The CommitmentBank:
Investigating projection in naturally occurring discourse[1]

Marie-Catherine de MARNEFFE — *The Ohio State University*
Mandy SIMONS — *Carnegie Mellon University*
Judith TONHAUSER — *The Ohio State University / University of Stuttgart*

Abstract. This paper describes a new resource, the CommitmentBank, developed for the empirical investigation of the projection of finite clausal complements. A clausal complement is said to project when its content is understood as a commitment of the speaker even though the clause occurs under the scope of an entailment canceling operator such as negation or a question. The study of projection is therefore part of the study of commitments expressed by speakers to non-asserted sentence content. The content of clausal complements has been a central case for the study of projection, as there is a long-standing claim that clause-taking predicates fall into two classes – factives and nonfactives – distinguished on the basis of whether the contents of their complements project. This claim identifies the embedding predicate as the primary determinant of the projection behavior of these contents. The CommitmentBank is a corpus of naturally occurring discourses whose final sentence contains a clause-embedding predicate under an entailment canceling operator. In this paper, we describe the CommitmentBank and present initial results of analyses designed to evaluate the factive/nonfactive distinction and to investigate additional factors which affect the projectivity of clausal complements.

Keywords: projective content, attitude predicates, factive predicates, corpus annotation.

1. Introduction

To understand the messages conveyed by speakers or writers, we need to recognize which propositions they intend to commit themselves to by their utterances. Most straightforwardly, speaker commitment is indicated by assertion, typically carried out by the utterance of a declarative clause. But of course, the picture is far more complex, because speakers may be understood to be committed to contents which are not asserted. This paper is concerned with the empirical investigation of one particular sub-class of cases where speakers may be understood to be committed to, or to be certain about, non-asserted content. We focus on the contents of finite clausal complements of clause-embedding predicates such as *think, accept, tell* and *know*.

For some clause-embedding predicates, speaker commitment to the content of the complement (CC, henceforth) seems straightforwardly explainable in terms of entailment. It is generally held that sentence (1) entails the CC of *know*.

(1)   Jane knows that it is snowing.

As speakers are generally assumed to be committed to the (obvious) entailments of their assertions, a speaker of (1) will be taken to be committed to the claim that it is snowing.

However, as is well-known, speaker commitment to the CC may remain even when the predicate (in our case *know*) is embedded under an entailment canceling operator. In general, speak-

---

ers are *not* committed to content which occurs under such an operator, or to its entailments. The entailment canceling environments usually considered are enshrined in the so-called Family of Sentences (see Chierchia and McConnell-Ginet 1990), illustrated in (2). The standard claim is that utterances of these sentences would typically commit the speaker to the truth of the CC of *know*, even though that content is no longer entailed by the sentence as a whole.

(2)   a. Jane doesn't know that it is snowing.
         b. Does Jane know that it is snowing?
         c. Jane may know that it is snowing.
         d. If Jane knows that it is snowing, she will wear her snow boots, hat and gloves.

Content which is expressed under the scope of an entailment canceling operator but which is nonetheless understood to be a commitment of the speaker is said to *project*. Content which has the potential to display this behavior is *projective*. The question of interest in this paper is this: Under what circumstances will the CC of a clause-embedding predicate project?[2]

On standard semantic accounts (e.g., Heim 1983; van der Sandt 1992), projection of the CC is a consequence of it being *presupposed*. Exactly what is meant by presupposition varies between different accounts, but all standard accounts agree that projection is a characteristic feature of presupposed contents, so much so that the tendency for content to project is very typically taken as a diagnostic for presuppositionality (but see Chierchia and McConnell-Ginet 1990 for early articulation of reservations about this diagnostic). In the case of clause-embedding predicates, there is a long-standing view (dating back to Kiparsky and Kiparsky 1970) that these predicates divide into two classes: factive embedding predicates, which lexically encode presuppositionality of their CCs; and nonfactive embedding predicates, which do not. The contents of factive complements, being presupposed, can project over entailment canceling operators. The contents of nonfactive complements are generally taken to be non-projective.

The factive/nonfactive categorization might seem to make straightforward predictions about projection of CCs. For a variety of reasons, this is an over-simplification. First, there is a long-standing observation that any presuppositional content, including CCs, may fail to project when projection would generate an inconsistency. For example, in the sequence in (3) the speaker cannot be taken to be committed to the CC, as she has just explicitly denied it.

(3)   It is not snowing, so of course Jane doesn't know that it's snowing.

Moreover, the classification of a predicate as nonfactive simply carries the claim that its CC will not project by virtue of being presuppositional; it does not follow that there are no discourse circumstances in which that content may project for other reasons. Indeed, the recent literature contains some passing suggestions regarding projective interpretations of complements of non-factives (see Anand and Hacquard 2014; Schlenker 2010). Further complicating the picture, Karttunen (1971) suggested that for at least some factives, which he dubbed "semi-factives," the person of the matrix subject or the tense of the matrix predicate might affect whether the complement projects. There are further relevant factors: Tonhauser et al. (2018) found that the projectivity of utterance content, including the CC of factive, semi-factive and nonfactive

predicates, is gradient and that content is more projective the more it is not at-issue with respect to the question under discussion. Tonhauser (2016) found that the prosodic realization of sentences with factive and semi-factive predicates influences the extent to which the complement projects. With these additional factors taken into account, it is clear that the factive/nonfactive categorization can be at best a very partial predictor of projection in actual cases. In fact, as we will argue below, this categorization has very little predictive power (see section 3.1.1).

Our goal with the CommitmentBank has been to create a resource for the empirically-based study of projection of CCs, using naturally occurring examples and basing analysis on judgments of projection provided by theoretically untrained speakers. The ultimate goal of the project is to provide a predictive account of projection of clausal contents. In this paper, we detail the process of creating the CommitmentBank, our methodology for collecting projection judgments, and the further annotations that we have carried out. We also present preliminary results concerning the predictive power of a variety of features of the discourses. The CommitmentBank corpus is available at `https://github.com/mcdm/CommitmentBank`.

2. The CommitmentBank corpus

In this section, we detail how we built the corpus and which annotations have been carried out.

2.1. Data

The CommitmentBank contains 1,200 examples of naturally occurring discourse segments extracted from three corpora of different genres: the Wall Street Journal (WSJ, news articles), the fiction component of the British National Corpus (BNC, fiction) and Switchboard (SWBD, dialogue). Each discourse consists of a target sentence with a clause-embedding predicate embedded under an entailment canceling operator (negation, modal, antecedent of conditional, or question) with up to 2 prior context sentences/turns. Examples are given in (4) and (5), where the target sentence is underlined. The alphanumeric code in brackets after each example is the internal code of the item in the CommitmentBank.

