Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Angela Cao April 10, 2022

An Analysis of Causal Language Constructions in Diverse Discourse Data

By

Angela Cao

Jinho D. Choi
Advisor

Linguistics

Jinho D. Choi
Advisor

Yun Kim
Committee Member

Marjorie Pak
Committee Member

David Zureick-Brown
Committee Member

2022

An Analysis of Causal Language Constructions in Diverse Discourse Data

By

Angela Cao

Jinho D. Choi
Advisor

An abstract of
A thesis submitted to the Faculty of the Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements for the degree of
Bachelor of Arts with Honors

Linguistics

2022

Abstract

An Analysis of Causal Language Constructions in Diverse Discourse Data
By Angela Cao

Creating datasets of manually annotated texts for relationships such as causality has been of interest to computational linguists. This thesis introduces the annotated **Constructions of CAUSE, ENABLE, and PREVENT** (CCEP) corpus to contribute to the field by systematizing the nuanced CAUSE, ENABLE, and PREVENT roles and enabling annotation of a wide variety of causal construction types. This corpus utilizes *constructions* as the basic unit of causal language, which is based on the linguistic paradigm entitled Construction Grammar (CxG) and manifests through the surface construction labeling (SCL) approach. In this project, I adapt a pre-identified bank of causal connectives (the Constructicon) from Dunietz (2018), which are used as triggers for annotation instances. Through high inter-annotator performance demonstrated in the corpus of 150 doubly-annotated documents based on the CCEP guidelines, I (1) support Wolff et al. (2005)'s causal aspectualization as psychologically real through high inter-annotator agreement of distinguishing such, (2) build upon previous annotation work that aim to embed this model of causation, and (3) provide a high quality dataset for understanding textual causality.

An Analysis of Causal Language Constructions in Diverse Discourse Data

By

Angela Cao

Jinho D. Choi
Advisor

A thesis submitted to the Faculty of the Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements for the degree of
Bachelor of Arts with Honors

Linguistics

2022

Acknowledgments

Thank you to:

Emory's Center for Mind, Brain, and Culture for their research grant.

Professors Marjorie Pak, Yun Kim, Lelia Glass, and David Zureick-Brown for their mentorship.

Atticus Geiger for introducing me to the study of causal language.

Professor Jinho D. Choi and Dr. Gregor Williamson for their advisory.

Gail Tynkov and Fizza Mahmood for making life a bit brighter.

My father, mother, and sister for their unconditional love and support.

It is an understatement to say that these people have contributed to the completion of my thesis; *I am only as much as the sum of my parts*, and these are the people who have made me everything that I am during my undergraduate career. I am infinitely grateful.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Overview

1.1 Introduction

Humans can decipher causal relations in conversation without a second thought. "I am going to the store because I need milk," I tell you. So you know that if *I did not need milk*, then *I would not have gone to the store*, and that there exists some notion of causation between these two events. This counterfactual theory of causation has its origins from Hume and Lewis (1975, 1973) among others. Because we can perform such reasoning almost instantaneously, there of course exists a subset of humans who are interested in studying this phenomenon. Linguists, especially, take great interest in studying what seems intuitive. By extension, computational linguists (who arguably subset linguists) are interested in collecting structural representations of commonsense human reasoning in the hopes of automating such in machines. The desired structures of the hoped-for data are specified in ANNOTATION SCHEMES – a "rigorously defined system of guidelines for layering interpretive information on top of the text" (Dunietz, 2018). In order to contribute to the study of causal language, this project aims to provide evidence for a high-quality annotation scheme designed for understanding textual manifestations of causal relations.

This thesis thus begins with an overview followed by three parts. I first begin by motivating our work through a review of past theoretical and computational linguistic work in causal language. Then, I introduce the CCEP (Constructions of CAUSE, ENABLE, and PREVENT) Annotation Scheme which builds mainly upon Wolff et al. (2005)'s model of causal language and Dunietz (2018)'s BECauSE corpus. Finally, I present results from a preliminary corpus of 150 manually annotated documents that lead us to future directions.

1.2 Motivation

The analysis of causal language has many applications. As humans, causal attribution provides the basis for many of our beliefs, inferences, and judgements. Consider that use and interpretation of assertions about causal relationships provide the basis for inferences we make as we obtain world knowledge. Furthermore, our daily interactions give rise to our own questions about reasons and explanations of political events, social interactions, academic knowledge, and *passing responsibility* based on causal attribution. Consider Pearl (2009)'s example of such from the Bible:

> When God asks: "Did you eat from that tree?"
>
> This is what Adam replies: "The woman whom you gave to be with me,
> She handed me the fruit from the tree; and I ate."
>
> Eve is just as skillful: "The serpent deceived me, and I ate."

As textual documents (such as the Bible) provide straightforward records of causal relationships, or at least what we as humans believe them to be, it is only natural that those interested in causal reasoning would utilize them as a resource for understanding causation in everyday occurrences.

From former work on creating datasets of causally annotated text, I identify three main limitations that I aim to improve upon in this thesis:

Simple counterfactual tests do not fully capture CEP categories

Most of the previous annotation schemes that aim to categorize causal relations into CAUSE, ENABLE, and PREVENT used simple counterfactual tests to discern between them. For example, consider CaTeRS' (discussed in Section 2.2) definitions of CAUSE, ENABLE, and PREVENT concepts using the following definitions:

- A CAUSE B: In the textual context, if A occurs, B most probably occurs as a result.
- A ENABLE B: In the textual context, if A does not occur, B most probably does not occur (not enabled to occur).
- A PREVENT B: In the textual context, if A occurs, B most probably does not occur as a result.

Evidently, these definitions are only concerned with one facet of the CEP categories – namely, necessity and sufficiency. However, and for good reason, Wolff et al. (2005) does not define *causal necessity* as a defining attribute of ENABLE and *causal sufficiency* for CAUSE or PREVENT. Not only are the notions of sufficiency and necessity a point of contention in literature (see Lauer and Nadathur (2020); Baglini and Siegal (2020); Bar-Asher Siegal and Boneh (2019)), but these hypothesized characteristics of CEP only arise as a byproduct of the core attributes of CAUSE, ENABLE, and PREVENT as shown in Table 2.2.

I admit that in accordance with previous CEP labelling projects, counterfactual reasoning was my own original approach to causal role labeling as discussed in Section 2.3. For good reason, this approach did not lead to satisfactory inter-coder agreement and so I advance to an improved methodology for our CCEP corpus.

Causal language encapsulates a wide variety of lexical items

Other previous work in annotation of causal language ties causal meaning to limited *triggers*. For example, the Penn Discourse Treebank's (PDTB) triggers are limited to conjunctions and adverbials (Prasad et al., 2008, 2006). Likewise, the PropBank scheme (Kingsbury and Palmer, 2003; Bonial et al., 2014) limits its annotation of causal language to the arguments of verbs. Thus, a richer representation of causal language enabled by a wide variety of identified triggers would improve the field's understanding of textual causal phenomena.

News datasets are not primed for causal relation annotation

This claim is supported by our finding that news article annotation actually led to the lowest IAA, albeit not by a significant amount ($p > 0.05$). Despite this, most previous causal annotation schemes were annotated on news articles because of their wide availability.

1.3 Thesis statement

This thesis provides methods for representing causal relations in cross-genre text documents with the goal of improving the three formerly identified motivations. I hypothesize that with the "*surface constriction labeling*" (SCL) approach to annotating, we will be able to manually annotate corpora with richer representations of the CAUSE, ENABLE, and PREVENT relations than are currently available. The SCL approach to annotation is grounded in CONSTRUCTION GRAMMAR, which argues that meaning is intrinsically tied with surface form (discussed in Section 2.4). I will demonstrate this hypothesis in three steps:

1. I first **define an annotation scheme** in which causal linguistic constructions are pre-defined in the Construsticon – a list of possible identifiable constructions.

I adapt the Constructicon from Dunietz (2018).

2. Then, I **train annotators** to become proficient in this task through training videos, scored training quizzes, and practice documents.

3. Finally, we **annotate a corpus** following this scheme, which spotlights as its main feature the categorization of its annotated causal instances as CAUSE, ENABLE, and PREVENT-types.

Our corpus of 150 documents and 870 instances of causal constructions achieve F_1 and κ inter-annotator agreement scores comparable to Dunietz (2018)'s. This result demonstrates that it is indeed possible to embed nuanced CAUSE, ENABLE, and PREVENT categorization into causal annotation tasks.

In the remainder of this thesis, I discuss each of these steps in greater detail.

1.4 Summary of contributions

The key contribution of this thesis is implementing the SCL approach for enabling a finer-grained distinction between CAUSE, ENABLE, and PREVENT categories in document annotation tasks than currently available. Through aspectual CEP tests that are provided to annotators, our CCEP corpus demonstrates some interesting insights into how authors of different document types manifest causal relations. These contributions are embodied in publicly downloadable annotated corpus[1].

[1] Available in the future, since the conference I am currently aiming for has an anonymity period.

Chapter 2

Background

2.1 Linguistic analysis of causal language

Causal language has long been a source of interests for linguists. Consider the following instances of such:

1. I heated up the leftovers for dinner.

2. Last Saturday, Pete made me sleep on the couch.

3. I smooshed the bug flat.

Proposed categorizations have commonly included lexical causatives (as in 1), periphrastic causatives (2), and resultative constructions (3). Like resultative constructions, transitive sentences that use a lexical causative lack an overt causative element. Bittner (1999) refers to these as 'concealed causatives'. They are distinguished in that the causing event is left implicit when using lexical causatives, while this is not so with resultatives. Consider 1, in which it is not clear if a microwave, oven, or some other method of heating was used for a warm dinner. In contrast, 3 makes explicit that a smooshing event led to the flattened bug. Finally, periphrastic causatives such as 2 denote an 'indirect' notion of causation not shared by the other two examples (Levin,

2019). Other works in linguistics seek to categorize periphrastic caustives based on notions of *sufficiency* and *necessity* (Lauer and Nadathur, 2020, 2018; Baglini and Siegal, 2020).

Cognitive approaches based on Talmy (1988)'s theory of force dynamics, such as Wolff et al. (2005) (which is of special interest to us), argue that causatives can be aspectually grouped into the types CAUSE, ENABLE, and PREVENT, or a combination of such. Computational projects such as Mirza et al. (2014); Prasad et al. (2006) have adapted this categorization when developing causally annotated corpora. Here, the notion of causation supersets the type CAUSE, which is then disparate from the lexical *cause*. Sloman et al. (2009) has linked ENABLE to *causal necessity* and PREVENT to a probabilistic relationship between cause and effect, which is similar to Lassiter (2018)'s argument that differentiates probabilistic indicatives from counterfactuals based on causal modelling. Building upon these works, Beller et al. (2020) provide a semantics for periphrastic causatives such as *affect*, *enable*, and *made no difference* using a causal judgement task. Similar cognitive work such as Gerstenberg et al. (2015, 2020), which are especially interested in how people make judgements about causation, propose that faulting a causal event takes into account both if the cause affects *how* the effect occurs as well as *whether* it did.

2.2 Annotation of causal language

Further efforts have been made to create corpora of causal relations. Causal language has long been of interest to linguists, cognitive scientists, and computational linguists. Cognitive approaches based on Talmy (1988)'s theory of force dynamics, such as Wolff et al. (2005), argue that causatives can be aspectually grouped into the types CAUSE, ENABLE, and PREVENT (CEP), or a combination of such. Consider Table 2.2 for Wolff et al.'s force dynamics characterization of such. As demonstrated in Table 2.1, 5 of

Annotation scheme	Manual annotation?	Pre-identified events?	Annotated temporal relations?	Annotated discourse relations?	Used CEP categories?
PDTB (Prasad et al., 2008, 2006)	✓			✓	
PropBank (Kingsbury and Palmer, 2003; Bonial et al., 2014)	✓			✓	
Causal TempEval-3 (Mirza et al., 2014)		✓	✓		✓
CATENA (Mirza and Tonelli, 2016)		✓	✓		✓
CaTeRS (Mostafazadeh et al., 2016b)	✓		✓		✓
Storyline Extraction (Caselli and Vossen, 2017)	✓	✓	✓		✓
BECauSE 2.1 (Dunietz et al., 2017b; Dunietz, 2018)	✓		✓		∗[1]

[1]BECauSE uses `Facilitate` and `Inhibit`, where `Facilitate` maps onto CAUSE/ENABLE and `Inhibit` to PREVENT.

Table 2.1: Characteristics of previous causal annotation schemes.

the 7 summarized schemes took inspiration from CEP categorization.

	Patient tendency toward result	Affector-Patient Concordance	Occurrence of result
CAUSE	N	N	Y
ENABLE	Y	Y	Y
PREVENT	Y	N	N

Table 2.2: Defining CAUSE, ENABLE, and PREVENT according to Wolff et al. (2005).

Computational projects such as Mirza et al. (2014); Mostafazadeh et al. (2016b) have adapted this categorization when developing causally annotated corpora. Here, the notion of causation supersets the type CAUSE, which is then disparate from the lexical *cause*. Sloman et al. (2009) has linked ENABLE to *causal necessity* and PREVENT to a probabilistic relationship between cause and effect. Building upon these works, Beller et al. (2020) provide a semantics for periphrastic causatives such as *affect*, *enable*, and *made no difference* using a causal judgement task. Similar cognitive work such as Gerstenberg et al. (2015, 2020), which are especially interested in how people make judgements about causation, propose that faulting a causal event takes into account both if the cause affects *how* the effect occurs as well as *whether* it did. Consider the following examples of corpora that annotate for causal relations.

PropBank In PropBank Bonial et al. (2010), causal discourse relations are annotated by predicate and argument, where ARGM-CAU is used to annotate "the reason got an action", for example: "*they* [PREDICATE ***moved***] *to London* [ARGM-CAU ***because of the baby***]" (Mirza et al., 2014). This is dissimilar from our approach as it is not fixated on implicit or explicit connectives.

CaTeRS More recently, the semantic annotation framework titled Causal and Temporal Relation Scheme was developed by Mostafazadeh et al. (2016b), which applied the scheme to 320 5-sentence short stories sampled from Mostafazadeh et al. (2016a)'s ROCStories corpus. The CaTeRS framework annotated causal and temporal

relations simultaneously, but was limited in that their scope of lexical causality could only capture events as argument primitives (Davidson, 1967).

2.2.1 The BECauSE corpus of causal language

The BECauSE 2.1 corpus of causal relations implements the BECauSE annotation schema developed by Dunietz (2018); Dunietz et al. (2015, 2017b). It includes annotated relations of 59 articles from Washington section of the New York Times (NYT) corpus (Sandhaus, 2008), 47 Wall Street Journal (WSJ) documents from the Penn Treebank (Marcus et al., 1994), 12 documents from the Manually Annotated Sub-Corpus (Ide et al., 2010), and 772 sentences transcribed from Congress' Dodd-Frank hearings (Smith et al., 2014). The causal relations in this combined corpus were annotated based on pre-identified connectives, which directed ARGCs (causes) to ARGEs (effects). The scheme also allows the presence of overlapping relations, such as `Temporal`, `Correlation`, `Hypothetical`, `Obligation/Permission`, `Creation/Termination`, `Extremity/Sufficiency`, and `Context`, which are all relations that frequently co-occur with causation. Notably, Dunietz expresses a desire to attempt more fine-grained distinctions based on Wolff et al. (2005)'s aforementioned categories, as well as extending their annotation schema to other relation types such as CONCESSION and COMPARISON.

