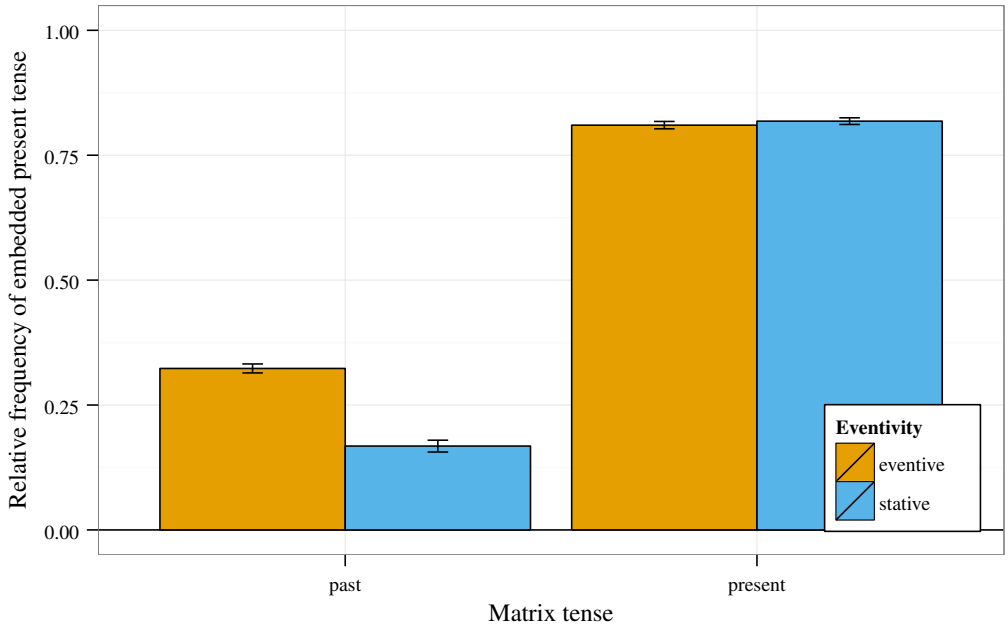

Figure 3 Conditional relative frequency of embedded tense given matrix tense and eventivity. Error bars give 95% confidence intervals computed by a nonparametric bootstrap with 10000 replicates.

One worry that remains here is that, by removing the five verbs marked *inconclusive*, we may have underestimated the uncertainty regarding the eventive preference for present-under-past. To remedy this, we utilize random regression imputation (see Gelman & Hill 2006: Ch. 25). A random effects logistic regression with EVENTIVITY as the dependent variable and random intercepts for DOCUMENT was fit to the data, excluding the *inconclusive* verbs. This model was then used to predict the probability of a particular EVENTIVITY value (*eventive* or *stative*) for each instance of the *inconclusive* verbs, using the fixed intercept estimate and the DOCUMENT random intercept BLUPs. If a document only contained *inconclusive* verbs in our dataset, the values for verbs in that document were set to the probability corresponding to the fixed intercept (estimate log-odds of *eventive*: 0.260).

These predicted probabilities were then used as the basis for a parametric bootstrap. In each iteration of this bootstrap, a value (*eventive* or *stative*) was sampled for each instance of an *inconclusive* verb based on that instance's predicted probability. For each of 10000 parametric bootstrap samples generated via this procedure, a nonparametric bootstrap sample was drawn over the dataset excluding the *inconclusive*

verbs—i.e. the same resampling procedure that generated the confidence intervals in Figure 3 was repeated for each parametric iteration. These parametric and non-parametric samples were then combined, and the relative frequencies computed. Of interest here, the mean relative frequency of embedded present for eventives with matrix past remains the same to three significant figures (0.323) as does the 95% confidence interval [0.314, 0.332]; in contrast, the mean relative frequency of embedded present for statives with matrix past rises slightly (from 0.168 to 0.176) and the 95% confidence interval concomitantly shifts upward without a change in size (from [0.156, 0.180] to [0.164, 0.187]). Nonetheless, controlling for matrix tense, the prevalence of present-under-past is still much higher with eventives than statives.