(4)   What fun to hear Artemis laugh. She's such a serious child.
        I didn't know she had a sense of humor. [BNC-1607]

(5)   A: Oh yes. Animals have a way of talking.
        B: Alfie did. I tell you if I could have gotten a hold of that cat that day.
        A: I don't know uh that I'd trade my dog in for the world. [SWBD-243]

These discourse segments were automatically extracted. First, using the Stanford dependency parser (Klein and Manning 2003; de Marneffe et al. 2006), we obtained dependency parses for each sentence in the three corpora, and extracted sentences in which a predicate has a clausal complement and the predicate is itself embedded under negation or a modal, or in a question or the antecedent of a conditional. To guarantee accuracy of the samples, the results were manually curated. The 1,200 examples represent 48 different clause-embedding predicates. Table 1 shows the number of discourses for each predicate, by embedding environment.

Predicate	Conditional	Modal	Negation	Question	Predicate	Conditional	Modal	Negation	Question	Predicate	Conditional	Modal	Negation	Question
accept	0	0	1	1	forget	0	4	7	2	recognize	0	0	1	0
admit	1	3	1	1	guarantee	0	2	0	0	remember	1	4	2	0
announce	1	1	0	1	guess	0	6	9	5	see	1	27	10	1
assume	1	7	1	2	hear	2	5	3	1	seem	0	0	2	0
believe	5	19	40	10	hope	0	17	1	2	say	21	40	39	14
bet	0	0	0	1	hypothesize	0	0	0	1	show	0	2	0	0
bother	0	0	1	0	imagine	2	14	12	1	signal	0	1	0	1
convince	0	3	4	0	insist	2	1	0	2	specify	0	1	0	0
decide	3	8	0	0	know	18	16	78	21	suggest	4	3	11	1
demand	0	1	1	0	learn	2	0	2	2	suppose	2	3	2	1
expect	0	1	4	0	mean	4	14	27	7	suspect	5	11	4	0
fear	2	1	0	0	notice	1	7	23	3	swear	0	1	0	0
feel	4	8	16	6	occur	0	0	1	0	take	0	0	0	1
figure	1	0	0	0	pretend	0	2	2	2	tell	6	21	7	4
find	6	9	1	5	prove	0	4	1	0	think	21	39	265	61
foresee	1	0	0	0	realize	0	3	20	6	understand	0	4	4	1

Table 1: Number of discourses by predicate in each embedding environment.

2.2. Projection (speaker commitment) annotations

In order to determine in each discourse segment whether the CC projects, we must determine whether interpreters judge the speaker of the target sentence to be committed to that content. We operationalize this by asking how certain the speaker/author is that this content is true. For each discourse segment, we elicited such judgments from at least eight self-reported native English speakers, using a questionnaire created on Amazon's Mechanical Turk Platform. (The elicitations were performed with IRB approval; distribution was restricted to IP addresses in the United States. Annotators were paid $1.00 for completing a questionnaire.)

Figure 1 shows an annotation task. The discourse is displayed, with a proper name in bold face identifying the implied speaker (or with speaker A/speaker B for the Switchboard dialogues). Annotators are asked for a judgment about the certainty of the speaker with respect to the CC. They respond on a Likert scale labeled at 3 points (+3/speaker is certain that the CC is true, 0/speaker is not certain whether the CC is true or false, -3/speaker is certain that the CC is false). We follow, i.a., Tonhauser 2016; Tonhauser et al. 2018 in using the 'certain that' diagnostic to tap into annotators' judgments of speaker commitment, but we have altered the scale to allow annotators to indicate a judgment that the speaker is committed to the negation of the CC. While the issue of "counterfactivity" has been largely neglected in the theoretical literature, it is just as important, from a natural language processing perspective, as the issue of factivity. For information extraction, for instance, it is important to recognize that in (5) above, the speaker is committed to **not** trading in their dog.

Each questionnaire contained eight discourses of interest and two constructed control dis-

Sally: While the rest of the gang dived into the pub opposite to use the toilets, I called in at H. R. Higgins (Coffee-man) Ltd and bought six gift boxes of coffee (assorted) and two of tea (scented). That was my Christmas shopping sewn up. Who said it was stressful?

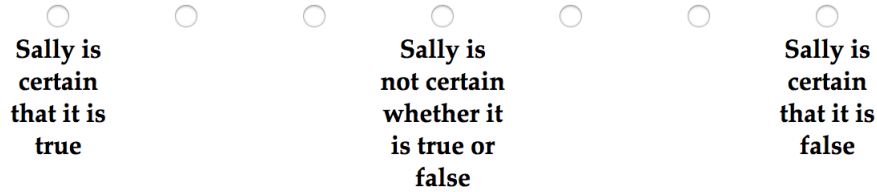Tell us how certain Sally is that Christmas shopping was stressful.

○	○	○	○	○	○	○
Sally is certain that it is true			**Sally is not certain whether it is true or false**			**Sally is certain that it is false**

Figure 1: Item display for projection annotation on Mechanical Turk.

courses used to assess annotators' attention. In each questionnaire, one control discourse clearly indicated speaker certainty that the CC is true (6), while the other clearly indicated speaker certainty about the negation of the CC (7). For the "true" control items, we accepted responses of +2 or +3; for the "false" ones, we accepted responses of -3 and -2. Data from annotators who gave other responses to at least one control item was excluded from analysis.[3]

(6) **John**: Tess was our star in the marathon this year. She's always trained with all her heart and soul. After all that training, she was happy to cross the finish line.
Prompt: Tell us how certain John is that Tess crossed the finish line.

(7) **A**: Did you hear anything about Olivia's chemistry test?
B: Well, she studied really hard. But even after putting in all that time and energy, she didn't manage to pass the test.
Prompt: Tell us how certain speaker B is that Olivia passed the test.

The complements *p* in the prompt to the annotators (*Tell us how certain the speaker is that p*) were manually constructed from the discourses to ensure that pronouns and temporal references were correctly resolved. For instance in (4), the complement is *she had a sense of humour*, but the pronoun *she* is resolved to *Artemis* in the prompt. We discarded discourses where the target sentence but not the context was contained in a direct quote (as in (8), as here the speaker of the target sentence is not the same as the speaker of the context). In addition, we discarded a wide variety of other cases where it was difficult to construct a brief and accurate paraphrase of the CC.[4] Example (9) shows a case where the clausal complement of *prove* contains a bound pronoun (*drug makers/their pill*), making paraphrase of this content impossible. We also discarded discourses where the automatic extraction was erroneous. Overall, we discarded 60% of the data extracted from WSJ, 70% of the data from the BNC, and 40% for SWBD.