It is from the BECauSE corpus that I take my widest inspiration in creating the CCEP scheme, although I am not concerned with non-causal relations. In fact, as I discuss in Section 2.3, since the CCEP scheme is part of a larger multi-layered semantic annotation project that already includes temporal and coreferential annotation.

2.3 A preliminary study

I refer to the first attempt at this project as Discourse Relation Annotation (DRA). The DRA project encapsulated two goals:

1. Annotate document-level discourse relations using simplified roles from the Penn Discourse Tree Bank (PDTB) (Prasad et al., 2008, 2006),

2. expand upon the `:cause` role in the PDTB to include the categories of CAUSE, ENABLE, and PREVENT, and

3. contribute to a larger multi-layered semantic annotation project (including temporal and coreference) through consistency in annotation subject and guideline design.

Although my DRA guidelines went into more detail about the nature of these roles, here I provide a tabular summary:

It is clear to me now, as it may be to readers, that the roles in these guidelines were poorly defined. Although the full guidelines did contain more specificity than space allows for in the Table 2.2, the scores shown in 2.4 are clearly abysmal. The F_1 metric is commonly used in evaluating annotation work, with discussions of annotator agreement requiring a satisfactory minimum of 0.80 (Bayerl and Paul, 2011), which the scores in Figure 2.4 are nowhere near. F_1 is defined as $2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$, where precision is the number of true positives over the total number of positives, while recall is the number of the number of true positives over the total number of relevant elements. For example, consider a search engine that returns 20 results of which 15 are relevant, while failing to return 10 other relevant results. Then, precision $= \frac{15}{20} = \frac{3}{4}$ and recall $= \frac{15}{25} = \frac{3}{5}$. This calculation is algorithmized in Figure 2.

From our first attempt at annotating causal and discourse relations, I came to the following conclusions:

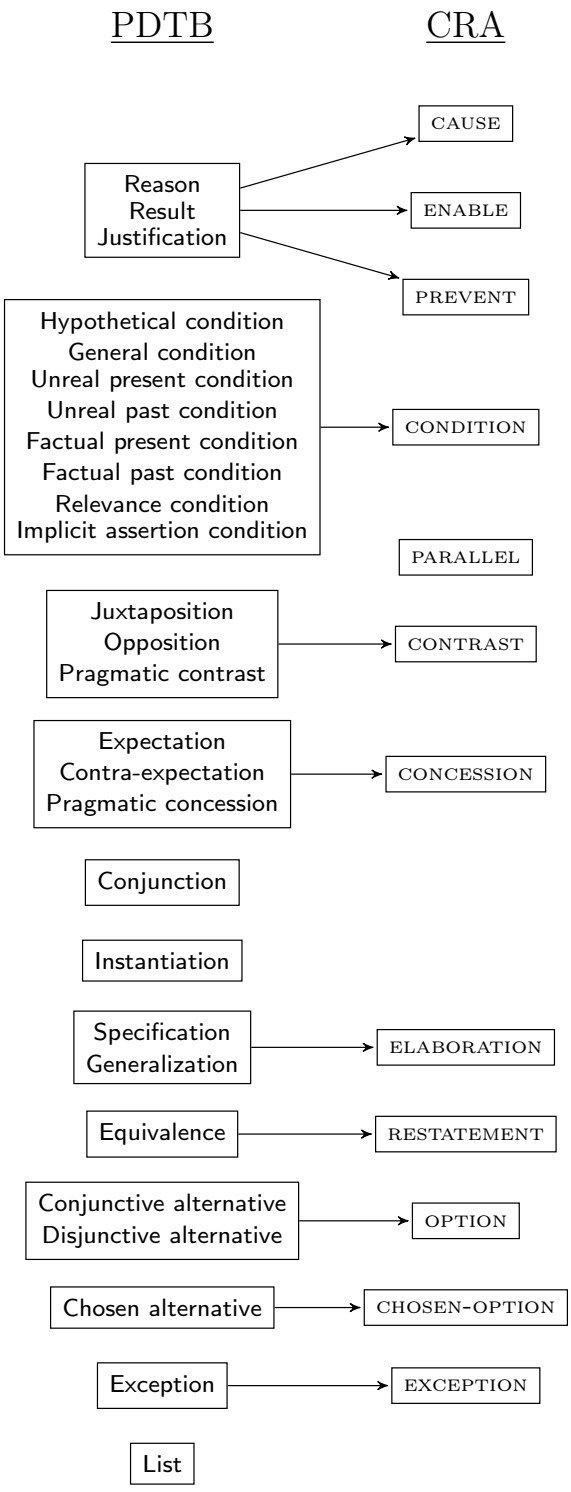

Figure 2.1: Comparison of PDTB to CRA roles.

Role	Situation	Annotation	'Alert' items
:cause	If A did not occur, B would not have either.	(A → B)	'because', 'so', 'then', 'and', 'for', 'thus', 'after', 'the reason', 'only when', 'since', 'in order to', 'then'
:enable	If A did not occur, B may have occurred anyway.	(A → B)	'so', 'then', 'and', 'affect', 'also', 'a reason', 'one of the reasons', 'then', 'lead'
:prevent	A reduces the likelihood of B occurring.	(A → B)	'stopped'
:condition	Only when A happens, can B happen.	(A → B)	'if'
:contrast	A and B share a significantly different predicate or property.	(A ↔ B)	'not like', 'opposite', 'however', 'more', 'less', 'than'
:parallel	A and B share a significantly similar predicate or property.	(A ↔ B)	'like', 'similar', 'reminds', 'likewise'
:concession	A should result in the occurrence of C, but B indicates the opposite of C occurring.	(B → A)	'although', 'but'
:exception	A is an exception to B; B is not true because of A.	(A → B)	'except', 'actually'
:elaboration	A is a more detailed description of B.	(A → B)	'specifically', 'actually', '—'
:restatement	A restates the semantic meaning of B.	(A → B)	'like', 'including', 'such as'
:option	A and B are alternative situations of each other.	(A ↔ B)	'or', 'also', ','
:chosenoption	A and B are alternative situations of each other, but the subject prefers A.	(B → A)	'preferred', 'would rather', 'want'

Figure 2.2: Summary of roles from the first DR guidelines.

1. Document-level relations often span the whole text

2. Large number of relations

3. Relations are quite conceptual, abstract, and/or subjective

4. Lack of training

With these problems of my Discourse Relation Annotation Guidelines in mind, I address each one in my revised annotation project, which I guide with our CCEP (Connectives of CAUSE, ENABLE, and PREVENT) Annotation Guidelines.

	Spans (F_1)	Pairs (F_1)	Relations (F_1)
Round 1 (Reddit)	0.482	0.249	0.160
Round 2 (Reddit)	0.541	0.268	0.192
Round 3 (Fables)	0.524	0.331	0.266

Figure 2.3: Our first attempt at annotating discourse and causal relations.

To first address point (1), our Discourse Relation Annotation Guidelines aimed to annotate discourse and causal relations at the document-level. This would have been a novel contribution indeed, as previous projects have focused on sentential or inter-sentential relations; for good reason, as I soon discovered. Causal relations, which are largely iterative in discourse, (i.e. *I went to the store to buy milk, but this caused me to miss my dentist's appointment, and now I have an unfilled cavity*) are further complicated when they appear in disjoint spans throughout a text. This is also significantly reliant on annotator's interpretations as they were not lexically grounded (the 'Alert' items were only a suggestion). In order to address this problem in our new CCEP guidelines, I adopt the *surface construction labeling* (SCL), which ties semantic relations to their surface form. I **implement SCL through the Constructicon**, which is discussed in Section 3. By doing so, I limit inter-coder disagreement that arises from difference of interpretation. SCL is discussed in more detail in Section 2.4.

Regarding point (2), previous studies such as Bayerl and Paul (2011) found significant negative relationships between a larger number of categories and inter-annotator agreement. Since this quite an easy problem to address, I approach this problem by only **implementing our three causal roles** in my new guidelines while disregarding discourse relations.

Per point (3), this problem is evident from the "situation" column in Table 2.2. While this was originally done intentionally to abstract away from lexical "triggers" (in order to enable document-level relations), this was a major problem for causal role annotation since the concepts of CAUSE, ENABLE, and PREVENT are quite abstract to begin with. In order to address this problem, I do away with simple counterfactual tests

(as presented in the 'Situation' column) in favor of **more nuanced categorization enabled by the Causal Relation Decision Tree** (see Figure A.1). This tool aims to ground causal categorization in concrete tests of traits of the CEP categories, instead of asking annotators to reason about the abstract categories by themselves.

Concerning point (4), as discussed in Section 4.1, I **implemented rigorous training and standards** that annotators were required to successfully pass before beginning actual annotation (excluding our pilot round discussed in Section 2.4). This including an introductory video, one hundred training questions contained in ten quizzes, and successful annotation of ten practice documents.

2.4 Advancements

From our previous discussion of our first annotation attempt, it is clear that most problems arose from an ambiguous and overly abstract guidelines. To combat this problem in our revised guidelines, I implement a "surface construction labeling" (SCL) approach, which is grounded in the theory of Construction Grammar (CxG). This was used by Dunietz (2018) in creating the BECauSE corpus.

Construction Grammar primer

Our project grounds its understanding of meaning in Construction Grammar (CxG). Construction Grammar takes the fundamental units of language as *constructions* (Goldberg, 2013). This approach to semantics "pairs patterns of surface forms **directly with meanings**," (Dunietz, 2018; Dunietz et al., 2017a) which contrasts with the Chomskyan tradition of Transformational Grammar (TG). Theories grounded in TG make emphasis on the difference between a sentence's Deep and Surface-structures (abbreviated as D and S-structures) as syntactically grounded *transformations* which act as multivalued functions mapping D-structures onto S-structures. This has been

referred to as *infinite use of finite means*, attributed to Humboldt and discussed in Chomsky (2015). For example, Wittgenstein (1953) notes that the command "Slab!" can mean the same thing as "Give me the slab," despite their different externalizations which were arrived at through different transformations applied to the same D-structure.

CxG stands in contrast to theories built upon TG. Instead, CxG ties the semantics of an utterance directly with its meaning. This was born from the realization that previous attempts to formalize grammar have done so by largely ignoring language use in the real world (consider, for example, Chomsky's *ideal speaker*). Thus, CxG places more importance on complex patterns of larger phrases than on individual morphological or syntactic elements. Per this theory, every conventionalized linguistic pattern, is just another *construction*, which may "include open slots that can be filled by other words or constructions – i.e., constructions can be *linguistically productive*" (Dunietz, 2018). To support this, empirical studies such as Tomasello (2001) have concluded that children create novel utterances not through transformations, but by modifying utterance-level schema through identifying 'slots' where conventionalized items may fit into (what he refers to as "cut and paste" operations). In accordance of this theory of usage-based syntactic operations, the use of the construction "Slab!" to mean "Give me the slab" thus arises from the speaker's intention to reproduce an entire goal-directed act simply through an alternative construction with regard to potential pragmatic constraints.

A methodological implementation of CxG in annotation work is *surface construction labeling* (SCL) which addresses many concerns that arose from the DRA scheme. By tying surface form with function, I argue that causal relations in language should be easily observable through specific lexical constructions. This manifests in the Constructicon, which is tool for annotation I discuss in greater detail in Sections 3 and A. The Constructicon provides to annotators 194 pre-identified causal construc-

tions, which they are to annotate manifestations of strictly by-the-book. Since many constructions are ambiguous between causal and non-causal ones, I provide the Causal Relation Decision Tree (Figure A.1) to aid annotators.

Pilot annotation round for CCEP

Prior to beginning our newly proposed training process for the CCEP guidelines, I had a pilot annotation round of which the results are summarized below. Annotators included myself and one other. This included annotation of five Reddit documents with word counts between 150 and 350, exclusive.

File	Spans (F_1)	Relations (F_1)
00	0.786	0.769
01	0.828	0.800
02	0.593	0.556
03	1.000	1.000
04	0.667	0.667
Average	0.775	0.758

Figure 2.4: Our first attempt at using the new CCEP guidelines.

While these scores are still below Bayerl and Paul (2011) "rule of thumb" of 80%, they are a significant improvement from our previous performance using the DRA Guidelines. I move forward after some revision of these guidelines with the expectation that the rigorous training process would raise agreement scores to a sufficient threshold.

Chapter 3

The CCEP Annotation Scheme

The Constructions of CAUSE, ENABLE, and PREVENT Annotation Scheme contains two subcomponents – the annotation guidelines and the Constructicon. Both are adapted from Dunietz (2018). The purpose of this section is highlight the main points of both the Constructicon and the Annotation Scheme, while motivating the design of both with insights into the role of annotators' cognition while annotating.

3.1 Our working definition of "causal language"

To define "causal language" within the CCEP scheme, I first briefly discuss theories of causal reasoning to understand annotators' decision-making processes while annotating. According to the focal sets models of causation, causal relations are inferred on the basis of *covariation*, which describes when two events are likely to co-occur (Cheng and Novick, 1991; Cheng, 1997; Cheng and Novick, 1992). Covariation is calculated by subtracting the probability of an effect e, in the presence of a candidate cause c, from the probability of the effect in the absence of c. This is defined as $\triangle\ P = P(e \mid c) - P(e \mid \neg c)$. A causal relation is inferred when $P(e \mid c) > P(e \mid \neg c)$. For example, the probability of cancer in the presence of smoking is greater than in the absence of smoking, licensing the statement "smoking causes cancer" (Wolff

and Song, 2003). It stands to reason, then, that while many textual occurrences of events may trigger annotators' cognition as being *causally related*, annotators do not annotate all such pairings of events.

To be more specific, annotating instances of "causal language" within the CCEP scheme refers to annotating arguments of clauses or phrases in which one event, state, action, or entity (the Cause) is **explicitly presented as** promoting or hindering another argument (the Effect). So, our primitive spans of arguments refers to propositional clauses or phrases that reference an entity and/or action. The Cause and Effect must be textually connected through an explicit trigger, which I refer to as the "connective".

As a consequence of this requirement for a lexical trigger, our work refrains from annotating:

1. Causal relationships with no lexical trigger: I.e., *A robber set upon them. They ran away.*

2. Connectives that encode the means or the result of the causation: I.e., *kill* can be interpreted as *cause to die*, but since this encodes the result, we do not annotate it. Furthermore, *paint* in *John painted the house red* encodes the Means, and so these guidelines exclude it.

3. Connectives that assert an unspecified causal relationship: I.e., *linked* in *Smoking is linked to cancer* does not explicitly specify the direction of causation, so annotators should not annotate it.

The decision to exclude the above instances of causal languages was made so that I may focus specifically on language that only expresses causation. If I were to include the above instances, the majority of transitive verbs in the English language would be considered causal, and it would be close to impossible to disentangle causation as a semantic phenomenon from its means or effect.

3.2 Parts of a causal instance in annotation

Annotation instances are triggered by the appearance of a causal connective, which relate up to three other spans of text and of which any may be disjoint. Consider the following descriptions:

- **The causal Connective**: The Connective acts as the basis of all annotation instances and which signifies the possibility of a causal construction when it appears in text. Some examples include *for... to*, *because*, and *after*.
- **The Cause**: The Cause span is generally an event or state involving an entity ideally expressed as a propositional clause or phrase (as opposed to a set of words that do not form a coherent grammatical unit).
- **The Effect**: The span of the Effect is generally an event or state, ideally expressed as a propositional clause or phrase.
- **The Means**: The Means span includes an action which serves the purpose of differentiating between the agent of the Cause and the action by which that agent induces the Effect, both of which may be disparately present in text. For example, *dropping a lit match* in *I caused a fire by dropping a lit match* would be annotated as the Means.

