5

Event Causality

Paramita Mirza

Abstract. A crucial aspect of understanding and reconstructing narratives is identifying the underlying causal chains, which explain why certain things happened and make a coherent story. In order to build such causal chains, or *causelines* (see Chapter 6), we need to identify causal links between events in the story, which may be expressed explicitly as well as understood implicitly using commonsense knowledge. This chapter reviews research efforts on the automated extraction of such event causality from natural language text. It starts with a brief review of existing causal models in psychology and psycholinguistics as a building block for understanding causation. These models are useful tools for guiding the annotation process to build corpora annotated with causal pairs. I then outline existing annotated resources, which are used to build and evaluate automated causality extraction systems. I focus on summarizing research efforts on automatic extraction systems for identifying causality between events triggered by lexical items: for instance, extracting a causal link between *hugged* and *felt* in 'He hugged her tight because he felt grateful'. Most extraction systems rely on the presence of explicit causal connectives such as *because*. However, causality between events may be expressed implicitly through adjacency in discourse. Furthermore, circumstantial events surrounding the causal complex are rarely expressed with language as they are part of human common sense. Therefore, discovering causal common sense is also important to fill the gaps in the causal chains, and I discuss existing work in this line of research.

5.1 Introduction

An important part of understanding text, narratives in particular, arises from understanding whether and how two events are related semantically. For instance, when given a sentence 'He hugged her tight because he felt grateful

https://doi.org/10.1017/9781108854221.007 Published online by Cambridge University Press

for her help', humans understand that there are three events, indicated by the words 'hugged', 'felt grateful' and '(her) help', and that the *helping* event results in him *feeling grateful*, which in turn leads to the *hugging* event, establishing a causal chain among them. Besides being important components of discourse understanding, automatic extraction of event causality and causal reasoning are important for various downstream tasks such as question answering, decision support and future event prediction given a chain of past events, among others.

As a starting point, we should first consider a definition of *causation* in the general sense. It is commonly agreed in philosophical and psychological literature that (1) causation is a relation between two events – *cause* and *effect*; (2) causation has a temporal dimension – the cause must precede the effect; and (3) causation is counterfactual: if the cause had not occurred, the effect would not have occurred either.

It is important to note that causation exists as a psychological principle for understanding the world independent of language, even though language can be used to talk about causation (Neeleman and van de Koot, 2012). Consider our previous example, 'He hugged her tight because he felt grateful'. As humans we understand that the *hugging* event is indeed triggered by him *feeling grateful*. However, it would not happen if certain conditions were not in place or certain events did not happen, for instance, both of them feel comfortable with hugging each other, they stand in close proximity, his arms were free and so on. Such a *mental model* (Neeleman and van de Koot, 2012) or *causal complex* (Hobbs, 2005) contains fine-grained information of eventualities (events or states) whose happening or holding entails that the effect will happen, and therefore, it cannot be trivially conveyed with a single linguistic expression.

When we intend to use knowledge about causal relations between events to reason formally about causation, we need to consider which eventualities are in or out of the causal complex for a particular effect; meanwhile, for language understanding and determining lexical semantics between concepts, it is often sufficient to distinguish eventualities that deserve to be called *causes* for a resulting event or state (Hobbs, 2005).

In this chapter I will first focus on the second inquiry, i.e., on understanding whether there exists a *causal relation* encoded in certain linguistic expressions, in terms of *causes* and resulting *effects*. I will start with a brief overview of existing attempts in the psychology field to model causation, which are useful for guiding causal annotation endeavours. I will also summarize available resources annotated with causal relations, followed by research efforts on automated extraction of event causality that leverage such resources. I will then discuss existing work on discovering causal common sense, which is important for constructing causal chains, taking into account triggering and enabling factors in the causal complex for an effect to happen.

5.2 Modelling Causal Relations

Counterfactual Model: The first inquiry previously mentioned, related to determining which eventualities are in or out of the causal complex, leads one to examine counterfactuals (Hobbs, 2005). According to the *counterfactual model* (Lewis, 1973), an event C is a cause of an event E if and only if it holds true that if C had not occurred, E would not have occurred. Considering our example where someone felt grateful and hugged his partner as a result, the situation is causal because if he had not felt grateful, he would not have hugged her.