(8) The small increase in the index of leading indicators, which had climbed 0.5% in August but was unchanged in July, does lend support to the view that the economy has

---

[3]All control items we used can be found online, at `https://github.com/mcdm/CommitmentBank`.
[4]These examples raises the question of how, as theorists, we should discuss or evaluate projection in such cases.

slowed noticeably. However, it doesn't give much of a clue as to whether a recession is on the horizon. "I don't think it provides much new information on the economy," said Richard Rippe, economist at Dean Witter Reynolds Inc.

(9)   But courts quickly tumbled down a slippery slope. Just as all plaintiffs are not alike, it turns out that DES defendants marketed the drugs differently and may have offered different warranties. The ultimate result came in Hymowitz v. Lilly, where the highest New York court expanded the market-share approach for the first time to say that drug makers that could prove Mindy Hymowitz's mother didn't use their pill must still pay their share of any damages.

2.3. Person and number of the subject of the clause-embedding predicate

We automatically annotated the person of the subject of the predicate. The subject of the predicate is found using the dependency graph, and its person and number are assigned using heuristics based on its lemma and part-of-speech tag. Person is "first", "second" or "third". Number is "singular", "plural", or "unknown" in the case of a "you" subject. We are currently manually annotating whether a "you" subject is impersonal or not.

2.4. Temporal reference of the matrix clause

The temporal reference of the matrix clause of each target sentence was manually annotated, as "present" (time of utterance is included in the temporal reference of the matrix clause), "past" (temporal reference of the matrix clause fully precedes time of utterance) or "future" (time of utterance fully precedes the temporal reference of the matrix clause). Manual annotation was required because temporal reference was not straightforwardly determinable from tense marking. The annotation was carried out by native speakers of English based on their intuitions about intended temporal reference.

2.5. Modality

The standard Family of Sentences used to diagnose projection includes embedding of content under an epistemic possibility modal. As noted, we automatically extracted sentences including possibility modals (including *could*, *may*, *might*, *probably*, *possibly*, *maybe*, *possible*, *probable*, *perhaps*). However, many of these can be used to signal multiple types of weak modality. In order to enable us to restrict analyses to just those cases involving the standard epistemic case, we manually annotated each modal item using the following categorization:

Epistemic: The modal expresses speaker uncertainty about whether an event did or will happen. It is an expression of the speaker's evaluation of the possibility of the event, as in (10).

(10)   Like now. The Community in Knockglen would defend Eve vociferously. Even some of the Sisters here in Dublin might see that the girl had a point. [BNC-1074]

Ability: The modal pertains to the ability of the subject to do something, as in (11).

(11) The car engine roared again and the red car moved off but it didn't go far. Stuart's Mum was reversing out of her space and her car was blocking the way. Stuart could see that the red car was going to turn back. [BNC-1069]

Deontic: The modal is an expression of permission, as in (12).

(12) Nick rolled his eyes upwards. "Not so bad then." She wished she could tell him that Mr Evans hadn't stolen the Will after all but Nick had never thought that he had so there was no point in it. [BNC-1158]

Circumstantial: Circumstantial modality concerns what is possible or necessary given a particular set of circumstances, as in (13). This covers a wide range of cases. In practice, if a modal did not fit any of the above three categories, the Circumstantial label was used.

(13) It was a complex language. Not written down but handed down. One might say it was peeled down. [BNC-1015]

2.6. Plausibility of the CC given the context

In some cases, the speaker of a discourse may be judged certain about the CC (or its negation) based on information contained in the preceding discourse (as in (14), where the existence of the Toacks is entailed by the first sentence) or based on world knowledge (as in (15), where credit card number is known to be sensitive information).

(14) The English read : IT IS FORBIDDEN TO CROSS THE TOACKS. That summer I came to know the Toacks - with their roots hooked under the earth's crust. On the right side of them you'd never guess they even existed but from where I was they were too deep to fathom and too tall to cross. [BNC-1450]

(15) But what we may not know is just what makes somebody a sucker. What makes people blurt out their credit-card numbers to a caller they've never heard of? Do they really believe that the number is just for verification and is simply a formality on the road to being a grand-prize winner? [WSJ-31]

To evaluate the role of information external to the target sentence in projection, we carried out an annotation task in which we collected certainty ratings about the CC in the absence of the target sentence. We used the same task as described in Section 2.2, but presented annotators with only the context sentences, *without* the target sentence. For instance, for the discourse presented in Figure 1, annotators would see exactly the same prompt, but without the last sentence of Sally's utterance, *Who said it was stressful?*, and were asked the same question ("How certain is the speaker that Christmas shopping was stressful?"). As before, judgments of annotators who did not answer as expected on control items were discarded. Such annotations have been obtained for 558 discourses from the WSJ and the BNC. We assume that these annotations are a proxy for the plausibility of the CC given the context.

3. Analysis of factors explaining projection

We use the CommitmentBank to analyze several factors put forth in the literature as having some role in projection and to quantitatively assess their contribution to projection, first independently and then combining them. The analyses were carried out on a slightly restricted
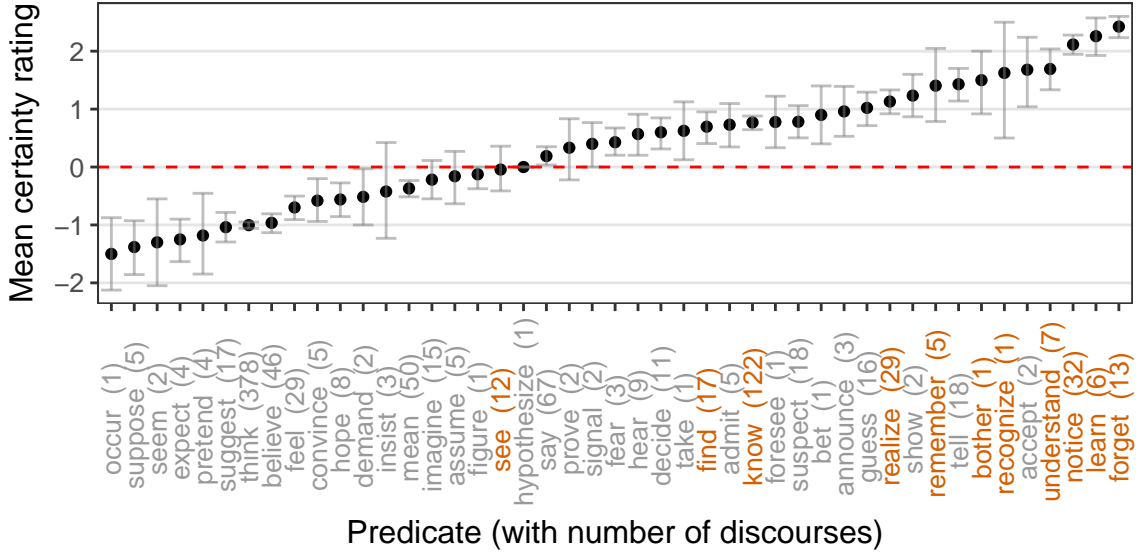
Figure 2: Mean certainty ratings for CCs, by predicate, restricting modal environment to epistemic modals. Number of discourses in parentheses. Error bars indicate bootstrapped 95% confidence intervals. Purported factive predicates are in orange, nonfactive predicates in gray.