In an annotation instance, it is possible for the Cause, Effect, or Means to be absent, although at least one of the Cause or Effect must be present for it to be annotatable.

3.3 The Constructicon

Possible causal connectives that annotators may annotate are all pre-identified in **the Constructicon**, of which a portion is available to view in Appendix B. This tool is

available to annotators as a searchable spreadsheet of 194 causal connective patterns, and was designed to minimize the decision-making burden placed on annotators. Consider the following examples of such, of which verbs are given in present tense and nouns/adjectives are given as copulas for readability):

- for <Effect> to <Effect>, <Cause>
- <Effect> because <Cause>
- <Cause>, so <Effect>

The Constructicon also includes five other columns per entry, detailed below:

1. **Variants**: Variants of annotatable connectives are specified here. For example, variants of *<Cause> forces <Effect>* include *<Cause> forces <Effect> to <Effect>* and *<Cause> forces <Effect> into <Effect>*.

2. **Word(s) to annotate as connective**: This column is especially useful for alignment issues as it explicitly identifies what needs to be present in order for a connective to be annotatable. For example, for the *<Cause> forces <Effect>* entry, this column specifies *force* and optionally *to* and *into* to be annotated as the connective.

3. **Type**: The 'Type' column is adapted from Dunietz (2018)'s categorizations of Degrees. As I will discuss in Section 3.4, the CCEP identifies three types of causal relations associated with each connective – CAUSE, ENABLE, and PREVENT. This column is especially of use to annotators because it specifies connectives to be either PREVENT or CAUSE/ENABLE-type, and thus annotators will explicitly be told when a causal relation is PREVENT-type. The Constructicon holds 151 CAUSE/ENABLE constructions and 25 PREVENT constructions.

4. **Comments**: Comments allow connective-specific directions to be communicated to annotators. For example, a frequently appearing construction is *<Effect> to*

<Cause>. The comments for this entry specify that the instance should only be annotated when it can be paraphrased as "in the hopes of," or "with the goal of."

5. **Example(s)**: As is self-explanatory, this comment provides examples for the respective causal construction. For instance, this column for the *<Cause>; therefore, <Effect>* construction includes *Therefore, we should consider how to most appropriately give the Fed the necessary authority.*

3.4 Types of causation

While Dunietz focuses on causal categories of (1) Purpose, Motivation, and Consequence, as well as (2) Facilitate and Inhibit I aim to extend the applicability of his tools to categorize CAUSE, ENABLE, and PREVENT, which is a more nuanced exploration of his second dimension. Dunietz (2018) did originally aim to have a 3x3 categorization including CAUSE, ENABLE, and PREVENT; unfortunately, they were unable to reach satisfactory IAA scores in formative attempts. Their solution was to collapse CAUSE and ENABLE into Facilitate, leaving PREVENT to map to Inhibit.

Recall our preliminary discussion of Wolff et al. (2005)'s causal theory of CAUSE, *enable*, and *prevent* in Section 2, of which a physical model of Table 2.2 is provided in Figure 3.1. These vector diagrams represent the various forces present in a causal relation, according to Wolff and Zettergren (2002). Recall that the dot product $A \cdot B$ measures the length of A's orthogonal projection onto B. So, the prototypical case of a patient having a *tendency* towards the endstate is defined as when the angle θ between P and E is $0°$. Similarly, an agent and patient act in *concordance* when the angle between P and A is $0°$. If I instead take $\theta = 90°$ as in CAUSE, $A \cdot B = 0$ because the vectors are already orthogonal. Finally, the result *occurs* when the angle between R and E is $0°$, where R is the vector sum of P and A. In PREVENT relations, R$\neq$ E, and so the endstate preferred by P cannot occur.

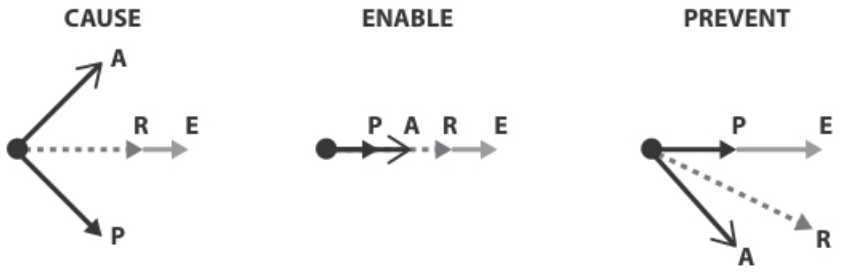

Figure 3.1: Representation of CAUSE, ENABLE, and PREVENT categories from Wolff (2007), where A = affector force, P = patient force, R = resultant force, and E = endstate.

While CEP are not disjoint categories and should be thought of as a scale of attributes, since forces of causation are very rarely at 0° or 90° relative to each other, the CCEP guidelines makes every attempt to distinguish one from the other in order to eliminate ambiguity for annotators. But first I provide more specific discussion of these categories according to Wolff et al. (2005), who writes that "Differences among the concepts are captured in terms of various patterns of tendency, relative strength, rest, and motion between an *affector* and a *patient*."

3.4.1 CAUSE

A is a CAUSE of B if (1) The patient does not have a tendency for B, (2) The affector and patient do not act in concordance, and (3) The result B actually occurs.

Consider the following examples of such:

1. Strong winds caused the bridge to collapse.

2. The grating noise eventually made my ears bleed.

3. That forced me to drop out of school.

In Example 1, readers may infer that (1) *the bridge* does not have a tendency to collapse, and (2) since the *strong winds* influence *the bridge* to collapse, that the *winds* and *bridge* do not act in accordance, and (3) *the bridge* actually did collapse.

3.4.2 ENABLE

A ENABLES B if (1) The patient has a tendency for B, (2) The affector and patient act in concordance, and (3) The result B actually occurs.

Consider the following examples of such:

1. Vitamin B enables the body to digest food.

2. My mother let me come to the dance.

3. I'll allow it.

In Example 2, it is apparent from it's textual context that (1) the narrator wanted to go to *the dance* in the first place, (2) the *mother*'s actions aligned with this want of the narrator's, and (3) the narrator does actually go to *the dance*.

3.4.3 PREVENT

A PREVENTS B if (1) The patient has a tendency for B, (2) The affector and patient do not act in concordance, and (3) The result B has a reduced likelihood of occurring due to A.

Consider the following examples of such:

1. Corn oil prevents butter from burning.

2. My mother stopped me from leaving the house.

3. I dissuaded her from texting him.

In Example 3, it is apparent that (1) the entity referred to as *her* wanted to text *him*, (2) the narrator takes action to realize their own want of *her* to not text *him*, and (3) the entity *her* does not end up texting *him*.

3.4.4 Differentiating CEP while annotating

To differentiate between CEP while annotating, I provide annotators with the Causal Relations Decision Tree (CRDT) as depicted in Figure A.1 of the Appendix. This flowchart is meant to ground the notions of CAUSE, ENABLE, and PREVENT so that annotators are not burdened with the task of understanding abstract concepts. While these tests are not fool-proof, they do systematize intuitions that previous researchers have solidified about the concepts.

Consider the following sentence:

> A deer suddenly sprinted out of left field, causing me to stomp on the brakes.

Here, it is easy to select the three attributes presented by Wolff et al. (2005) in Figure 3.1 and Table 2.2.

- Patient tendency for result: No
- Affector-patient concordance: No
- Occurrence of result: Yes

Thus, I can finalize this as a CAUSE-type causal instance. However, there are many instances where selecting the three attributes is not so trivial. Consider the following sentence:

> For the United States to continue to lead the world's capital markets, we must continue to encourage innovation.

Here, the question cannot be answered: did the result of *continuing to lead the world's capital markets* actually occur? In order to standardize ambiguities like this one, in the CCEP guidelines, I do not ask annotators to rely on *patient tendency*

for result, *affector-patient concordance*, and *occurrence of result* to categorize the causal relations. Instead, I introduce the Causal Relation Decision Tree (as formerly discussed) in order to test for specific attributes of CAUSE, ENABLE, and PREVENT.

Recall that the Constructicon specifies to annotators when a connective is PREVENT-type. So, once a causal construction is identified, annotators need only to distinguish between CAUSE and ENABLE-types.

3.5 CAUSE vs. ENABLE

The central question for annotators of the CCEP scheme is to thus distinguish between instances of CAUSE and ENABLE. For a conceptual discussion of this difference, consider a world described by three variables:

- FF for forest fire, where $FF = 1$ if there is a forest fire, and $FF = 0$ otherwise;
- L for lightning, where $L = 1$ if lightning occurred and $L = 0$ otherwise;
- O for oxygen, where $O = 1$ if oxygen is present and $O = 0$ otherwise; (Halpern, 2016; Halpern and Pearl, 2013)

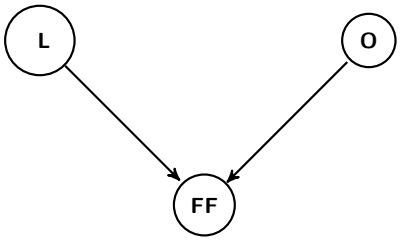

Figure 3.2: Graphical representation of FF example.

Why does *Lightning causes fire* sound fine while *Oxygen causes fire* does not (**?**Ni, 2012)? After all, it is clear from Figure 3.2 that both values of lightning and oxygen factor into whether the forest fire occurs. To distinguish the notion of CAUSE from ENABLE, I introduce *focal sets*. Focal sets are defined as contextually determined

sets of events that reasoners use when determining whether two events are causally linked (Cheng and Novick, 1991, 1992). A causal relation is perceived as ENABLING when the candidate causal factor is constantly present in the reasoner's focal set of events, which makes $P(e \mid \neg c) = \frac{P(\neg c|e) \times P(e)}{0}$ undefined since there would be no possible world in which $\neg c$ is the case. However, the candidate causal factor must also covary positively in another focal set. Referring back to our oxygen and lightning example, oxygen is present in both the case where a forest fire occurs and when it doesn't. However, in another focal set such as oxygen-free chambers in chemistry labs, the presence of oxygen does in fact covary with the hypothetical occurrence of a fire. Since oxygen covaries with fires in this other focal set of events, oxygen ENABLES rather than CAUSES fires.

However, since it would be unreasonable to ask annotators to reason through $P(e \mid \neg c)$ every time they came across one of the 161 CAUSE or ENABLE constructions in the Constructicon, I instead provide the following CAUSE vs. ENABLE tests:

Test 3.1. If the relation can be restated as "Cause {with the goal of / in the hopes of} Effect," is the Effect fully realized or only hoped-for? If it is only hoped-for, it is likely a CAUSE relation.

Test 3.2. Is the Cause presented as both necessary and sufficient for the Effect? If so, it is likely a CAUSE relation.

Test 3.3. Is the instance easily restated as "Cause enabled Effect" without changing the semantics? If so, it is likely a ENABLE relation.

Test 3.4. If the Cause did not occur, is the Effect presented as being able to occur anyway? If so, it is likely a ENABLE relation.

Test 3.5. If the Cause and Effect have agents, do the agents of the Cause and Effect act in agreement? If so, it is likely an ENABLE relation.

These tests are placed into the annotator's decision flow as depicted in Figure A.1, the Causal Relation Decision Tree. Note that these tests are meant to be ordered hierarchichally; passing test 3.1. holds more weight than passing test 3.5.

Test 3.1. is intended to capture causal relations of `Purpose`, i.e. *I tied my shoe so that it wouldn't fall off.* This may not intuitively seem to be a CAUSE-type relation, since as Dunietz (2018) observes, the causal chain is more something like:

A desires S and believes X will cause $S \longrightarrow A$ performs action $X \longrightarrow S$ may obtain

A statement of `Purpose` may be describing either $\longrightarrow$ relation on this chain. I argue that instances of `Purpose` are subset those of type CAUSE, because in the textual context in which an Effect manifests solely out of a desire to realize the Cause, the Effect and Cause do not act in concordance.

Test 3.2. may be controversial because literature such as Wolff (2007); Cheng and Novick (1991); Einhorn and Hogarth (1986) have noted that the concepts of CAUSE and ENABLE cannot be characterized in terms of causal necessity or sufficiency. However, I instate this test because (1) most annotators have laymen understanding of what necessity and sufficiency generally entail, and (2) causation of ENABLE-type requires at least two forces acting towards the same goal, while CAUSE does not. So while a single entity in an ENABLE relation may be necessary for the Effect to occur, it cannot be sufficient alone. Thus, the test follows this line of reasoning:

1. The annotator knows that the relation between the Cause and Effect is CAUSE or ENABLE (thus why the annotator is utilizing these tests).

2. If, in the context of the text, both the force in the Cause and the Effect prior to the Effect occurring tends toward the endstate occurring, then these forces

are both necessary and together sufficient for the endstate to occur (see next paragraph for a note on *necessity*).

3. Since the span of the Cause does not encapsulate both forces, its force by itself cannot guarantee the occurrence of the Effect (insufficiency).

4. Wolff's attribution of A {CAUSES/ENABLES} B when B actually occurs is weaker for ENABLE. Consider the following:

   (a) *I caused her to go to the store but she didn't end up going.

   (b) ?I enabled her to go to the store but she didn't end up going.

   While the cancellation in 4a is clearly unsuccessful, the felicity of 4b is less clear, especially if using another ENABLE verb such as *let*.

5. So when the Cause is alone insufficient for the Effect, it is more likely an ENABLE relation.

6. By contrapositive, if it not an ENABLE relation, then the Cause is sufficient for the Effect.

This follows from the conversation about INUS (Insufficient but Necessary alone, Unnecessary but Sufficient together) causation, which argues that only *sets* of events may be sufficient in a causal relation (Baglini and Siegal, 2020). However, while Baglini and Siegal (2020) discusses causation in a formal context, readers must recall that in the context of a small portion of annotatable text such as a Reddit document, Causes are generally presented as contextually necessary for the Effect to occur. Annotators are explicitly instructed not to reason beyond the context of the text. If the narrator of an annotatable text whines, "I failed the test only because the professor dislikes me," the span of *the professor dislikes me* is to be interpreted as the sole Cause that is sufficient for bringing about the failure, as the narrator so claims.

Test 3.3. arises from the observation that while not all instances of the use of lexical *cause* are of CAUSE-type, uses of *enable* are generally of ENABLE-type (for example, consider *A cause of her death were her poor eating habits*) (Lakoff and Johnson, 1999).

Test 3.4. arises from similar reasoning to the point made for Test 3.3., but holds for cases where a force relevant to the causal relation is not captured within the span of the Cause or Effect (i.e., mentioned elsewhere in the document). If all relevant forces act with the same endstate captured with the Effect in mind, it should be possible for one of the forces to account for the lack of another alternate force moving in the same direction.

Test 3.5. is similarly interested in cases where the agent of the Cause and Effect act in concordance, thus following Wolff's original attribution of ENABLE.

To conclude, while these diagnostic tests do not provide perfect mappings to Wolff et al. (2005) three-dimensional attributive classification of CEP, they do aid in standardizing notions of CEP for annotators in a way that still sufficiently retains the original notions of CAUSE, ENABLE, and PREVENT. This is demonstrated by our IAA scores in Section 4.2.1.

3.6 The annotation tool

To visualize the annotator decision process, consider interface depicted in Figure 3.3. This is the INCEpTION tool discussed in Section 4.1).