Using this same method, we also assessed whether the EVENTIVITY effect was reliable when controlling for effects of VERB and DOCUMENT. The full eventivity model described above was fit to each of 1000 new resampled datasets and the coefficient for EVENTIVITY extracted. Consonant with the earlier likelihood ratio test, the distribution of this coefficient shows a reliable increase of present-under-past for eventives compared to statives among matrix past sentences (mean increase in log-odds: 0.456, 95% CI: [0.016, 0.900]), suggesting that controlling for uncertainty introduced by the eventivity annotation as well as the possibility of verb-specific and document-specific effects, eventives still prefer present-under-past more than statives.

3.3 Cessation and parentheticality in our corpus

Our proposal predicts that double access should only be ruled out when parentheticality clashes with cessation. We thus expect to find double access sentences with stative embedders, so long as they involve either no cessation, or if cessation, then no parentheticality:

(20) a. no cessation, no parentheticality
b. no cessation, parentheticality
c. cessation, no parentheticality
d. #cessation, parentheticality

To test (20), we inspected a small sample of past matrix sentences under the statives *think*, *believe*, *feel*, and *know*.[9] We observed that most cases had neither cessation nor parentheticality (20a). We found a handful of cases of no cessation with parentheticality (20b) and cessation with no parentheticality (20c), but no cases

9 20 random instances of past-under-past, and 20 instances of present-under-past for each of these verbs, and of the eventives *say* and *tell*.

of cessation and parentheticality (20d). Below, we provide an instance of each kind, leading to some discussion about why (20b) and (20c) were relatively rare.

We begin with the discourse below, where we have bolded the attitude report under consideration:

(21) The response on the subject of the current student numbers and the government's aim of achieving a graduate population of 50% was mixed. **A number of responses felt that there are currently excessive numbers of students and courses**. However it was also communicated that to maintain a stable and diverse society there should be a varied range of courses and equal opportunities for students to benefit from a university education.

We note that there is no cessation implicature in the bolded report: the discourse conveys nothing about current feelings. This is likely due to the fact that the reference time for *felt* is a time that is set prior to the utterance time by the past tensed *was mixed*. Moreover, there is no parentheticality: the report elaborates on the fact that the results were mixed; what is at issue here is how different respondents felt.

Example (22) is a case of cessation with no parentheticality:

(22) Researchers lead by a team at the UK's Wellcome Trust Sanger Institute have published a detailed analysis of the human X-chromosome. An accompanying study uncovered the surprise finding that women have two active copies of many X-chromosome genes. **Previously, scientists believed that one of the two X-chromosomes present in every cell of a female embryo is effectively 'shut down' early in development**.[10]

The presence of *previously* makes cessation clear. We infer that scientists had a certain belief prior to the discovery, which they no longer hold, given this discovery. As in (21), (22) does not exemplify parentheticality because the report establishes a contrast between what is known now and what the scientists used to believe.

Example (23) involves no cessation, but it arguably exemplifies parentheticality:

(23) Following consultation among colleagues, it has been agreed to hold a contacts conference on 18th April 2002 at Scarman House, Warwick University. The conference is intended for departmental contacts for a representative if the contact is not free on that day. **We felt that Warwick is fairly central, and is within easy reach of Birmingham International for colleagues who fly in from Scotland or Northern Ireland**.[11]

10 http://www.bionews.org.uk/page_12293.asp
11 http://escalate.ac.uk/1248

The discourse conveys nothing about how the author currently feels (no cessation). This is likely due to the mention of a consultation that occurred in the past, setting the reference time for *felt* at the time of the consultation. As for parentheticality, one may argue that what is at issue is the location of the conference. In this case, the main point would be carried by the complement clause, which establishes that Warwick is central and accessible to all.

A much clearer case of parentheticality and no cessation is given in (24), which involves an eventive matrix verb (which cannot trigger cessation) and *S-lifting* (Ross 1973), where the complement clause precedes the matrix (which triggers parentheticality):

(24) "Obviously we are very troubled by the Russians' decision. ... The move has serious implications for U.S. security interests and those of our friends and allies in the Middle East." **Between now and Dec. 1, Washington hopes to persuade Russian officials to retract their decision to break the deal, U.S. officials said**.[12]

In sum, both parentheticality and cessation are found in our corpus. However, both are rare with past tensed, stative attitudes. This may be, in part, an artifact of our choice of corpus: we suspect that both cessation and parentheticality might be more frequent in spoken corpora. The rarity of *cessation* with statives, for instance, may stem from the fact that many of our sentences are embedded within a past narrative, where the reference times do not overlap the utterance time. As for *parentheticality*, while such uses were rare with statives, they were relatively frequent with eventives like *say* or *tell*. This asymmetry echoes Hooper's (1975) intuition that parenthetical readings are more easier with past tense *strong assertives* (e.g., *say, report*) than with *weak assertives* (e.g., *think, believe*); see Anand & Hacquard 2014 for more discussion of the strong vs. weak contrast.