subset of the data. As noted above, our data includes clause-embedding predicates occurring under a variety of weak modals. The standard literature, however, only considers weak epistemic modals as providing a diagnostic for projection. In order to ensure that we are analyzing CCs which, on standard views, have the potential to project, we eliminated examples with non-epistemic modals from analysis. The resulting dataset consists of 982 discourses (100 from WSJ, 461 from BNC, 421 from SWBD), with 45 different clause-embedding predicates.[5]

3.1. Evaluating the role of the predicate

The first factor we consider is the predicate. Figure 2 shows the mean certainty ratings for all items by predicate, arranged in order from lowest mean certainty rating (reflecting judgments that the speaker is certain that the CC is false) to highest mean certainty rating (reflecting judgments that the speaker is certain that the CC is true). Factive predicates are shown in orange.[6] We observe by-predicate variability: for instance, the mean certainty rating for *occur* is lowest, at -1.5, whereas the mean certainty rating for *forget* is highest, at 2.43. We now consider how classifying predicates as factive/nonfactive contributes to the projectivity of the CC, and then consider how individual predicates contribute.

---

[5]Compared to Table 1, the following three verbs are not present in the restricted dataset: *guarantee*, *swear*, *specify*.
[6]The verb *see* has been noted to have an evidential use in which the CC does not routinely project over entailment canceling operators. We hand checked the *see* examples in this subset of the data and determined that none involved an evidential use of the type identified in, e.g., Simons 2007.

3.1.1. Factivity

As discussed in the introduction, two classes of clause-embedding predicates are standardly distinguished: factives, which signal that the CC is presupposed, and hence projective; and nonfactives, which do not signal anything about the CC. The CommitmentBank includes both types of predicates; the factive verbs found in CommitmentBank are listed in (16):

(16)   *bother, find, forget, know, learn, notice, realize, recognize, remember, see, understand*

On a simplistic interpretation, the standard view – projection is a consequence of presupposition, and presupposition is determined by the embedding predicate – leads us to expect that all items with factive predicates will be judged highly projective (i.e., receive high speaker certainty ratings), while items with nonfactive predicates will have certainty ratings around zero. In other words, whether the predicate is factive or not should be a strong predictor of projectivity.

Figure 2 shows a picture not entirely out of line with the predictions of this simple interpretation: there is indeed a clustering of factive predicates at the right hand side of the graph. However, there is no clean separation of factive and nonfactive predicates: some of the nonfactive predicates have higher mean certainty ratings than some factives, with *know* – perhaps the most cited example of a clearly factive verb – relatively low compared with others in that class. Nonetheless, the CC of *know* is still overall judged projective, as predicted by the standard account. Note that our results confirm the claims of Anand and Hacquard 2014 and Schlenker 2010 mentioned above, that at least some nonfactive predicates can give rise to projection.

The impression created by Figure 2 that semantic factivity is a predictor of projection, albeit a weak one, is confirmed by an ordinal mixed-effects model[7] predicting the certainty ratings from a fixed effect of 'factivity' (a binary factor indicating whether the predicate is in the list in (16), with 'nonfactive' the reference level). The model included random by-annotator intercepts (capturing differences in projectivity between annotators). The coefficient (β) for 'factivity' is positive ($\beta = 1.60$, $SE = 0.05$, $z = 34.69$, $p < 0.001$), indicating, as expected, that overall the CCs of factive predicates received higher certainty ratings than those of nonfactives, i.e., are more likely to project. However the Nagelkerke R^2 of this model is similar to a model which uses the person of the subject of the predicate as fixed effect (0.126 vs. 0.124). Nagelkerke R^2 is a measure of goodness of fit of the model, which can be thought of as quantifying how much of the variance in the data is captured by the model: an R^2 of 0 means that the model does not account for any of the variance in the data, whereas 1 indicates a perfect fit. In other words, the person of the matrix subject accounts for the same amount of variance in the data as the traditional characterization of a predicate as factive or nonfactive. (We return to the contribution of person in section 3.2.)

Of course, theories that account for projection in terms of lexically encoded presupposition allow for the absence of projection in particular cases involving factive predicates. Advocates of such accounts might propose additional pragmatic accounts of the fact that theoretically untrained speakers appear to treat projection as a gradient rather than a binary feature. Also, as we have already noted, standard analyses do not rule out projection of the CCs of nonfactive

---

[7]We used the `ordinal` package (version 2018.8-25, Christensen (2018)) in R (version 3.5.0, R Development Core Team (2016)).

Factor	Nagelkerke R^2
Genre	0.062
Predicate tense	0.078
Embedding	0.091
Person of predicate subject	0.124
Factivity	0.126
Predicate lemma	0.228

Table 2: Nagelkerke R^2 of ordinal mixed-effects models, predicting certainty ratings of the 982 discourses, from 'Factor' as fixed effect and random by-annotator intercepts.

predicates, but only requires that this projection be accounted for without invoking lexically encoded presuppositionality. Nevertheless, the patterns of projection judgments in the CommitmentBank raise a question about how the binary, categorical distinction between factives and nonfactives is to be drawn. Kiparsky and Kiparsky (1970) originally distinguished factives from nonfactives on the basis of a range of syntactic and semantic/pragmatic observations, including but not limited to judgments of projection. But currently, factives are standardly distinguished from nonfactives on the basis of the projectivity of the CC and whether the CC is entailed (e.g., Gazdar 1979; Chierchia and McConnell-Ginet 1990; Schlenker 2010; Anand and Hacquard 2014).[8] The judgments on which the distinction rests are generally those of theorists considering extremely simple sentences like those in (2), considered without any context. But is the binary, categorical distinction between factives and nonfactives supported by more rigorous testing? A first challenge comes from Tonhauser et al.'s (2018) finding that there is significant variability in how projective the CCs of factive predicates are: the CC of *discover*, for instance, was significantly less projective than that of *know*.