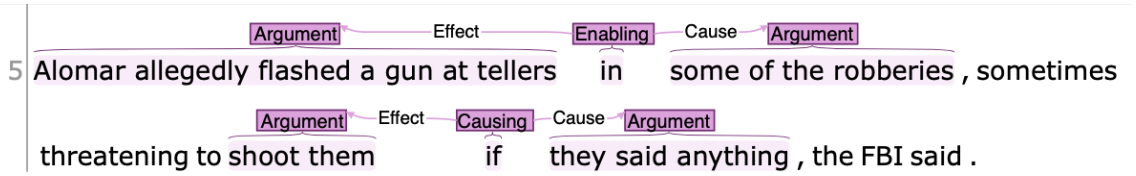

Figure 3.3: Sample annotation instances from cnn-41.txt.

The annotator of the Figure 3.3 would have first recognized that *in* and *if* are both causal connectives identified in the Constructicon. Then, they would have delimited

potential argument spans in accordance with the annotation guidelines. Finally, they would have followed the path of the CRDT in order to ascertain that (1) it is indeed a causal instance, and (2) the type of causation that the connective should be labeled as.

Chapter 4

The CCEP for Causal Language

In this chapter, I discuss the process of obtaining the CCEP corpus of annotated causal language.

4.1 Methodology

Data

The data was taken from three origins:

1. Aesops Fables from Project Gutenberg [1]
2. CNN data from cnn_dailymail corpus [2]
3. Reddit data[3] accessed on 14 February 2022

All data from these sources were first tokenized using the ELIT Tokenizer[4] and then filtered to a length between 100 and 200 tokens. This range was chosen in order to ensure a sufficient number of fables were available, since they are relatively short in length by default.

[1] https://www.gutenberg.org/cache/epub/21/pg21.txt
[2] https://huggingface.co/datasets/cnn_dailymail
[3] https://github.com/emorynlp/RedditData
[4] https://github.com/emorynlp/elit-tokenizer

Reddit posts were taken from popular college subreddits including r/College, r/GradSchool, r/CollegeRant, and r/ApplyingToCollege. Posts using profanity were removed using the Profanity-Check Python library[5]. Reddit data was also filtered to those that had > 5 comments.

Training

To ensure that annotators understand the guidelines and meet a standard of performance, they undergo extensive training prior to actual annotation. They are required to

1. **read the guidelines** and **view an instructional video**,
2. **take 10 online quizzes** consisting of 10 questions each on relation and span identification (see Appendix C for an example), and
3. **achieve satisfactory inter-annotator agreement** scores on 10 test documents.

I began the training process with three annotators (excluding myself), which consisted of two undergraduate students and a postdoctoral scholar who have all had annotation experience. Of these three, only one annotator achieved a $> 70\%$ average score across all 10 quizzes, and so I moved into the annotation process with one annotator other than myself.

Annotation

Annotators are also instructed to rotate through the various data sources in batches of 5 to ensure that any difference in IAA scores is not a result of familiarity with the annotation tool or experience following the annotation scheme.

[5] `https://github.com/vzhou842/profanity-check`

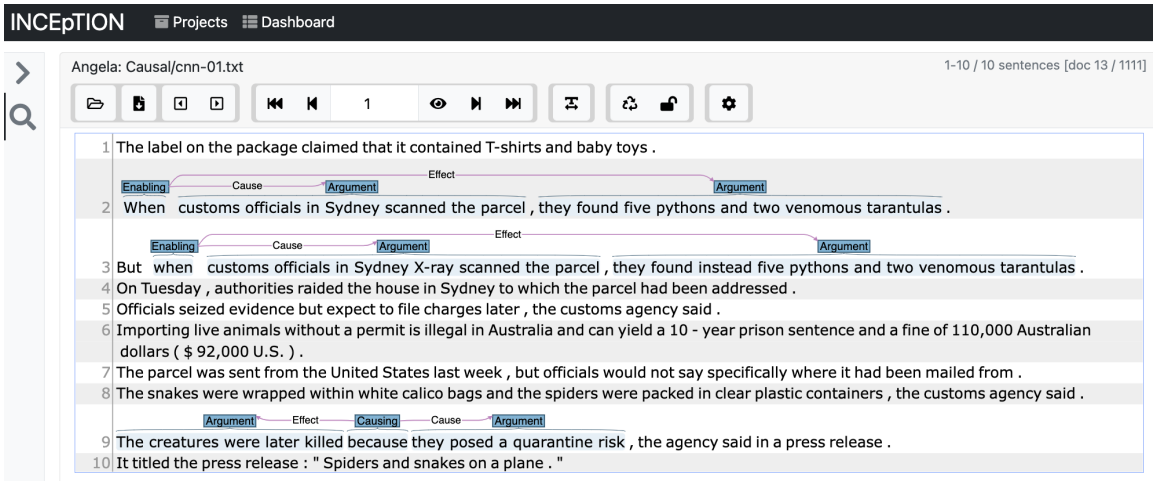

Figure 4.1: The INCEpTION annotator interface.

Annotation was done in the INCEpTION tool[6] developed by Technische Universität Darmstadt. This tool enabled our Causal Relation Annotation Project to be coordinated with two other parallel annotation projects – coreference and temporal annotation.

As demonstrated in Figure 4.1, the interface provides a lucid experience of selecting and labeling spans of text, and then choosing a relation between them.

4.2 Overview of the CCEP Corpus

4.2.1 Inter-Annotator Agreement

Our main motivation for using F_1 to measure span agreement and κ to measure causation type and argument labels was to be able to compare our performance to Dunietz (2018)'s, as shown in Tables 4.3 and 4.1. Since I briefly discussed the F_1 measure in Section 2.3, I now discuss the Cohen's Kappa (κ) score. κ is measured as $\frac{p_o - p_e}{1 - p_e}$, where p_o is the relative observed agreement among annotators and p_e is the hypothetical probability of chance agreement (Artstein and Poesio, 2008). Our implementation of F_1 is shown in Figure 2 as well as for κ in Figure 14.

[6] `https://inception-project.github.io/`

Annotation scheme	Relation types	Arguments IAA	Arguments metric	Connectives IAA	Connectives metric	Relation IAA	Relation metric	Corpus size
PDTB	1	0.90*	Percent	n/a	n/a	0.53†	F_1	2499 (news)
PropBank	1	0.93	Cohen's Kappa	0.93	Cohen's Kappa	0.91	Cohen's Kappa	2499 (news)
Causal TimeEval-3	3	n/a	n/a	0.55	F_1	0.3	F_1	20 (news)
CATENA	3	n/a	n/a	n/a	n/a	0.622	F_1	276 (news)
CaTeRS	9**	0.91	Fleiss' Kappa	n/a	n/a	0.51	Fleiss' Kappa	320 (stories) Mostafazadeh et al. (2016a)
StoryLine Extraction	2	n/a	n/a	n/a	n/a	0.638	Dice Coefficient	258 (news)
BECauSE 2.1	5	0.86‡	F_1	0.70	F_1	0.80	Cohen's Kappa	>116 (news) Sandhaus (2008) Marcus et al. (1994) Ide et al. (2010) Smith et al. (2014)

* Calculated for 3717 tokens.
† Only for CONTINGENCY relations.
** Only 4 of 9 are causal.
‡ Calculated for arguments (not including connective spans).

Table 4.1: Results from previous causal annotation studies.

Algorithm 1: Calculating κ scores, which is only done for agreed spans.

Assuming: length($annotator_1$) == length($annotator_2$)

number of agreement $\leftarrow 0$;
while x *in* $annotator_1$ **do**
  if x *in* $annotator_2$ **then**
    number of agreement ++ ;
  end
end
$p_o \leftarrow \frac{\text{number of agreement}}{\text{length}(annotator_1)}$;
for $y_{i...j}$ *in layers*[1] **do**
  $annotator_1$_probability $\leftarrow \frac{\text{total } y_{i...j} \text{ in } annotator_1}{\text{length}(annotator_1)}$;
  $annotator_2$_probability $\leftarrow \frac{\text{total } y_{i...j} \text{ in } annotator_2}{\text{length}(annotator_2)}$;
end
$p_e \leftarrow annotator_1$_probability + $annotator_2$_probability ;
$\kappa \leftarrow \frac{p_o - p_e}{1 - p_e}$;
return κ

[1]Here, *layers* refers to either the labels of spans (argument labels) or the types of causation identified with the connective (CAUSE, ENABLE, or PREVENT).

Algorithm 2: Calculating F_1 scores.

$t_p, f_p, f_n \leftarrow 0$;
while *x in annotator*$_2$ **do**
  if *x in annotator*$_1$ **then**
    t_p++ ;
  end
end
while *x in annotator*$_2$ **do**
  if *x not in annotator*$_1$ **then**
    f_p++ ;
  end
end
while *x in annotator*$_1$ **do**
  if *x not in annotator*$_2$ **then**
    f_n++ ;
  end
end
precision $\leftarrow \frac{t_p}{t_p+f_p}$;
recall $\leftarrow \frac{t_p}{t_p+f_n}$;
$f_1 \leftarrow 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision}+\text{recall}}$;
return f_1

As demonstrated in Figure 4.2, our overall corpus of causal annotations yields an F_1 score of 0.77 for Connective identification excluding cases of partial overlap, which is improved from the 0.70 of Dunietz (2018). This score of 0.77 also exactly matches the score achieved for BECauSE 2.0 as discussed in Dunietz et al. (2017b). For our

	Reddit	**News**	**Fables**	**Overall**
Spans (F_1)	0.81	0.74	0.72	0.75
Argument labels (κ)	0.93	0.86	0.91	0.90
Connective spans (F_1)	0.82	0.75	0.75	0.77
Types of causation (κ)	0.78	0.89	0.82	0.83

Table 4.2: Comparison of causal relation annotation performance on different text types using the same guidelines. κ indicates Cohen's kappa, which was only calculated for agreed spans (in line with Dunietz (2018)).

	IAA
Connectives (F_1)	0.70
Degrees (κ)	1.0
Causation types (κ)	0.80
Argument spans (F_1)	0.86
Argument labels (κ)	0.97

Table 4.3: Dunietz (2018)'s IAA scores excluding partially overlapping spans.

F_1 scores, I calculated the micro-average (also known as unitizing), meaning that there is no averaging over documents; rather, the annotations are concatenated into a single long document before scoring. This was due to the irregular appearances of connectives; while some documents contained upwards of a dozen instances of causal connectives, there were also 22 of our 300 doubly-annotated documents that did not have any annotations at all.

Furthermore, for agreed connective spans, the corpus also yielded a κ score of 0.83 for types of causation. This is similar to Dunietz' 0.80 for the causation catgories of Purpose, Motivation, and Consequence. However, our overall span score was lower than Dunietz' 0.86 at 0.75. This was likely due to argument length disagreement, as all three document types contained very different writing, ranging from the wordy rant-like style of Reddit documents to factual News reporting. This is in contrast to Dunietz (2018)'s corpus which only contained formal writing such as news reports and congressional hearings, allowing annotators to become used to one style of writing.

4.2.2 Statistics

The analysis of our annotated corpus provides some interesting insights. The corpus contains a total of 150 doubly-annotated documents, of which there were 870 annotations of causal constructions. Of these 300 annotated documents, 22 of them did not contain any annotated instances of causal language. Furthermore, note that a one-way ANOVA that compares macro-F_1's across different document types yields a p-value of 0.29 which is not significant. This demonstrates the robustness of our guidelines across genres, which included specific instructions for genre-specific idiosyncrasies such as the appearances of texting shorthands in Reddit posts.

As shown in Tables 4.4 and Figures 4.2 and 4.3, CAUSE-type instances dominated all instances of annotated causal language. This was to be expected – as shown in our CRDT, 3.2. which tests for CAUSE-type causation asks annotators whether the textual context *presents* the Cause as necessary and sufficient for the Effect. In the limited context of a 200-token document, many writers of these texts use causal language to identify and point out causal relationships, thus delimiting the Cause as contextually necessary and sufficient in some way for the Effect to occur.

	Reddit		**News**		**Fables**		
Category	n	**Percent**	n	**Percent**	n	**Percent**	**Total Count**
CAUSE	218	78.99%	182	71.94%	258	75.66%	658
ENABLE	56	20.29%	63	24.90%	77	22.58%	196
PREVENT	2	0.72%	8	3.16%	6	1.76%	16
Total	276	100%	253	100%	341	100%	870

Table 4.4: Counts of CAUSE, ENABLE, and PREVENT annotations in different text types.

Furthermore, consider Table 4.5, which depicts the most popular connectives across the different document types. While our findings generally align with Dunietz' counts of connective patterns in the BECauSE corpus (our most frequent five appear in his top seven), it is interesting to note that their frequencies vary across document types. For example, consider that the conditional only appears 8 times in the CNN documents,

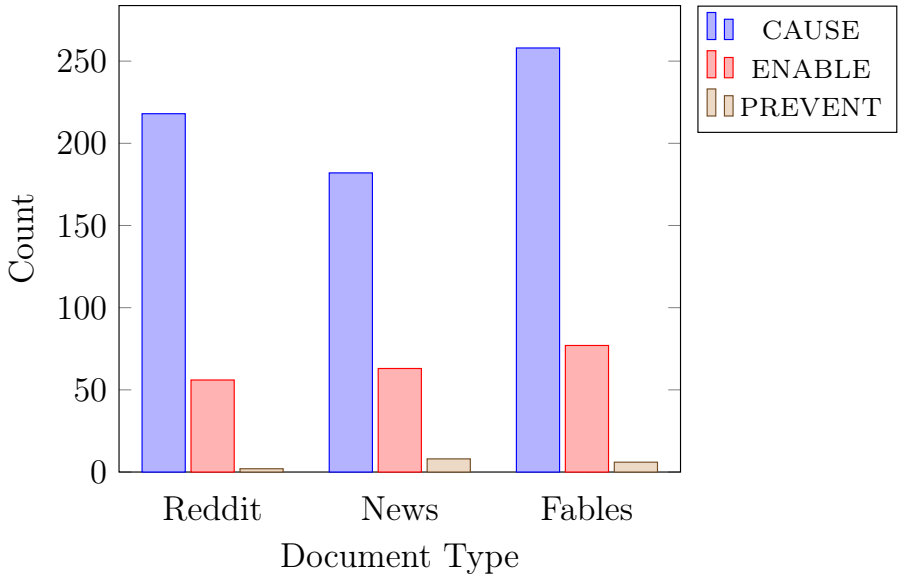

Figure 4.2: Visualization of causal counts presented in Table 4.4.

denoting the factual and unhypothetical nature of reporting news. Furthermore, consider that while “after" appears as our fourth most popular connective pattern, these appearances are actually solely dominated by the News (with a count of 41, compared to Reddits' 4 and Fables' 6). Finally, I observe that these most frequent five connectives account for approximately half of the instances of all annotated causal language in our corpus.

	Reddit		**News**		**Fables**			
Connective	n	**Frequency**	n	**Frequency**	n	**Frequency**	**Total** n	**Total %**
to	48	17.39%	24	9.49%	46	13.49%	118	13.56%
for	29	10.51%	30	11.86%	42	12.32%	101	11.61%
if	30	10.87%	8	3.16%	47	13.78%	85	9.77%
after	4	1.45%	41	16.21%	2	0.59%	47	5.40%
because	35	12.68%	4	1.58%	6	1.76%	45	5.17%
Total	146	52.90%	107	42.30%	143	41.94%	396	45.52%

Table 4.5: Comparison of popular connectives across different document types.