Finally, it is important to note that our corpus extractions revealed that even when the conditions on cessation and parentheticality are satisfied, there are interpretations of present-under-past that are distinct from double access. For example, we found several examples of the following variety:

(25) My next appointment after that is 31/01 when hopefully we will find out if all this pain and suffering has done its stuff. I'll admit to being VERY nervous now! **I was thinking that the radiotherapy is now finished so everything will sort itself out.**

Here we see the embedded present tensed *is now finished* gets a purely relative interpretation. That is, the speaker shifts the perspective to his now, recalling how

12 http://www.casi.org.uk/discuss/2000/msg01208.html

the radiotherapy was finished at the time of his thinking. Nothing is said about the status of the radiotherapy at the time that (25) was uttered. In effect, one could argue that this is a case of an embedded historical present (Bary & Altshuler 2014), yielding an interpretation of the embedded present that is found in Russian and Hebrew present-under-past reports (see Schlenker 1999, Sharvit 2003 for a discussion of such cases).

3.4 Do cessation and parentheticality always clash?

Consider the following discourse (Guillaume Thomas p.c.):

(26) a. Where is the Holy Grail hidden?
b. The esteemed and late professor von Klech believed that it is buried under Notre Dame de Paris.

At first blush, one may wonder why this discourse is acceptable even though the matrix in (26b) is clearly parenthetical (given the question in (26a)) and clearly exemplifies cessation (since the attitude holder is understood to be dead). Note, however, that the death of the professor is not a good reason to doubt its validity as a reliable source of information. Hence there is no pragmatic clash. In other words, cessation and parentheticality does not automatically result in a pragmatic clash if the cessation does not entail that the attitude holder has changed her mind.

4 Conclusion

This paper addressed two questions: (i) Does double access arise in naturalistic settings? And (ii) what conditions its appearance? Aided by a corpus study, we have argued that double access is not a fringe phenemonon, but that its acceptability is modulated by two pragmatic factors: cessation and parentheticality.

In this way, we hope to have shed light on the disagreement in previous literature about acceptability of double access. We think that the root of this disagreement was (i) not considering a wide range of data and (ii) not controlling for cessation and parentheticality. With respect to (i), we saw that many of our examples involved embedded clauses with generic statements, which may lead one to think that double access is only good with "universal truths." However, when one looks at naturally occurring discourses, it becomes clear that expressing universal truths is not a necessary condition for double access. With respect to (ii), we believe that a pragmatic clash between cessation and parentheticality may have led to the queasiness that some linguists felt towards the classic example *John believed that Mary is pregnant*, in Abusch's context in which we are discussing Mary's recent weight gain. This context makes the complement clause at issue (parentheticality), but the past tense

on *believe* triggers cessation.

We have thus argued that double access is grammatical, but that it can lead to infelicity when cessation and parentheticality clash. However, this clash is not specific to double access *per se*: any attitude report, including past-under-past, should be infelicitous when parentheticality and cessation conflict.[13] This appears to be borne out. Sue's response seems odd, but would be much more felicitous with a present tense on *believe* or with *said* instead.

(27) a. *Conversation at 10am in the mall*
Sue to John: Why didn't Mary come to the show last night?
John to Sue: She was sick.
b. *Conversation at 3pm on the same day at the beach*
Bill to Sue: Why didn't Mary come to the show last night?
Sue to Bill: **#John believed that she was sick**.

References

Abusch, Dorit. 1997. Sequence of tense and temporal *de re*. *Linguistics & Philosophy* 20. 1–50.

Altshuler, Daniel & Roger Schwarzschild. 2012. Moment of change, cessation implicatures and simultaneous readings. In E. Chemla, V. Homer & G. Winterstein (eds.), *Proceedings of Sinn und Bedeutung 17*, 45–62. Paris.

Altshuler, Daniel & Roger Schwarzschild. 2013. Correlating cessation with double access. In Maria Aloni, Michael Franke & Floris Roelofsen (eds.), *Proceedings of the 19th Amsterdam Colloquium*, 43–50. Amsterdam.