A further challenge comes from the naturally occurring discourses of the CommitmentBank. By modifying the aforementioned model, we can quantify the degree to which the projection behavior of each predicate differs from that predicted by its factive/nonfactive categorization. This is done by adding random by-predicate intercepts (capturing differences in projectivity between predicates) to the ordinal mixed-effects model which predicts the certainty ratings from the fixed effect of 'factivity' (now: $\beta = 1.57$, $SE = 0.30$, $z = 5.32$, $p < 0.001$). Table 3 gives the random effect coefficients for each predicate. These values tell us how the intercept for each predicate needs to be adjusted: positive coefficients mean that the predicate leads to higher certainty ratings than is predicted by its factive/nonfactive classification, and negative coefficients mean that the predicate leads to lower certainty ratings. For instance, the CCs of *know* and *realize*, with negative coefficients, are less projective than the overall mean projectivity of the factive CCs. On the other hand, the CCs of *accept* and *tell*, with positive coefficients, project more than the CCs of other nonfactive predicates. The variability observed between the coefficients of factive predicates suggests that the projection behavior of factive predicates is more heterogeneous than assumed and, hence, may challenge the assumption that the factive/nonfactive distinction is central to understanding projectivity.

---

[8]Kiparsky and Kiparsky (1970) did not require factives to entail the CC.

Predicate	Coef.	Predicate	Coef.	Predicate	Coef.	Predicate	Coef.
accept	1.48	feel	-0.62	**learn**	0.92	show	0.94
admit	0.57	figure	0.11	mean	-0.33	signal	0.30
announce	0.83	**find**	-0.84	**notice**	0.75	suggest	-0.98
assume	-0.10	foresee	0.70	occur	-0.77	suppose	-1.28
believe	-0.99	**forget**	1.20	pretend	-1.25	suspect	0.71
bet	0.71	guess	1.01	prove	0.19	take	0.42
bother	-0.18	hear	0.51	**realize**	-0.34	tell	1.48
convince	-0.40	hope	-0.44	**recognize**	0.13	think	-0.87
decide	0.49	hypothesize	0.19	**remember**	0.24	**understand**	0.23
demand	-0.38	imagine	-0.19	say	0.20		
expect	-1.10	insist	-0.43	**see**	-1.44		
fear	0.29	**know**	-0.71	seem	-0.96		

Table 3: Random effect coefficients for each predicate in an ordinal mixed-effect model with 'factivity' as fixed effect and random by-predicate and by-annotator intercepts. Predicates assumed to be factive are in bold.

3.1.2. Predicting projection from predicate lemma

Figure 2 showed that there is between-predicate variability in the projectivity of the CC. Here we consider a second way in which the predicate may be taken to contribute to the projectivity of the CC. Specifically, we consider an ordinal mixed-effects model in which the certainty ratings are predicted from a fixed effect of predicate lemma (with *accept* as the reference level and treatment coding) and random by-annotator intercepts. The last row of Table 2 gives the Nagelkerke R^2 of the model: 0.228. This model thus captures more variation than the model which uses semantic factivity as a fixed effect. This is not totally surprising, given that semantic factivity is a coarse-grained binary classification of the predicates and given that there is significant projection variability between factive predicates (Tonhauser et al. 2018). In the context of these observations, the CommitmentBank data suggests that properties of individual predicates are of greater predictive value with respect to projection than the factive/nonfactive distinction, and so call for a careful analysis of the lexical properties which affect projection. We also note, however, that the Nagelkerke R^2 of 0.228 indicate that there is still a lot of variance in the data that the model does not capture. We consider some other factors below.

3.1.3. Between-item and between-annotator variability

The previous sections showed that there is between-predicate variation, even within the classes of factive and nonfactive predicates. Our results show variation along two further dimensions: between the items of predicates and between annotators (see also Tonhauser et al. 2018). As illustration of between-item variation, consider Figure 3, which shows the mean certainty ratings for each item with *know*, *believe* and *tell*. As is evident, the CC of none of these predicates has a uniform projection profile: certainty ratings are highly item-sensitive, for both the canonical factive *know* and the nonfactives *believe* and *tell*.
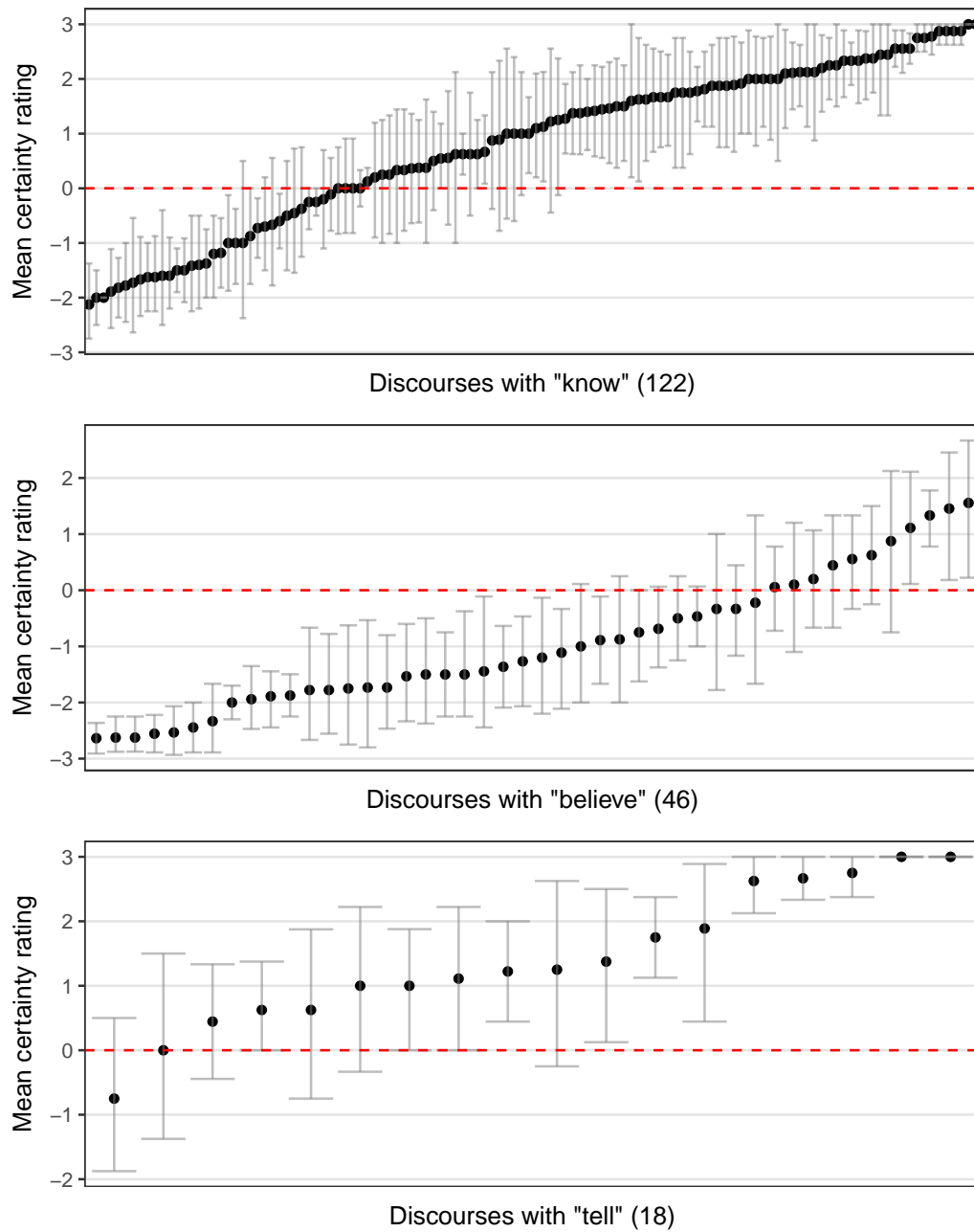
Figure 3: Mean certainty ratings for CCs of *know*, *believe* and *tell*. Number of discourses in parentheses. Error bars indicate bootstrapped 95% confidence intervals.