Finally, Table 4.4 is of interest to us because it demonstrates that Fables had the most annotations of causal language, while News contained the least. I hypothesize that this is because of the clear temporal and successive event-driven structure of Fables, which have been popularly used for temporal annotations for this reason

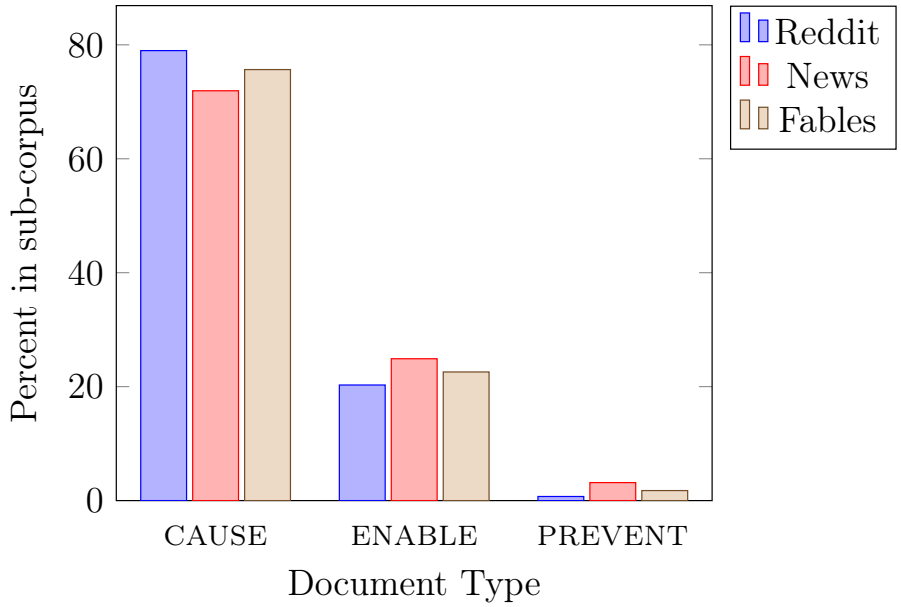

Figure 4.3: Visualization of normalized causal counts presented in Table 4.4.

(Bethard et al., 2012). The same reason may explain News' less frequent causal relations – News articles are more concerned with reporting states of affairs than explaining a sequence of causally related events to its audience.

4.3 Key findings

To conclude our analysis of the CCEP corpus, I summarize the main findings:

1. This project reached IAA scores of $F_1 = 0.75$ for overall spans, $F_1 = 0.77$ for connective spans, $\kappa = 0.90$ for argument labels, and $\kappa = 0.83$ of causation categorization for connectives.

    - Our results demonstrate an improvement in connective agreement from Dunietz (2018) but a drop in agreement for overall span agreement. The connective agreement matches Dunietz et al. (2017b) BECauSE 2.0 project which developed from the original BECauSE, but included overlapping relations.

- Our causation categorization score of 0.83 is comparable to Dunietz (2018)'s causation categorization IAA score of 0.80.

2. The most frequently annotated connectives in our corpus aligned with the most frequently annotated in Dunietz (2018)'s BECauSE corpus.

3. The sub-corpus of Fables demonstrated the most occurrences of causal language, while News had the least.

    - This is possibly because of the clear temporal and causal structure of short stories.

4. CAUSE-type relations dominated all three sub-corpora.

    - This aligns with findings from Mostafazadeh et al. (2016b)'s CaTeRS corpus.

Chapter 5

Future Outlook

5.1 Lessons learned

As discussed in previous chapters, my annotation project offers to readers takeaways for future annotation work.

I address that the SCL approach to annotation does make some aspects more difficult for annotators. Consider that many of the connectives in the Constructicon, such as *to* and *for* appear in non-causal usage. Thus, tying annotators to use these connectives as *triggers* may lead to some mistaken annotation of non-causal connectives.

Furthermore, causal relations may manifest in language through pragmatic interpretation rather than through a connective trigger. Consider the following sequence of events: "Two soldiers were walking down a road when a robber showed up. They ran away very quickly."

Finally, the SCL approach also prevents disjoint document-level causal relation annotation. Consider the following text taken from `reddit-030.txt`:

> Apparently I "plagiarized" about 65% of a bio lab report that I definitely wrote all by myself. My TA gave me a 0 but did give me the option of

> redoing the assignment in 24 hours. I am apparently at risk of getting reported to the university. My grade dropped a horrific amount. I'm pretty much being called a liar and a cheat. Happened to anyone else? I literally cried when my TA told me. This is my 3rd year at college and nothing like this has ever happened to me.

From my knowledge of the guidelines, the accusation of plagiarism identified in the first sentence is a CAUSE of the narrator "literally crying." However, this causal relation is not annotatable according to our guidelines because (1) it is not demarcated by a lexical connective, and (2) even with the insertion of a connective such as "so" before "I literally cried...", the hypothetical span of:

> I'm pretty much being called a liar and a cheat. Happened to anyone else? So, I literally cried when my TA told me.

is not enough to fit into the construction of *<Cause>, so <Effect>* as the leftward slot of "so" is not filled by the accusation of plagiarism.

5.2 Summary of contributions

To begin, my methodological contribution supports the "surface construction labeling" approach to annotation, which was introduced by Dunietz (2018). Through grounding relations in constructions, this project was able to attain high IAA. My artifactual contribution manifests in the freely available CCEP corpus, which I will continue to grow past this thesis as discussed in Section 5.3.2. Furthermore, this corpus contributes to the ongoing conversation about the aspectual categories of CAUSE, ENABLE, and PREVENT. While Wolff et al. (2005) emphasize that these groupings are not exclusive and are more a collection of attributes on a scale (which has been a source of criticism – see Ni (2012), among others) our high IAA scores affirm that these groupings are in fact measuring concepts that are psychologically real.

5.3 Future directions

5.3.1 Linguistic extensions

Specifically of interest to linguists may be future directions involving annotation using the SCL approach but in another dimension – for example, direct or indirect causation (Baglini and Siegal, 2020). Some have speculated that the cognitive distinction people make among causatives is not through CAUSE, ENABLE, and PREVENT, but rather through depths of causation (for example, compare *I got Sarah to go to the gym*, which suggests some intermediary action, versus *I killed Bob*).

In the realm of causality in language, it may also be of interest to have manual annotation of *normality* judgements of the Cause. Work such as Icard et al. (2017); Alicke et al. (2011) has suggested that people's judgements of actual causation is mainly attributed to how *normal* they regard events in the Cause or Effect span as, and an annotation task may be able to measure this at a large scale. However, the challenge for this sort of task may be from standardizing alignment rules, as annotators would also be tasked with annotating causal relations that may not at first glance seem causal, i.e. the *bacon* causing deliciousness in *This bacon pizza is delicious.*

On a more narrow scope, human annotation of gradable connectives such as *cause*, *got*, *make*, and *force* may provide empirical evidence for linguists to make distinctions between such.

5.3.2 Annotation extensions

Given that the CCEP corpus of 150 documents achieved high inter-annotator agreement, the natural follow-up to this thesis is to collect more annotations with the goal of obtaining enough data to train a shallow semantic parser to automatically tag instances of causal language in text. As a member of the Computational Linguistics group of the Emory NLP Lab, I have been training undergraduate annotators with

the goal of obtaining causally 1000 annotated Reddit documents by the end of the school year. Reddit documents are of special interest to us because of their widespread availability and conversational nature between posters, commenters, and repliers to these commenters. Annoting Reddit documents would be a novel contribution to the field because (1) most previous annotation projects have relied on corpora of News articles, and (2) semantic parsers trained on Reddit documents would be able to automatically tag more casual usages of language than those trained on annotated news articles, which are typically formal and carefully edited.

Beyond the constraints of the CCEP guidelines, it would also be of interest to those interested in the annotation of causal language to constructions involving punctuation rather than lexical items, such as the em-dash or semi-colon. Annotation of text written by the original author allows the unique advantage of every idiosyncratic punctuation being chosen for a reason.

I hope that future researchers interested in this field of study may find these suggestions useful.

Appendix A

The CCEP Annotation Guidelines

The purpose of this project is to generate consistent cross-sentential annotation of causal relations based on pre-identified connectives in the Constructicon and Wolff et al. (2005) categories of causation. This project is heavily inspired by Dunietz (2018) annotation guidelines for the BECAUSE (Bank of Effects and Causes Stated Explicitly) corpus.

A.1 Overview of causal linguistic constructions

We first define an *affector* as an entity that acts on another entity, and a *patient* as an entity that is acted on by another entity. Previous work in causal language has arranged causal linguistic constructions on a spectrum of specificity. At a high level, from least to most specific, this includes:

1. *Affect* verbs such as *affect*, *influence*, and *determine*, which specify only the occurrence of some change.

2. *Link* verbs such as *link to*, which differ from *affect* verbs in that they specify that a result was achieved.

3. Causal connectives and prepositions including *because*, *after*, and *when*.

4. Periphrastic (indirect) causatives such as *cause to* and *prevent from*, which are typically matrix verbs that take an object and clausal complement, sometimes add a specification of affector-patient concordance (i.e., *allow* specifies that the patient's tendency was towards the result encouraged by the affector).

5. Lexical causatives such as *kill* and bake which imply direct causation.

6. Resultative constructions such as *hammered the metal flat* which also incorporate the means of causation along with a periphrastic or lexical causative.

The majority of annotation instances indicated by the Constructicon is concerned with the middle of this spectrum. While much theoretical work is concerned with lexical causatives, we exclude these relations from our annotations. As with Dunietz (2018), this is because our definition of **Annotatable Causal Language** refers to clauses or phrases in which one event, state, action, or entity (the Cause) is **explicitly presented as** promoting or hindering another (the Effect). Both "explicit" and "presented as" are separately essential to our definition. The Cause and Effect must be deliberately related by an explicit trigger, which we term the Connective.

A.2 Annotatable units

- The **Causal Connective** is a word or series of words that encodes the causal relationship. Generally, the connective consists of a fixed construction with some open slots (e.g. *because*). Modifiers of the connective are not annotated (i.e., only *inhibits* in *severely inhibits* is annotated). Every causal connective should also be annotated with the instance's classification information – specifically, its degree of causation. The arguments of the causal connective, each of which may or may not be present, form the rest of each annotation instance. We define an **annotation instance** as each instance of a Causal Connective that has either

(or both) a Cause and Effect, and may also include a Means. The connective spans are pre-identified in the Constructicon.

- The **Effect** is the span that presents an outcome. It should be either a complete phrase or clause, though the clause may be non-finite (i.e., the subject may be missing and the verb may be a participle, gerund, or infinitive). Only in exception 4.2.9. below can the Effect argument not be a complete phrase or clause.

- The **Cause** is the span of text that presents an event, state, or entity that produces the effect. This should also be either a complete phrase or clause.

- The **Means** of causation is annotated when the activity by which the Cause (an entity or event) produced the Effect is also specified. I.e., in "The men kept the fire from spreading by clearing large ditches around it", *the men* should be annotated as the Cause, and *clearing large ditches around it* as the Means.

  Conceptually, the Means is a part of the Cause, but we annotate it separately because Means clauses make use of distinct linguistic machinery. Means arguments should only be annotated for:

  1. *By* or *via* clauses

     (a) [You] can [make] [it easier to eat home-cooked meals] by [meal-prepping on the weekends $_{\text{Means}}$].

     (b) [I] [guaranteed] [her the results she wanted] via [legal contract $_{\text{Means}}$].

     (c) By [always leaving the window open $_{\text{Means}}$], [she] accidentally [fostered] [the growth of mold].

  2. *How* phrases

     (a) [That $_{\text{Means}}$]'s how [I] almost [caused] [a war].

  3. *When* phrases

(a) [I] [caused] [a fire] when [I dropped the match $_{\text{Means}}$].

4. Certain other dependent clauses

(a) [I] [forced] [myself] [to] [study regularly in college] [through incentives like Starbucks $_{\text{Means}}$].

(b) [Singing loudly $_{\text{Means}}$, [she] [caused] [pain for everyone around].

As in examples 1.1.-1.3., *by* or *via* are not included in the span of the Means. Annotators may observe that *by* is also an entry in the Constructicon. *By* clauses thus include Means when the arguments are not in a causal relation – i.e., in example 1.1., *meal-prepping* is temporally instantaneous with *making it easier to eat home-cooked meals.*

In the case that any of the above instances of Means occur without a separate connective that does not trigger a Means, we do not annotate it. I.e., we would not annotate *By screaming too hard, she ruptured her vocal chords.*

A.3 Causation classification

The main task of this project is to classify annotation instances into three categories of causation: CAUSE, ENABLE, and PREVENT. To simplify this judgment task, we offer (1) the Constructicon, which pre-identifies the majority of connectives that *may* indicate the presence of a causal relation, and (2) the following tests.

1. **The "why" test**: After reading the sentence, could a reader reasonably be expected to answer a "why" question about the potential Effect argument? If not, it is probably not causal.

2. **The counterfactuality test**: Would the Effect have been just as probable to occur had the Cause not happened? If so, it is probably not causal.

3. **The ontological asymmetry test**: Could you just as easily claim the Cause and Effect are reversed? If so, it is probably not causal.

4. **The linguistic test**: Can the sentence be rephrased as "It is because (of) X that Y?" If so, it is likely to be causal.

Now, we discuss our subcategories of causation. Generally, there is a temporal constraint on causal roles; i.e. in A causes/enables/prevents B, some part of A must occur before B. The only exception to this case are Purpose cases (more on Purpose cases in section 4). Consider the following:

> I work a job because I need to pay rent and tuition myself.

Although it is not abundantly clear here whether "work" or the "need" occurred first, since we may reason that the sole purpose of *working a job* is because the patient *needs to pay rent and tuition*, we would annotate this as:

[I work a job $_{\text{Effect}}$] [because $_{\text{Causing}}$] [I need to pay rent and tuition myself $_{\text{Cause}}$].

A.3.1 Overview of categories

We include the below representations of CAUSE, ENABLE, and PREVENT to be of use to annotators.

	Patient tendency toward result	Affector-Patient Concordance	Occurrence of result
CAUSE	N	N	Y
ENABLE	Y	Y	Y
PREVENT	Y	N	N

Table A.1: Defining CAUSE, ENABLE, and PREVENT according to Wolff et al. (2005).

CAUSE

To reiterate, A is a CAUSE of B if:

1. The patient does not have a tendency for B
2. The affector and patient do not act in concordance
3. The result B actually occurs

Let's discuss this in layman's terms. Consider (1) in terms of this example:

> A deer suddenly sprinted out of left field, causing me to stomp on the brakes.

The patient in this situation is the narrator, who in the context of this piece of text, does not have a proclivity for *stomping on the brakes*; it is the action of the deer, the affector, which guarantees the result. Since the affector and the patient are at odds with each other, they (2) do not act in concordance. Finally, from the context of the situation, it is clear that (3) the result of *stomping on the brakes* actually occurs. Thus, this is a CAUSE relation. We would annotate it like so:

> [A deer suddenly sprinted out of left field $_{\text{Cause}}$], [causing $_{\text{Causing}}$] [me to stomp on the brakes $_{\text{Effect}}$].

Note that there cases of CAUSE that may be tricky to identify because of point 3. Consider the following statement, which depicts a Purpose instance of CAUSE:

> I searched for jobs in order to make money.
>
> [I searched for jobs $_{\text{Cause}}$] [in order to $_{\text{Causing}}$] [make money $_{\text{Effect}}$].

Annotators may see that it is not entirely certain that the Effect actually occurs. However, if they restrict their reasoning to only include possible worlds where *the*

effect actually occurs, it is easy to see that in these worlds, the job search was a CAUSE of making money.

For annotators' sake, we introduce several **CAUSE tests**:

1. If the relation can be restated as "Cause *{in the hopes of/with the goal of} Effect*", so that the purpose of Cause is to bring about the Effect, it is likely a CAUSE-type relation.