Anand, Pranav & Valentine Hacquard. 2014. Factivity, belief and discourse. In L. Crnic & U. Sauerland (eds.), *The Art and Craft of Semantics: A Festschrift for Irene Heim*, 69–90. Cambridge, MA: MITWPL.

Bar-Lev, Moshe. 2014. *Sequence of Tense in English and Hebrew and the de re interpretation of tenses*: The Hebrew University of Jerusalem M.A. Thesis.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The wacky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3). 209–226.

Bary, Corien & Daniel Altshuler. 2014. Double Access. In Eva Csipak & Hedde Zeijlstra (eds.), *Proceedings of Sinn und Bedeutung 19*, Semantics Archive.

Gelman, Andrew & Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Heim, Irene. 1994. Comments on Abusch's theory of tense. In Hans Kamp (ed.), *Ellipsis, Tense and Questions*, 143–170. Amsterdam: University of Amsterdam.

13 Thanks to Maribel Romero for raising this point.

Hooper, Joan. 1975. On assertive predicates. In J. Kimbell (ed.), *Syntax and Semantics 4*, 91–124. Academy Press.

Klein, Wolfgang. 1994. *Time in Language*. London: Routledge.

Kratzer, Angelika. 1998. More structural analogies between pronouns and tenses. In *Semantics and Linguistic Theory (SALT) 8*, 1–22.

Nivre, Joakim, Johan Hall, Jens Nilsson, Atanas Chanev, Gülsen Eryigit, Sandra Kübler, Svetoslav Marinov & Erwin Marsi. 2007. MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering* 13(02). 95–135.

Ogihara, Toshiyuki. 1989. *Temporal reference in English and Japanese*: University of Texas, Austin dissertation.

Ogihara, Toshiyuki. 1995. Double-access sentences and reference to states. *Natural Language Semantics* 3. 177–210.

Ross, John Robert. 1973. Slifting. In Maurice Gross, Morris Halle & Marcel-Paul Schützenberger (eds.), *The Formal Analysis of Natural Languages*, 133–169. The Hague: Mouton de Gruyter.

Schlenker, Philippe. 1999. *Propositional attitudes and indexicality: A cross-categorial approach.*: MIT dissertation.

Schmid, Helmut. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference on New Methods in Language Processing*, vol. 12, 44–49. Citeseer.

Sharvit, Yael. 2003. Embedded tense and universal grammar. *Linguistic Inquiry* 34(4). 669–681.

Simons, Mandy. 2007. Observations on embedding verbs, evidentiality and presupposition. *Lingua* (117). 1037–1056.

Smith, Carlota. 1978. The syntax and interpretation of temporal expressions in English. *Linguistics & Philosophy* 2. 43–99.

von Stechow, Arnim. 1982. Structured propositions. *Report of Sonderforschungsbereich* 99.

Urmson, J.O. 1952. Parenthetical verbs. *Mind* (61). 480–496.

von Stechow, Arnim. 1995. On the proper treatment of tense. In Mandy Simons & Teresa Galloway (eds.), *Semantics and Linguistic Theory (SALT) 5*, 362–386. Ithaca, New York: Cornell University.

Author contributions

Altshuler and Hacquard contributed the main theoretical proposals that structure the paper as well as the qualitative corpus analysis (sections 3.3 and 3.4). Altshuler furthermore conducted the verb eventivity annotation with Roberts (section 3.2.2).

Roberts conducted corpus filtering (section 3.1) and extracted sentences and contexts for the qualitative analysis (Section 3.3). White wrote the corpus API to PukWaC and the code built on top of that API for searching dependency parses and extracting sentences (section 3.1). He conducted all statistical analysis (section 3.2) besides the interannotator agreement analysis in section 3.2.2, conducted by Roberts.

Daniel Altshuler
Institute for Language and Information
Heinrich-Heine-Universität Düsseldorf
Universitätsstraße 1
40225 Düsseldorf, Germany
daltshul@gmail.com

Valentine Hacquard
University of Maryland
1401 Marie Mount Hall
College Park, MD 20742-7505
hacquard@umd.edu

Thomas Roberts
University of Maryland
1401 Marie Mount Hall
College Park, MD 20742-7505
rotom@umd.edu

Aaron Steven White
University of Maryland
1401 Marie Mount Hall
College Park, MD 20742-7505
aswhite@umd.edu