The examples in (17) illustrate the contrasts between discourses for which the CC was highly projective (17a, 17c) and discourses for which the CC was not projective (17b,17d).

(17)   a.   At the heart of the universe there is cruelty. We are predators and are preyed upon every living thing. Did you know that <u>wasps lay their eggs in ladybirds piercing the weak spot in their armour</u>? [BNC-2375, mean: +3]

        b.   "Rather a long shot wasn't it? Twenty years? How do you know <u>the baby was born here</u>?" [BNC-2394, mean: -0.25]

        c.   The Susweca. It means "dragonfly" in Sioux you know. Did I ever tell you <u>that's where Paul and I met</u>? [BNC-2630, mean: +3]

        d.   His reaction to the news had been partly predictable and partly complex and more disturbing. There had been the natural initial shock of disbelief at hearing of the unexpected death of any person even casually known. He would have felt no less if <u>he'd been told that Berowne was dead of a coronary or killed in a car smash</u>. [BNC-428, mean: -0.75]

The plots in Figure 3 illustrate that in addition to between-item variability, there is also between-annotator variability: large confidence intervals indicate that annotators vary in their ratings. Examples (17a) and (17c) are cases where there was no such variability: all annotators judged the speaker to be certain that the CC is true (+3). But in other examples, there was variability in the certainty ratings. In (18), 5 annotators said 0, 3 annotators replied negatively -1, -2 and -3, and 4 annotators replied positively (2 saying +1 and 2 saying +2).

(18)   A: Yeah I had a friend whose brother did steroids and as soon as he stopped working out he just ballooned out. It all turned to fat and he just really looked bad.
B: Really. Ugh. I didn't know <u>it did that</u>. [SWBD-251]

In (19), all annotators replied negatively, but in different degrees: 2 replied -3, 4 said -2 and 2 replied -1.

(19)   B: It's just a great cat but.
A: Well my personal preference is a dog. Uh I don't know uh that <u>I would ever want a cat</u>. [SWBD-114]

In roughly a third of the 982 discourses, annotators agreed on whether the CC projects but disagreed in the degree of (non)projection, as in (19). In two thirds of the discourses, some annotators gave projecting and others gave non-projecting responses. Between-annotator variation in our naturally occurring items is in line with the findings of Tonhauser et al. 2018, who also found that participants gave significantly different certainty ratings.

In summary, we have found, unsurprisingly, that knowing the identity of the embedding predicate contributes positively to predictions of projection, but a great deal of additional information is also required. Knowing the categorization of the predicate as factive or nonfactive, however, has comparatively less predictive power.
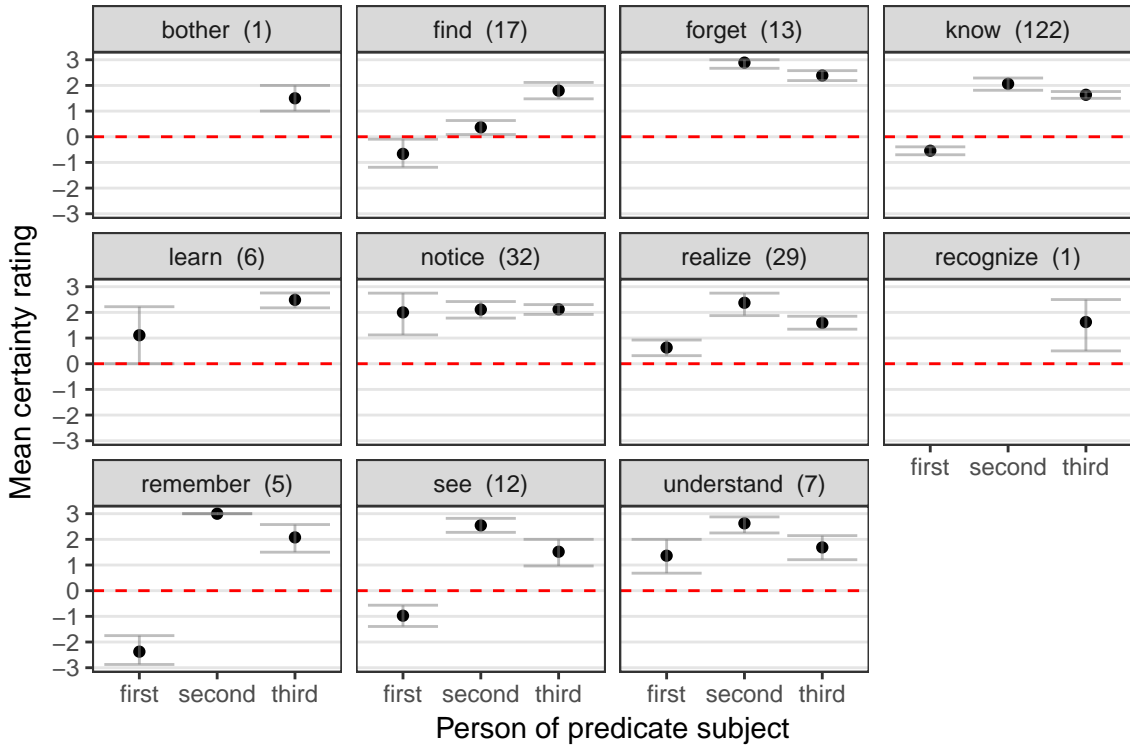
Figure 4: Mean certainty ratings for CCs by person subject of factive predicates. Number of discourses in parentheses. Error bars indicate bootstrapped 95% confidence intervals.

3.2. Subject and tense of the predicate

There are claims in the literature on factives that the person of the subject of the factive predicate may affect whether or not the CC projects (i.a., Karttunen, 1971; Gazdar, 1979; Stalnaker, 1974). The effect of person can be illustrated by pairs such as (20) and (21):

(20) If I discover that the data has been falsified, I will report it to the authorities.