2. Is the Cause presented as both necessary and sufficient for the Effect? If so, it is likely a CAUSE-type relation.

ENABLE

We reiterate that A ENABLES B if

1. The patient has a tendency for B

2. The affector and patient act in concordance

3. The result B actually occurs

Consider the following example:

> Yesterday, Betty allowed me to go to the store.

Yesterday, [Betty $_{\text{Cause}}$] [allowed $_{\text{Enabling}}$] [me $_{\text{Effect}}$] [to $_{\text{Enabling}}$] [go to the store $_{\text{Effect}}$].

In the above example, *yesterday* is not annotated as it does not modify a particular argument, but rather the entire annotation instance. Now consider a similar version:

> Betty allowed me to go to the store, and so yesterday I did.

Note that since the Cause argument itself includes a causal instance, annotating this statement would entail annotating an embedded annotation instance (*Betty allowed me to go to the store*) within a larger annotation instance (only the outer instance is demonstrated below).

> [Betty allowed me to go to the store $_{\text{Cause}}$], and [so $_{\text{Enabling}}$] [yesterday I did $_{\text{Effect}}$].

We introduce several **ENABLE tests**:

1. If the Cause did not occur, is it possible that the Effect may have occurred anyway? If so, it is likely an ENABLE-type relation.
2. If the Cause and Effect have agents, do the agents of the Cause and Effect act in agreement? (i.e., both *I*'s in If I go to the store, I will buy milk have the same will to buy milk if they go to the store.) If so, it is likely an ENABLE-type relation.
3. Is the instance easily restated as "Cause enabled Effect" without any change in semantics? If so, it is likely an ENABLE-type relation.

PREVENT

To reiterate, A PREVENTS B if

1. The patient has a tendency for B
2. The affector and patient do not act in concordance
3. The result B has a reduced likelihood of occurring due to A

Consider the following example:

> I was studying until Jacky began screaming.

[I was studying $_{\text{Effect}}$] [until $_{\text{Preventing}}$] [Jacky began screaming $_{\text{Cause}}$].

We introduce the **PREVENT test**: A PREVENTS B if A reduces the likelihood of B occurring. Consider the following exemplary scenario:

> Her sense of modesty deters her from speaking up.

[Her sense of modesty $_{\text{Cause}}$] [deters $_{\text{Preventing}}$] [her $_{\text{Effect}}$] [from $_{\text{Preventing}}$] [speaking up $_{\text{Effect}}$].

The following example is tricky:

> I unsuccessfully attempted to dissuade my father from purchasing another car.

Annotators may be tempted to annotate it, since there is a similar entry in the Constructicon for "discourages". However, negated connectives are tricky. As per 4.1.1., sometimes connectives are negated to indicate that a causal relationship does not hold, which is the case in the former example. So, we would not annotate this. However, the following example provides a negated connective in which there is a causal relationship.

> The feeder discourages squirrels from stealing seeds.

[The feeder $_{\text{Cause}}$] [discourages $_{\text{Preventing}}$] [squirrels $_{\text{Effect}}$] [from $_{\text{Preventing}}$] [stealing seeds $_{\text{Effect}}$].

Note that it is not the case that negation in the Effect guarantees that the relation is PREVENT-type. Causal classification should thus be dependent on both the Causal Connective and the annotator's reasoning using the Decision Tree provided in Figure A.1.

A.3.2 Decision tree for causation classification

Consider the below tree for a quick summary of the above discussion:

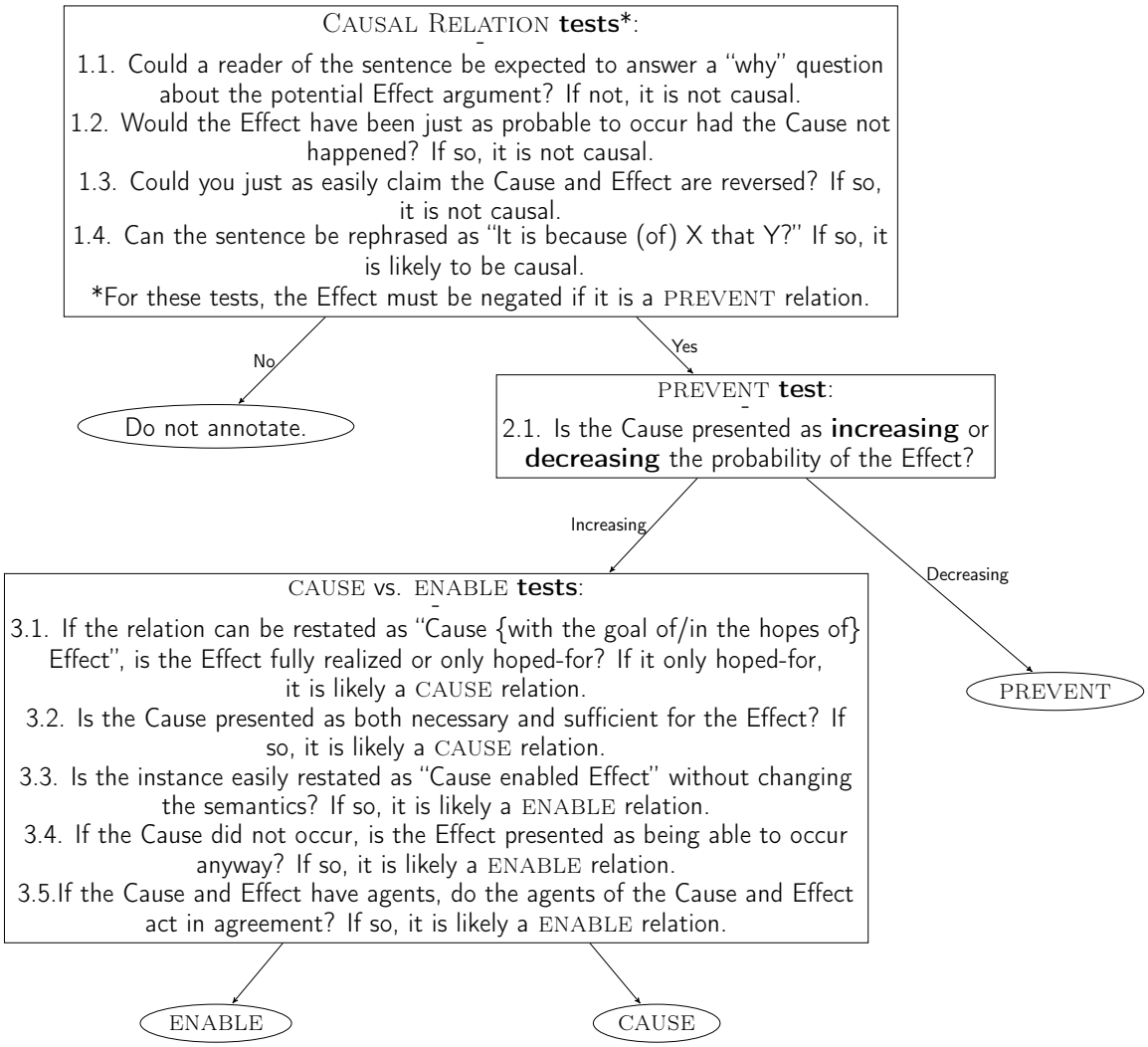

Figure A.1: Decision tree for causation categorization (the CRDT).

Note that tests 3.1.-3.5. are arranged hierarchically. For example, if a causal instances passes both 3.1. and 3.3., we prioritize the positive result from 3.1. (thus, CAUSE) because 3.1. holds more weight than 3.3.

Further note that for testing whether a PREVENT connective is causal using questions 1.1.-1.4. the Effect must be negated before being entered into the subsequent span in order for the test to work. (I.e. in *Her sense of modesty forbids her from*

speaking up, test 1.1. would be: Why did she not speak up? Because of her sense of modesty.)

We also provide fill-in versions of the tests in the tree in order to aid annotators' reasoning.

1. Causal relation tests

   (1.) Could you ask Q: *Why did [Effect] occur?* with the expected response A: *Because of [Cause]*?

   (2.) [Effect] {is/is not} just as likely to occur if [Cause] did not.

   (3.) Does *[Effect] [Connective] [Cause]* make sense?

   (4.) Can you rephrase it as *It is because (of) [Cause] that [Effect]*?

2. PREVENT test

   (1.) [Cause] {increases/decreases} the probability of [Effect].

3. CAUSE vs. ENABLE tests

   (1.) If you can rephrase it as *[Cause] {in the hopes of/with the goal of} [Effect]*, is the Effect {fully realized/hoped-for}?

   (2.) Is it true that *[Cause] is necessary and sufficient for [Effect]*?

   (3.) Can you rephrase it as *[Cause]* enabled [Effect]?

   (4.) Is it true that *It is possible for [Effect] to occur without [Cause]*?

   (5.) If Cause and Effect have agents, is it true that *The agent of [Cause] and the agent of [Effect] have similar goals in mind*?

Here we provide an example of using this tree and its fill-in frames. Consider the following statement and its annotation:

> Their belligerence provoked a war.

[Their belligerence $_{\text{Cause}}$] [provoked $_{?}$] [a war $_{\text{Effect}}$].

Strictly following the decision tree, we start with tests 1.1. to 1.4. Note that if it fails a binary test, the result is **inconclusive**.

1. Causal relation tests

   (1.) Why was there a war? Because of their belligerence. ⟶ Yes, it is causal.

   (2.) The Effect is not as likely to occur without the occurrence of the Cause. ⟶ Yes, it is causal.

   (3.) *A war provoked their belligerence* does not mean the same thing. ⟶ Yes, it is causal.

   (4.) The sentence can be rephrased as *It is because of their belligerence that a war occurred.* ⟶ Yes, it is causal.

2. PREVENT test

   (1.) *Their belligerence* increases the probability of *a war.* ⟶ This is a CAUSE or ENABLE relation.

3. CAUSE vs. ENABLE tests

   (1.) This cannot be rephrased as *They were belligerent {in the hopes of/with the goal of} a war.* ⟶ likely CAUSE

   (2.) *Their belligerence* is presented as both necessary and sufficient for *a war.* ⟶ Likely CAUSE

   (3.) *Their belligerence enabled a war* does not have the same meaning. ⟶ Inconclusive

(4.) If *their belligerence* did not occur, the statement does not present *a war* as able to occur anyway. $\longrightarrow$ Inconclusive

(5.) This test does not hold because it is not clear that the affector and patient are the same.

It is clear through the progression of the tests that CAUSE is the most likely relation. So, we finalize the annotation as:

[Their belligerence $_{\text{Cause}}$] [provoked $_{\text{Causing}}$] [a war $_{\text{Effect}}$].

Note that in practice, the process of annotating is much simpler. Once a Causal Relation has been established using the Constructicon and tests 1.1.-1.4., annotators are then able to consult the Constructicon for whether the relation is PREVENT or CAUSE/ENABLE. So, annotators thus only have to decide between CAUSE and ENABLE once a causal relation is established.

A.4 Edge cases

First of all, note that for both connectives and arguments, punctuation is not included in the span of the annotation unless it cannot be avoided (i.e., it is within an argument as in example 6.2.).

A.4.1 Special cases of connectives

1. Negations

Connectives may be negated to say that the indicated causal relationship does not hold – for example, *This will not lead to the same disastrous consequences*. However, if even with the negated connective, a causal relation still holds, then these negations should be ignored for the purposes of annotation; the negation is simply another modifier.

2. Conjunctions

If there are two different connectives related by a conjunction, two different annotations should be created – one for each connective. See example 6.4.

- This does not apply to arguments – if there is a single connective but one or both of the arguments consist of two or more phrases or clauses connected by a conjunction, the entire conjoined argument phrase should be annotated as the argument of a single causation instance. The same holds for disjunctions (*or*).

3. Complementizers

Arguments may be introduced by complementizers such as *that*, sometimes immediately after the connective – for example, *We must ensure that this does not happen.* When they occur after verbs, these complementizers should be annotated as part of the argument, not the connective.

- Complementizers not following verbs may be part of the connective if they are always present (e.g., *on the grounds that*). When the complementizer is optional after a non-verbal connective (e.g., *so that*), the *that* should be omitted from both spans.

4. Ambiguity in the number of connectives

If a connective could in principle either be split into multiple connectives or combined into a single one, it should be annotated as multiple connectives. For example, in *This is necessary to prevent war*, *necessary to prevent* could be considered a single larger connective of PREVENT-type. Nonetheless, it should be split into the connectives *necessary to* (ENABLE-type) and *prevent* (PREVENT-type). Also consider example 6.4.

5. Pragmatic discourse markers

Occasionally, causal words will be used to express a **pragmatic cause** – e.g., not *X is true because...*, but *X, and I'm saying this because...* These should not be annotated.

6. Inference markers

Occasionally, causal words will be used to indicate evidence – for example, *The car was driven recently, because the hood was still hot.* Like pragmatic discourse markers, these should not be annotated.

- Causal language used to talk about whether something meets a particular definition (e.g., *This is not a square, because its sides are uneven*) does not fall into this category. Such language describes the reason for something being true, not merely the reason for believing something to be true.

7. Nominal and adjectival connectives

Many connectives are adjectives or nouns. These can appear embedded within many different linguistic constructions. For example, they can appear as arguments of a copula (*The cause of the fire was a cigarette butt*), complements (*Their support seems essential to the organization's continuity*), appositives (*He researches E. coli, the cause of many an infection*), prepositional arguments (*He pointed to his predecessor's mistakes as the cause of the current crisis*), and more. Only the noun or adjective and the function words consistently used to introduce their arguments should be annotated as part of the connective. In the above examples, *cause* and *essential to* should be annotated as the connectives.

8. Nominalized verbs

A nominalization of the verb in the construction should be annotated as a connective (e.g., *prevention of*). When the object of the verb appears with an *of* phrase, as in

prevention of, the *of* should be annotated as part of the connective, in keeping with the practice of annotating verb argument words.

9. *To* indicating a verb argument

Occasionally, *to* is used in a way that can be rephrased as *in order to* – e.g., *I used caulk to fix the leak.* This distinction is difficult to make reliably, and has few implications for downstream processing. Therefore, such cases should be annotated in the same fashion as other *to*'s used to indicate Purpose. (These cases should not be annotated if they cannot be rephrased as *in order to*.)

10. The word *for*

One particularly difficult case is the preposition *for*. It is included in the Constructicon, but below is a more complete list that includes possible meanings that are not considered causal:

Sense	**Examples**	**Causal?**
Exchange of goods	*Buy **for** $5, swap X **for** Y*	No
Topic	*My ideas **for** a better world*	No
Purpose of existence (i.e. existence of the putative EFFECT)	*a vase **for** the flowers, a forward-facing camera **for** video chat*	No
Precipitating action	*He thanked the crowd **for** listening, I attacked him **for** slandering me, I'm reporting them to the BBB **for** horrible customer relations*	Yes
Purpose of benefit	*I'm running **for** prostate research.*	Yes
Precipitating situation or need	*I go to the mall **for** the crowds, I went to the store **for** a bag of carrots*	Yes

Table A.2: The different causal and non-causal senses of *for*.

A.4.2 Special cases of arguments

1. Coreferent nouns/pronouns

If a Cause or Effect argument consists entirely of a pronoun that is coreferent with another noun, the pronoun should still be annotated as the Cause or Effect.

2. Missing arguments

The Cause or Effect may be missing entirely, particularly in passive sentences. For example, no Cause is given in the sentence *The hedging of business risks could well be discouraged.* In these cases, the missing argument should simply be omitted from the annotation. (Note that this is relatively rare.)