(21) If he discovers that the data has been falsified, he will report it to the authorities.

While (21), considered with no additional context, suggests that the speaker believes that the data has been falsified, this implication is absent in (20). Stalnaker (1974) offers a pragmatic account of this difference. Beaver (2010), providing a much more thorough discussion of effects of person based on naturally occurring data, also argues, on the basis of careful examination of individual cases, that the effects of person are essentially pragmatic, and interact with additional factors in determining presence or absence of projection. The CommitmentBank provides a further resource for such examination of specific cases, but here we will ask a more global question: to what extent is the person of the factive subject a predictor of projection?

Figure 4 plots the mean ratings per person for all 11 factive predicates in the data. For most predicates, the CCs of predicates with first person subjects are associated with lower certainty ratings than those of predicates with non-first person subjects.

As noted above, the person of the predicate subject captures as much variance as the factivity
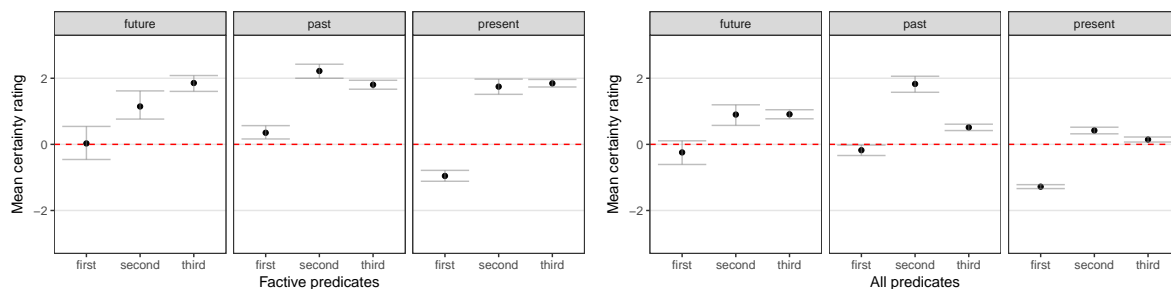
Figure 5: Means certainty ratings by tense and subject person, for CCs of factives on the left, for CCs of all predicates on the right. Error bars indicate bootstrapped 95% confidence intervals.

classification. Table 2 revealed that the tense of the predicate alone only captures a trivial amount of variance (0.078). But the CommitmentBank data suggests that there is an interaction between the tense of the predicate and the person of the subject, as previously proposed in the literature. The left panel of Figure 5 shows the certainty ratings by tense and subject person for the CCs of factive predicates. A similar picture emerges for all predicates, but with lower ratings overall, as shown in the right panel of Figure 5. Log-likelihood comparisons of an ordinal mixed-effects model with fixed effects for tense and person, and their interaction, and a model without the interaction confirm that the interaction is significant ($p < 0.001$). This is true for models fitted to the factive predicates and models fitted to all predicates. (All models included random by-annotator intercepts.) This observation raises the question of whether separate projection analyses for factive and nonfactive predicates are empirically justified.

In summary, our results indicate that information about person of the subject and tense of the predicate are relevant, if small, factors in predicting projectivity for examples in context.

3.3. Embedding and genre

Karttunen (1971) pointed out that for at least some factive predicates, projection may differ across entailment-canceling environments. For example, he notes that when *discover*, *find out* and *see* occur in a question, the CCs do not necessarily project: he observed that (22) "can also be understood as a sincere request for information. The speaker is not sure about the truth of the complement and is prepared to accept the addressee's discovery as a fact" (p.63).

(22)   Did you discover that you had not told the truth?

There has also been some experimental evidence suggesting that embedding environments lead to different projection behavior. In particular, Smith and Hall (2014) investigated projection under negation and antecedent of a conditional, finding a difference in degree of projection for the CC of *know* (projecting more under negation than under the antecedent of a conditional).

Figure 6 illustrates that embedding environments might indeed lead to different projection behavior. The much lower certainty rating for negation in SWBD is probably due to a high proportion of discourses with *believe* and *know* in the first person which receive a Neg Raising reading, such as (5) above, or (23) for which all annotators interpret the speaker to be committed to the negation of the CC (all 8 annotators gave -2).
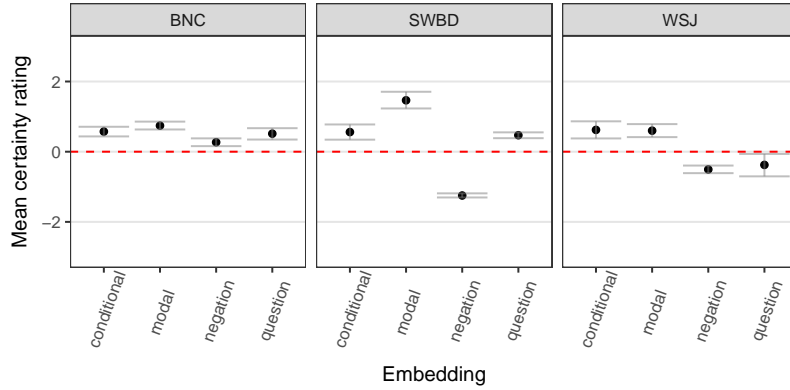
Figure 6: Mean certainty ratings for CCs of all predicates, by embedding and genre. Error bars indicate bootstrapped 95% confidence intervals.
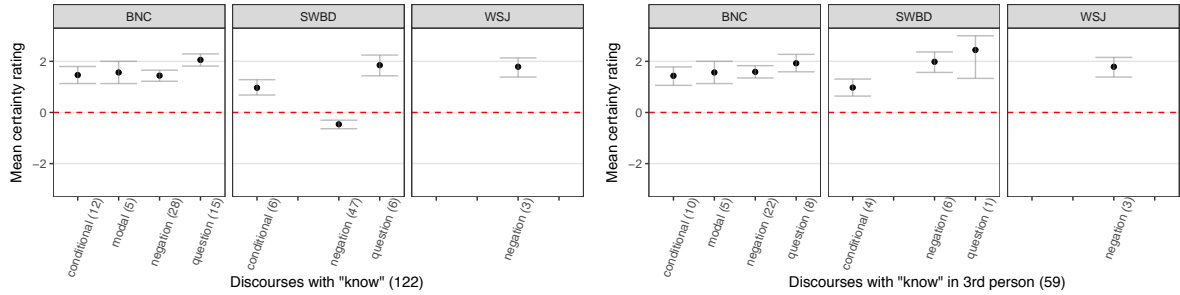

Figure 7: Mean certainty ratings by embedding and genre, for CCs of *know* on the left panel, and for CCs of *know* in the third person on the right. Error bars indicate bootstrapped 95% confidence intervals.

(23) A: but at the same time I think it would do them a world of good.
      B: Yeah.
      A: But there's a
      B: I don't know that you could require everyone yeah to do it for a whole year or two years or something like that. [SWBD-245]

Figure 7 concentrates solely on *know*: the left panel plots all 122 discourses with *know*, the right panel concentrates on discourses in the third person (59 discourses). As can been seen, when restricting the data to third person discourses, discourses under negation in SWBD project more, but there still appears to be an effect of embedding. Table 2 suggests that genre and embedding capture only a small amount of the variance in the data. More experimental evidence is needed to fully understand the impact of embedding and genre on projection.

3.4. Plausibility of the CC given the context

Tonhauser et al. (2018) hypothesized that the prior plausibility of the CCs influences projectivity. The mean plausibility ratings of the CCs (see Section 2.6) give an indication of whether there is contextual information in favor of the truth of the CC (positive mean) or the falsity of the CC (negative mean), or whether the context is uninformative regarding the CC (mean

around 0). In the 558 WSJ and BNC discourses for which we have such annotations, 9% have a mean smaller or equal to -1, 74% a mean between -1 and 1, and 17% a mean equal to or above 1. We fitted an ordinal mixed-effects model that predicted the certainty ratings for the target sentences from these plausibility means (for the 558 BNC and WSJ discourses). The Nagelkerke R^2 of this model is low, at 0.036, probably because for most items (74%), there is no contextual information against or in favor of the truth of the CC.

The CommitmentBank data and the annotations for the plausibility of the CC given the context allow for further analyses. Which predicates are used to embed the complement in the target sentence of cases where the context or the CC already prime projection? In which cases does the full target sentence alter the certainty ratings compared to the ones obtained when not seeing the target sentence? We leave these questions for future research.

3.5. Summary analyses

In the previous sections, we examined several factors which have been claimed to be relevant to projection of CCs, and assessed their individual contributions to explaining projection behavior. Unsurprisingly, a model with predicate as fixed effect accounts for more variation in the data than any other, including a model with factivity as fixed effect. We now bring all of the factors together into a single model. We fitted an ordinal mixed effects model that predicts the certainty ratings from fixed effects of embedding and genre, and their interaction, as well as tense and person, and their interaction, and the predicate lemma. We again included random by-annotator intercepts. (More complex models did not converge.) All of the fixed effects were significant ($p < 0.001$), as established by log-likelihood model comparisons. This finding confirms claims in the theoretical literature that these factors play a role in projection. However, the model still only has a Nagelkerke R^2 of 0.35 (if we replace predicate lemma by factivity, the Nagelkerke R^2 is 0.29). We also fitted a variant of this model which includes the plausibility means of the CCs as a fixed effect, for the 558 BNC and WSJ discourses for which we have such annotations. In that model, genre is not a significant factor (the certainty ratings for BNC and WSJ discourses do not differ much). The Nagelkerke R^2 of the model is 0.31 (if we replace predicate lemma by factivity, the Nagelkerke R^2 is 0.24). These findings suggest that there is more variability in the data to be accounted for.

Overall, the picture that arises from the CommitmentBank suggests that an account of projectivity requires the integration of many factors. For theorists and those with practical concerns who are interested in predicting whether, in a given utterance, the CC will be understood as a commitment of the speaker, there is clearly more work to be done in uncovering the detailed features of contexts and utterances which contribute to this evaluation.

4. Conclusion

We presented a new resource, the CommitmentBank, for studying projection behavior of the contents of complements of clause-embedding predicates in naturally occurring data. Our main finding is that there is significant variability in projection judgments in such data. An empirically adequate theory of projection will need to identify and incorporate the factors that contribute to that variability. The CommitmentBank can be used to continue to investigate factors that have been hypothesized to impact projection, such as the question under discussion and

information structure. We are currently annotating the data for additional pragmatic factors that might play a role in projection, including coherence relations between the target sentence and the context. More analyses of the CommitmentBank will shed light on the linguistic and extra-linguistic factors at play in projection, and we hope that the data will be used as a testbed for future theories of projection.

References

Anand, P. and V. Hacquard (2014). Factivity, belief and discourse. In *The Art and Craft of Semantics: A Festschrift for Irene Heim*, pp. 69–90. MIT Working Papers in Linguistics.

Beaver, D. (2010). Have you noticed that your belly button lint colour is related to the colour of your clothing? In R. Bäuerle, U. Reyle, and T. E. Zimmermann (Eds.), *Presuppositions and Discourse: Essays Offered to Hans Kamp*, pp. 65–99. Elsevier.

Chierchia, G. and S. McConnell-Ginet (1990). *Meaning and Grammar*. Cambridge, MA: MIT Press.

Christensen, R. H. B. (2018). "ordinal": Regression models for ordinal data. R package version 2018.8-25. http://www.cran.r-project.org/package=ordinal/.

de Marneffe, M.-C., B. MacCartney, and C. D. Manning (2006). Generating typed dependency parses from phrase structure parses. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pp. 449–454.

Gazdar, G. (1979). A solution to the projection problem. In C.-K. Oh and D. A. Dinneen (Eds.), *Syntax and Semantics*, Volume 11: Presupposition, pp. 57–89. New York: Academic Press.

Heim, I. (1983). On the projection problem for presuppositions. In M. Barlow, D. Flickinger, and M. Westcoat (Eds.), *West Coast Conference on Formal Linguistics (WCCFL) 2*, pp. 114–125.

Karttunen, L. (1971). Some observations on factivity. *Papers in Linguistics 4*, 55–69.

Kiparsky, P. and C. Kiparsky (1970). Facts. In M. Bierwisch and K. E. Heidolph (Eds.), *Progress in linguistics*, The Hague, Paris. Mouton.

Klein, D. and C. D. Manning (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st Meeting of the Association of Computational Linguistics*, pp. 423–430.

R Development Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Schlenker, P. (2010). Local contexts and local meanings. *Philosophical Studies 151*, 115–142.

Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua 117*(6), 1034–1056.

Smith, E. A. S. and K. C. Hall (2014). The relationship between projection and embedding environment. In *48th Annual Meeting of the Chicago Linguistic Society*.

Stalnaker, R. C. (1974). Pragmatic presuppositions. In M. Munitz and P. Unger (Eds.), *Semantics and Philosophy*, pp. 197–213. New York: New York University Press.

Tonhauser, J. (2016). Prosodic cues to presupposition projection. *Semantics and Linguistic Theory 26*, 934–960.

Tonhauser, J., D. Beaver, and J. Degen (2018). How projective is projective content? Gradience in projectivity and at-issueness. *Journal of Semantics 35*, 495–542.

van der Sandt, R. (1992). Presupposition projection as anaphora resolution. *Journal of Semantics 9*, 333–377.