3. Conjunctions with shared constituents

Coordinate structures may lead to a piece of an argument being shared between two parts of the sentence. Only the relevant subspan of the coordination should be annotated as part of the argument. For example, the bolded portion of the following would be annotated as the Cause: ***The product*** *was widely praised but* ***came with no assembly instructions****, leading to many complaints.*

4. Attachment ambiguities

When there is an attachment ambiguity that cannot be resolved even semantically, low attachment should be preferred. For example, consider the sentence *He watched her sing with enthusiasm.* Here *with enthusiasm* could be taken to modify *watched* or *sing*, as demonstrated in the two possible syntax trees below.

Preferring low attachment means *with enthusiasm* should be read as modifying *sing*; thus, we would take the latter tree to be the accurate structural representation of the sentence.

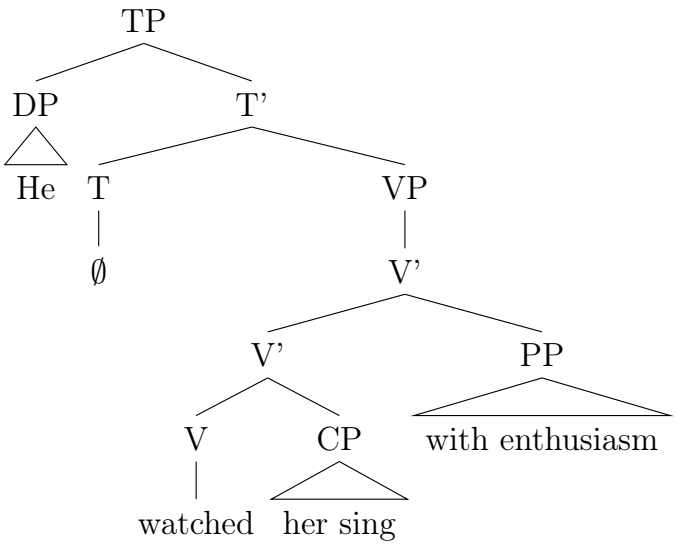

Figure A.2: Tree where *with enthusiasm* modifies *watched*.

5. Coreference ambiguities

Similarly, if there are two possible chunks that could be annotated as the antecedent of a coreferent pronoun, and the ambiguity cannot be resolved even semantically, the smaller chunk should be annotated as the antecedent.

6. Purpose-type Cause spans

In instances of the Cause argument where it depicts a Purpose, the controlling subject of the effect Clause will often be a subspan of the action whose purpose is being stated. For example, in *Maggie went to the store to buy eggs*, *Maggie* is the controlling subject of *to buy eggs*, but it is Maggie's trip to the store whose purpose is to buy eggs. The entire action – in this example, *Maggie went to the store* – should be annotated as the Cause.

7. Arguments of non-finite verbal connectives

Verbal connectives can appear without an explicitly named subject (e.g., *It is important to prevent abuse*). In some cases, however, the coreferenced subject is known (e.g., *It*

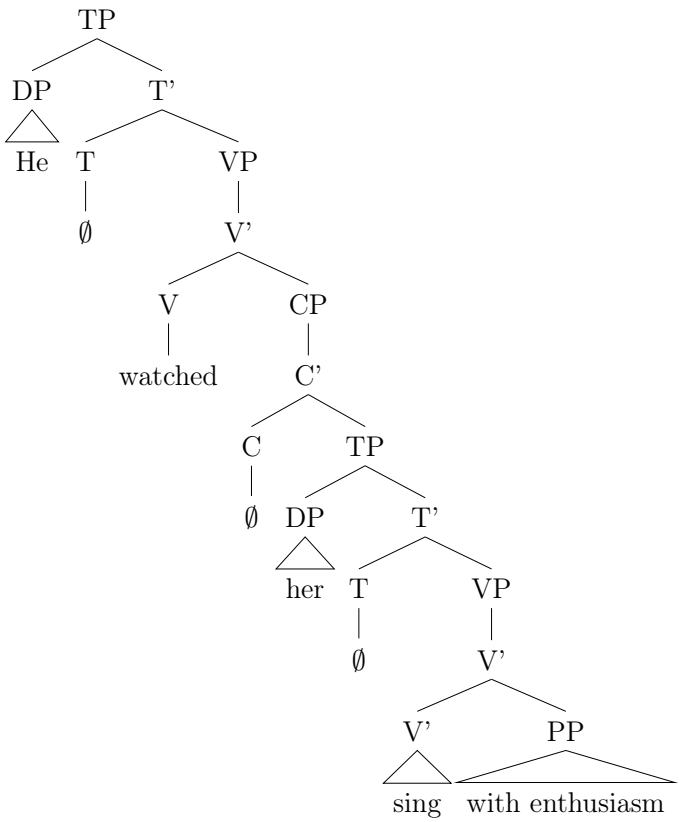

Figure A.3: Tree where *with enthusiasm* modifies *sing*.

is important for us to prevent abuse). If the subject can definitively be established, and no action is specified, the implied subject should be annotated as the Cause (*us*, in the above example), or the subject's action if it is explicitly given. If the subject cannot be definitively established, no Cause should be annotated.

- The test for a known subject is whether there is exactly one reflexive pronoun which can be added to the end of the clause. In the above example, *ourselves* can be added to the end of the *prevent* clause, but no other reflexive pronoun can, because *us* is implicitly the subject of *prevent*.
  - This case also covers *how to* or *why to* constructions (e.g., *We must determine how to prevent abuse*).

8. Hedges

Hedges should not be annotated because they generally appear in inferential environments (e.g., *It probably won't work, I don't think it'll work*).

9. Speech acts

The Effect argument may be a speech act – for example, *Jeremy can't make it, so can you bring some wine?* The instance should be classified considering the Effect to be the speaker performing the speech act, i.e. *I am asking you to bring some wine.*

10. Participials

There may be a participial phrase that seems to be an extension of the Cause or Effect but is separated from it in the sentence. For example, in *He walked warily because of the mud, skirting the lake at a distance*, the final participial is understood to apply to *he walked warily.* Such participials should not be annotated as part of either argument.

- If the connective itself is a participle, as in *The fire swept through the county,* ***causing*** *extensive damage*, the entire clause modified by the participle (here, *the fire swept through the county*) should be annotated as the argument, rather than just the subject of that clause (*the fire*).

11. Parentheticals, prepositional phrases, and other modifiers and interruptions

It is not always clear whether phrases like prepositional phrases and parentheticals should be included in an argument span. In general, the rule is that any language modifying or describing the argument (but not modifying the other argument or the entire relationship) should be included. This is constrained by the natural boundaries of the sentence (unless the next sentence begins with an *and*, in which case the next

sentence should be included as well). Subspans that **should** be included in arguments include (with the relevant argument span in brackets in the examples):

1. **Prepositions that modify only the argument span**: For example, [*Activists' efforts since 1985*] *have led to few changes.*

2. **Relative clauses**: For example, [*The First Lady, who traditionally chooses the decorations,*] *caused quite a stir with her selection.*

3. **Appositives**: For example, [*The First Lady, traditionally the decorator of record,*] *caused a stir with her selection.* Subspans that should NOT be included in arguments include:

   (a) **Parenthetical matrix clauses**: For example, [*The park,*] *he said,* [*was flooded*] *because of the rain*; *Due to the rain,* [*the park,*] *as she expected,* [*was flooded*]; *Due to the rain,* [*the park was flooded,*] *as she expected.* For consistency, the same standard applies even in cases where the parenthetical is arguably the matrix clause only for the argument which it interrupts or is juxtaposed to.

   (b) **Prepositions that modify the entire relationship**: For example, *Since 1985,* [*Activists' efforts*] *have led to few changes.*

12. Effects with modality of obligation

When an Effect has a *should* in it – e.g., *I didn't get cake, so you should give me the next piece* – we interpret the statement as indicating that, in the speaker's opinion, an obligation exists in the world as a result of the Cause.

13. Connective embeded within an argument

Sometimes the connective may be embedded within an apparent argument, as in *The regulatory restraints that many experts regard as a necessary condition of technological*

processes are largely unnecessary, where *a necessary condition of* is embedded within the subject of the sentence. In these cases, the argument span should be minimized to the longest possible propositional phrase or clause that excludes the connective. So here, *The regulatory restraints* would be annotated as Cause, *a necessary condition of* would be annotated as an ENABLING connective, and *technological processes* as the Effect, while everything else is excluded.

A.4.3 Specifications for Reddit posts

These are exceptions that apply to all documents; we merely observe them most frequently in Reddit posts.

1. Abbreviations

Many users on Reddit will use texting shorthand, such as "lol" or "fyi". These are not annotated because it is usually ambiguous whether it modifies a single argument or the entire causal statement. Consider the following examples:

- [They almost never do $_{\text{Effect}}$] btw, [because $_{\text{Causing}}$] [the school wants them to do research or whatever $_{\text{Cause}}$].
- [The dog $_{\text{Cause}}$] lol [at my parents' house $_{\text{Cause}}$] [caused $_{\text{Causing}}$] [me $_{\text{Effect}}$] [to $_{\text{Causing}}$] [lose track of time $_{\text{Effect}}$].

2. Long chains of events

Annotators may find that Reddit users may use long chains of events, i.e. *Grad studies makes it so we meet lots of people from lots of different places and then they move away and that sucks.* As specified in section 4.1.2., the entire causal event chain on the right side of *so* should be annotated as the Effect. Please review section 4.1.2. for more guidance on this point.

3. Unnecessary modifiers

Annotators will encounter cases where there seems to be an excessive use of modifiers/adjectival phrases. In accordance with section 4.2.12., these modifiers should be included when they modify only a single argument (but we never annotate modifiers of connectives, as per section 2.). We do this because in most cases, the natural boundaries of the sentence will help standardize the annotations.

- [I've ran out of a room crying once $_{\text{Effect}}$] [when $_{\text{Causing}}$] [everyone looked at me to answer a question in class $_{\text{Cause}}$].

4. Punctuation pairs

Occasionally, we come across "pairs" of punctuation markers, i.e. " " or ". While we do not include punctuation in our annotations, it is fine to leave one of the pair within a span if it is unavoidable. Consider the following example:

- [You did not get a chance to bond with your classmates $_{\text{Effect}}$] [because $_{\text{Causing}}$] [they shut down the chat to keep everyone "focused $_{\text{Cause}}$]". (Note that in this example, there is another embedded causal instance with the connective *to*.)

5. Connective shorthands

In casual text settings in some Reddit posts, annotators may notice that connectives such as "because" are abbreviated to "cause", "cus", "cuz", etc. These are to be annotated when it is abundantly clear exactly which connective it refers to, such as *I went to the store cus I needed some eggs*. An example of this occurrence in Fables would be "till", as in *I was a great dancer till I broke my foot*.

A.5 Suggestions for the annotation process

Before annotation rounds begin:

1. Familiarize yourself with (1) the connectives in the Constructicon and (2) the guidelines.

2. Send Angela any questions that you have about either.

While annotating:

1. During your first pass through the reading, identify and annotate spans that resemble connectives from the Constructicon.

2. Then, using the Constructicon, identify the Cause, Effect, and Means argument spans for each where they appear.

3. After, use tests 1.1.-1.4. to ensure that the annotation instances are indeed Causal. Remove those that fail the test(s).

4. Next, using the Constructicon, identify the PREVENT-type connectives.

5. With the remaining non-PREVENT-type connectives, use tests 3.1.-3.5. of the Decision tree to differentiate between CAUSE and ENABLE connectives.

6. After steps 1-4 are finished, CTRL-F for the following instances of easy-to-miss and frequently appearing connectives:

   (1.) after (*after*, *in the aftermath of*, *the aftermath of ... is*, *comes after*)

   (2.) as (*as*, *as long as*, *so as to*)

   (3.) at

   (4.) before

   (5.) for (*for*, *is responsible for*)

   (6.) once

   (7.) since

(8.) to (*to*, *obliges ... to*, *is critical to*, *is essential to*, etc.)

(9.) until

(10.) when (*when*, *whenever*)

(11.) where

(12.) with (*with*, *with* DET *goal of*, *with* DET *objective of*)

7. If missed connectives are found, go through steps 2 through 5 for the newfound connectives.

More generally, annotations should be strictly *by the book*; only connectives that appear in the Constructicon should be annotated, and only connective spans identified in the Constructicon are to be annotated.

If annotators believe that they have come across a causal construction that does not currently exist in the Constructicon, do not annotate it but submit it for consideration into the Constructicon here.

A.6 Example annotations

1. > Some borrowers opted for nontraditional mortgages because that was their only way to get a foothold in the California housing market.

   [Some borrowers opted for nontraditional mortgages $_{\text{Effect}}$] [because $_{\text{Causing}}$] [that was their only way to get a foothold in the California housing market $_{\text{Cause}}$].

2. > If they are regulated entities, yes, we can see their code and they need to freeze their code if asked.

[If $_{\text{Causing}}$] [they are regulated entities $_{\text{Cause}}$], [yes, we can see their code and they need to freeze their code if asked $_{\text{Effect}}$].

3. > My brother caused a fire by dropping a lit match.

[My brother $_{\text{Cause}}$] [caused $_{\text{Causing}}$] [a fire $_{\text{Effect}}$] by [dropping a lit match $_{\text{Means}}$].

4. > With that one signature, the President sparked hundreds of protests.

[With $_{\text{Causing}}$] [that one signature $_{\text{Means}}$], [the President $_{\text{Cause}}$] sparked [hundreds of protests $_{\text{Effect}}$].

With [that one signature $_{\text{Means}}$], [the President $_{\text{Cause}}$] [sparked $_{\text{Causing}}$] [hundreds of protests $_{\text{Effect}}$].

Since we are using the Inception annotation tool, consider that since *with* and *sparked* are two disparate connectives, annotators should not include an arrow from *with* to *sparked* as they would for *For* and *to* (which comprise a single connective) in *For Sally to pass the class, she must get higher than an 80 on the final.*

5. > They come up with a common standard so that they are all busting trades at the same level.

[They come up with a common standard $_{\text{Cause}}$] [so $_{\text{Causing}}$] that [they are all busting trades at the same level $_{\text{Effect}}$].

6. > Low interest rates and widely available capital were prerequisites for the creation of a credit bubble.

[Low interest rates and widely available capital $_{\text{Cause}}$] were [prerequisites for $_{\text{Enabling}}$] [the creation of a credit bubble $_{\text{Effect}}$].

7. > For the United States to continue to lead the world's capital markets, we must continue to encourage innovation.

[For $_{\text{Enabling}}$] [the United States $_{\text{Effect}}$] [to $_{\text{Enabling}}$] [continue to lead the world's capital markets $_{\text{Effect}}$], [we $_{\text{Cause}}$] must [continue to encourage innovation $_{\text{Cause}}$].

8. > A judgement in favor of the United States shall stop the defendant from denying the allegations of the offense in any subsequent civil proceeding brought by the United States.

[A judgement in favor of the United States $_{\text{Cause}}$] shall [stop $_{\text{Preventing}}$] [the defendant $_{\text{Effect}}$] [from $_{\text{Preventing}}$] [denying the allegations of the offense in any subsequent civil proceeding brought by the United States $_{\text{Effect}}$].

9. > What are your recommendations for creating a system that would prevent or discourage banks from becoming "too big to fail"?

What are your recommendations for creating [a system that $_{\text{Cause}}$] would prevent or [discourage $_{\text{Preventing}}$] [banks $_{\text{Effect}}$] [from $_{\text{Preventing}}$] [becoming "too big to fail $_{\text{Effect}}$]"?

What are your recommendations for creating [a system that $_{\text{Cause}}$] would [prevent $_{\text{Preventing}}$] or discourage [banks $_{\text{Effect}}$] [from $_{\text{Preventing}}$]

[becoming "too big to fail $_{\text{Effect}}$]"?

10. > Without better regulation, the economy will not recover and we can expect further crisis.

[Without $_{\text{Preventing}}$] [better regulation $_{\text{Cause}}$], [the economy will not recover and we can expect further crisis $_{\text{Effect}}$].

11. > Wine without food makes my head hurt, and with it makes my stomach hurt.

[Wine without food $_{\text{Cause}}$] [makes $_{\text{Causing}}$] [my head hurt $_{\text{Effect}}$], and with it makes my stomach hurt.

[Wine $_{\text{Cause}}$] without food makes my head hurt, and [with it $_{\text{Cause}}$] [makes $_{\text{Causing}}$] [my stomach hurt $_{\text{Effect}}$].

Appendix B

Sample of the Constructicon

Table B.1: A sample of the Constructicon, which is available to the annotators as a searchable Google sheet.

Connective pattern (verbs given in present tense and nouns/adjectives given as copulas for readability)	**Variants (not including passives, infinitives, or nominalizations of verbs)**	**Words to annotate as connective ([ ] = may not be present)**	**Type**	**Comments**	**Example(s)**
<Cause>where <Effect>	Where <Cause>, <Effect>	where	CAUSE/ENABLE	Should not be annotated when it refers to physical location. It should only be annotated when it is a metaphorical "where" referring to circumstance – i.e., where it could just as easily have been "when." Don't annotate when it can be replaced with "in which" (e.g., "a meeting where 10 people showed up").	Where regulators cozy up with CEOs, corruption abounds.
<Effect>amid <Cause>	<Effect>amidst <Cause> amid <Cause>, <Effect>	amid	CAUSE/ENABLE	In the more abstract sense of "in the context of," not "surrounded by" in a physical sense.	Banks and stores closed yesterday amid growing fears of violence.
With <Cause>, <Effect>	<Effect>, with <Cause>	with	CAUSE/ENABLE	Should not be annotated when used in a possessive or accompaniment sense (e.g., "With thousands of unique species, the island is an evolutionary biologist's dream.").	With supplies running low, we didn't even make a fire that night.
Without <Cause>, <Effect>		without	PREVENT	With the meaning "in the absence of."	Without better regulation, the same problem will recur.
Having <Cause>, <Effect>		having	CAUSE/ENABLE		Having looked at the older document, the faculty believed that a middle ground might be appropriate.

Appendix C

Sample of training quizzes: CR

Training Quiz 2

For each statement, choose the correct annotation based on the Causal Relation Annotation Guidelines. Mark only one choice per question.

Email: _____

1. The Dayton Democrat will spend three days visiting hospitals and other facilities to seek understanding why aid has been ineffective in stemming malnourishment and other medical problems. (10 points)
   - ◯ Do not annotate.
   - ◯ [The Dayton Democrat will spend three days visiting hospitals and other facilities to seek understanding {Effect}] [why {Causing}] [aid has been ineffective in stemming malnourishment and other medical problems {Cause}].
   - ◯ The Dayton Democrat will spend three days visiting hospitals and other facilities to [seek understanding {Effect}] [why {Causing}] [aid has been ineffective in stemming malnourishment and other medical problems {Cause}].
   - ◯ The Dayton Democrat will spend three days visiting hospitals and other facilities [to seek understanding {Effect}] [why {Causing}] [aid has been ineffective {Cause}] in stemming malnourishment and other medical problems.

2. They aren't sure why aid has been ineffective in preventing malnourishment and other medical problems. (10 points)
   - ◯ Do not annotate.
   - ◯ They aren't sure why [aid {Cause}] has been ineffective in [preventing {Preventing}] [malnourishment and other medical problems {Effect}].
   - ◯ They aren't sure why [aid has been ineffective {Cause}] in [preventing {Preventing}] [malnourishment and other medical problems {Effect}].
   - ◯ They aren't sure why [aid {Cause}] has been [ineffective in preventing {Preventing}] [malnourishment and other medical problems {Effect}].

3. My birth defects were linked to my mother's past smoking habits. (10 points)
   - ◯ Do not annotate.
   - ◯ [My birth defects {Effect}] were [linked {Causing}] to [my mother's past smoking habits {Cause}].
   - ◯ [My birth defects {Effect}] [were linked to {Causing}] [my mother's past smoking habits {Cause}].
   - ◯ [My birth defects {Effect}] were [linked {Enabling}] to [my mother's past smoking habits {Cause}].

4. On the grounds that I had stolen from Costco, Emory rescinded my offer. (10 points)
   - ◯ Do not annotate.
   - ◯ [On the grounds that {Causing}] [I had stolen from Walmart {Cause}], [Fulbright rescinded my offer {Effect}].
   - ◯ [On the grounds {Causing}] that [I had stolen from Walmart {Cause}], [Fulbright rescinded my offer {Effect}].
   - ◯ On [the grounds {Causing}] that [I had stolen from Walmart {Cause}], [Fulbright rescinded my offer {Effect}].

5. We must eliminate corporate taxes for the good of the nation. (10 points)
   - ◯ Do not annotate.
   - ◯ [We {Cause}] must [eliminate {Preventing}] [corporate taxes {Effect}] for the good of the nation.
   - ◯ [We {Cause}] must [eliminate {Preventing}] [corporate taxes for the good of the nation. {Effect}]
   - ◯ [We {Cause}] [must eliminate {Preventing}] [corporate taxes {Effect}] for the good of the nation.

6. We must eliminate corporate taxes for the good of the nation. (10 points)
   - ◯ Do not annotate.
   - ◯ [We must eliminate corporate taxes {Cause}] [for {Enabling}] [the good of the nation {Effect}].
   - ◯ [We must eliminate corporate taxes {Cause}] [for {Causing}] [the good of the nation {Effect}].
   - ◯ We must [eliminate corporate taxes {Cause}] [for {Causing}] [the good of the nation {Effect}].

7. My birth defects were not caused by my mother's smoking habits. (10 points)
   - ◯ Do not annotate.
   - ◯ [My birth defects {Effect}] were [not caused {Preventing}] by [my mother's smoking habits {Cause}].
   - ◯ [My birth defects {Effect}] [were not caused {Causing}] by [my mother's smoking habits {Cause}].
   - ◯ [My birth defects {Effect}] were not [caused {Causing}] by [my mother's smoking habits {Cause}].

8. As Americans, we must ensure that America remains in power. (10 points)
   - ◯ Do not annotate.
   - ◯ As Americans, [we {Cause}] must [ensure {Enabling}] that [America remains in power {Effect}].
   - ◯ As Americans, [we {Cause}] must [ensure {Causing}] that [America remains in power {Effect}].
   - ◯ [As Americans, we {Cause}] must [ensure {Enabling}] that [America remains in power {Effect}].

9. I got so hungry that I went to the store. (10 points)
   - ◯ Do not annotate.
   - ◯ [I got so hungry {Cause}] [that {Enabling}] [I went to the store {Effect}].
   - ◯ [I got so hungry {Cause}] [that {Causing}] [I went to the store {Effect}].
   - ◯ [I got so hungry that I went {Cause}] [to {Enabling}] [the store {Effect}].

10. I walked briskly to school as I was tardy, taking the shorter route. (10 points)
    - ◯ [I walked briskly to school {Effect}] [as {Causing}] [I was tardy {Cause}], taking the shorter route.
    - ◯ Do not annotate.
    - ◯ [I walked briskly to school {Effect}] [as {Causing}] [I was tardy, taking the shorter route {Cause}].
    - ◯ [I walked briskly to school {Effect}] [as {Enabling}] [I was tardy, taking the shorter route {Cause}].

Bibliography

Alicke, M. D., Rose, D., and Bloom, D. (2011). Causation, norm violation, and culpable control. *The Journal of Philosophy*, 108(12):670–696.

Artstein, R. and Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Comput. Linguist.*, 34(4):555–596.

Baglini, R. and Siegal, E. A. B.-A. (2020). Direct causation: A new approach to an old question. *University of Pennsylvania Working Papers in Linguistics*, 26:19–28.

Bar-Asher Siegal, E. and Boneh, N. (2019). Sufficient and necessary conditions for a non-unified analysis of causation. *Proceedings of the 36th West Coast Conference on Formal Linguistics*, pages 55–60.

Bayerl, P. S. and Paul, K. I. (2011). What determines inter-coder agreement in manual annotations? a meta-analytic investigation. *Computational Linguistics*, 37(4):699–725.

Beller, A., Bennett, E., and Gerstenberg, T. (2020). The language of causation. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*.

Bethard, S., Kolomiyets, O., and Moens, M.-F. (2012). Annotating story timelines as temporal dependency structures. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2721–2726, Istanbul, Turkey. European Language Resources Association (ELRA).

Bittner, M. (1999). Concealed causatives. *Natural Language Semantics*, 7(1):1–78.

Bonial, C., Babko-Malaya, O., Choi, J., Hwang, J., and Palmer, M. (2010). Propbank annotation guidelines.

Bonial, C., Bonn, J., Conger, K., Hwang, J. D., and Palmer, M. (2014). PropBank: Semantics of new predicate types. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3013–3019, Reykjavik, Iceland. European Language Resources Association (ELRA).

Caselli, T. and Vossen, P. (2017). The event StoryLine corpus: A new benchmark for causal and temporal relation extraction. In *Proceedings of the Events and Stories in the News Workshop*, pages 77–86, Vancouver, Canada. Association for Computational Linguistics.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104(2):367–405.

Cheng, P. W. and Novick, L. R. (1991). Causes versus enabling conditions. *Cognition*, 40(1-2):83–120.

Cheng, P. W. and Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, 99(22):365–382.

Chomsky, N. (2015). *What kind of creatures are we?* Columbia University Press.

Davidson, D. (1967). Causal relations. *Journal of Philosophy*, 64(21):691–703.

Dunietz, J. (2018). *Annotating and Automatically Tagging Constructions of Causal Language*. PhD thesis, Carnegie Mellon University.

Dunietz, J., Levin, L., and Carbonell, J. (2017a). Automatically Tagging Constructions of Causation and Their Slot-Fillers. *Transactions of the Association for Computational Linguistics*, 5:117–133.

Dunietz, J., Levin, L., and Carbonell, J. (2017b). The BECauSE corpus 2.0: Annotating causality and overlapping relations. In *Proceedings of the 11th Linguistic Annotation Workshop*, pages 95–104, Valencia, Spain. Association for Computational Linguistics.

Dunietz, J., Levin, L., and Carbonell, J. G. (2015). Annotating causal language using corpus lexicography of constructions. In *Proceedings of The 9th Linguistic Annotation Workshop*, pages 188–196.

Einhorn, H. and Hogarth, R. (1986). Judging probable cause. *Psychological Bulletin*, 99:3–19.

Gerstenberg, T., Goodman, N. D., Lagnado, D., and Tenenbaum, J. (2015). How, whether, why: Causal judgements as counterfactual contrasts. *Cognitive Science*, pages 1–6.

Gerstenberg, T., Goodman, N. D., Lagnado, D., and Tenenbaum, J. (2020). A counterfactual simulation model of causal judgments for physical events.

Goldberg, A. E. (2013). *Constructionist approaches*. Hoffmann and Trousdale.

Halpern, J. Y. (2016). *Actual Causality*. The MIT Press.

Halpern, J. Y. and Pearl, J. (2013). Causes and explanations: A structural-model approach — part 1: Causes.

Icard, T., Kominsky, J. F. K., and Knobe, J. (2017). Normality and actual causal strength. *Cognition*, 161:80–93.

Ide, N., Baker, C., Fellbaum, C., and Passonneau, R. (2010). The manually annotated sub-corpus: A community resource for and by the people. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 68–73, Uppsala, Sweden. Association for Computational Linguistics.

Kingsbury, P. R. and Palmer, M. (2003). Propbank: the next level of treebank.

Lakoff, G. and Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*.

Lassiter, D. (2018). Causation and probability in indicative and counterfactual conditionals. *Unpublished manuscript*, pages 1–27.

Lauer, S. and Nadathur, P. (2018). Sufficiency causatives. Unpublished manuscript.

Lauer, S. and Nadathur, P. (2020). Causal necessity, causal sufficiency, and the implications of causative verbs. *Glossa: a journal of general linguistics*, 5:49–105.

Levin, B. (2019). Resultatives and causation. *Unpublished manuscript*, pages 1–33.

Lewis, D. (1973). Causation. *Journal of Philosophy*, 70(17):556–567.

Lewis, D. (1975). Counterfactuals. *Foundations of Language*, 13(1):145–151.

Marcus, M., Kim, G., Marcinkiewicz, M. A., MacIntyre, R., Bies, A., Ferguson, M., Katz, K., and Schasberger, B. (1994). The Penn Treebank: Annotating predicate argument structure. In *Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey, March 8-11, 1994*.

Mirza, P., Sprugnoli, R., Tonelli, S., and Speranza, M. (2014). Annotating causality in the TempEval-3 corpus. In *Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL)*, pages 10–19, Gothenburg, Sweden. Association for Computational Linguistics.

Mirza, P. and Tonelli, S. (2016). CATENA: CAusal and TEmporal relation extraction from NAtural language texts. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 64–75, Osaka, Japan. The COLING 2016 Organizing Committee.

Mostafazadeh, N., Chambers, N., He, X., Parikh, D., Batra, D., Vanderwende, L., Kohli, P., and Allen, J. (2016a). A corpus and cloze evaluation for deeper understanding of commonsense stories. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 839–849, San Diego, California. Association for Computational Linguistics.

Mostafazadeh, N., Grealish, A., Chambers, N., Allen, J., and Vanderwende, L. (2016b). CaTeRS: Causal and temporal relation scheme for semantic annotation of event structures. In *Proceedings of the Fourth Workshop on Events*, pages 51–61, San Diego, California. Association for Computational Linguistics.

Ni, Y. (2012). Categories of causative verbs: a corpus study of mandarin chinese.

Pearl, J. (2009). *Causality*. Cambridge University Press, Cambridge, UK, 2 edition.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., and Webber, B. (2008). The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A. K., Robaldo, L., and Webber, B. L. (2006). The penn discourse treebank 2.0 annotation manual.

Sandhaus, E. (2008). The New York Times Annotated Corpus.

Sloman, S., Barbey, A., and Hotaling, J. (2009). A causal model theory of the meaning of cause, enable, and prevent. *Cognitive Science*, 33(1):21–50.

Smith, N. A., Cardie, C., Washington, A., and Wilkerson, J. (2014). Overview of the 2014 NLP unshared task in PoliInformatics. In *Proceedings of the ACL 2014*

Workshop on Language Technologies and Computational Social Science, pages 5–7, Baltimore, MD, USA. Association for Computational Linguistics.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, 12(1):49–100.

Tomasello, M. (2001). First steps toward a usage-based theory of language acquisition. 11(1-2):61–82.

Wittgenstein, L. (1953). *Philosophical Investigations*. Basil Blackwell, Oxford.

Wolff, P. (2007). Representing causation. *Journal of experimental psychology. General*, 136:82–111.

Wolff, P., Klettke, B., Ventura, T., and Song, G. (2005). Expressing causation in english and other languages.

Wolff, P. and Song, G. (2003). Models of causation and causal verbs. *Cognitive Psychology*, 47:276–332.

Wolff, P. and Zettergren, M. (2002). A vector model of causal meaning. In *Proceedings of the twenty-fifth annual conference of the cognitive science society*. Erlbaum.