There are several problems of the counterfactual model pointed out in Wolff (2007). One critical problem is that it might consider noncausal factors as causal (Wolff and Song, 2003; Neeleman and van de Koot, 2012). For instance, if both of his arms had not been free to move (or the extreme case, if he had not been born), then he would not have hugged her. Another issue with the counterfactual model is *overdetermination* (Sloman, 2005; Wolff, 2007). Consider an example where he hugged her not only because he felt grateful but also because he missed her presence. Based on the counterfactual principle, neither should be considered a cause because if he had not felt grateful, he would still have hugged her. To conclude, although counterfactual thinking can influence the causal reasoning, i.e., distinguishing eventualities that are in and out of causal complex, causal relationships cannot be reduced to counterfactual conditionals (Wolff, 2007).

Probabilistic Causation: Probabilistic causation designates a group of theories that aim to characterize the relationship between *causes* and *effects* using the tools of *probability theory* (Hitchcock, 2018). The central idea behind these theories is that causes *raise* the probabilities of their effects, which can be expressed formally using conditional probability:

$$P(E|C) > P(E|\neg C), \tag{5.1}$$

meaning that the probability that E occurs, given that C occurs, is higher than the probability that E occurs given that C does not occur (e.g., Reichenbach, 1956; Suppes, 1970).

https://doi.org/10.1017/9781108854221.007 Published online by Cambridge University Press

The *probabilistic contrast model* (Cheng and Novick, 1991) introduces the notion of *covariation*, $\Delta P = P(E|C) - P(E| \neg C)$, which is computed over a *focal set*, i.e., a set of events implied by the context E. When ΔP is greater than some (empirically determined) criterion, there should be a causal attribution to event C. In the case of event C constantly appearing in a focal set, leading to division by zero in computing the probability of the effect in the absence of event C, the causal status of such an event can only be determined by events in other focal sets, i.e., event C is (1) an *enabling condition* if it does co-vary with the effect in another focal set – namely, a set of events selected under another context, but (2) causally irrelevant if it does not co-vary with the effect in any other focal sets. Furthermore, this model distinguishes two main types of causal relationships: positive contrasts specify *facilitatory* causes (e.g., smoking *causes* cancer), whereas negative contrasts specify *inhibitory* causes (e.g., antioxidants *prevent* cancer).

The causation models based on probabilistic theory have a problem in establishing causal relationships on the basis of a single observation, because reliable probabilities depend on multiple observations (Tenenbaum and Griffiths, 2001). In the case of understanding causal relationships between events in natural language texts, the models are only applicable when we have sufficient explicitly correlated events. Such correlated events can be mined from large collections of narratives (e.g., news corpora); however, the reporting bias commonly present in news may distort the account of causation.

Dynamics Model: Another major approach towards causation involves *physicalist models*, which are built upon the assumption that causation can be described in terms of physical quantities such as energy, momentum, impact forces, chemical forces and electrical forces, among others. The *dynamics model* (Wolff and Song, 2003; Wolff et al., 2005; Wolff, 2007) is one of the prominent ones, which is based on Talmy's *force dynamic* account on causality (Talmy, 1988).

According to the dynamics model, causation involves interactions between two main entities: an *affector* and a *patient*. An affector is the entity that *acts on* a patient. The concept of causation, along with its related concepts, is captured in terms of three dimensions: (1) the patient tendency for the result; (2) the presence of concordance between the affector and the patient; and (3) the occurrence of the result (Wolff and Song, 2003). Table 5.1 shows the representation of Cause, Enable and Prevent concepts by the dynamics model.

In a Cause situation, such as 'Strong wind caused the bridge to collapse', the tendency of the bridge (patient) is not to collapse (result), but the strong wind (affector) does not act in concordance with the tendency, and the result

https://doi.org/10.1017/9781108854221.007 Published online by Cambridge University Press

Table 5.1. *Causality concepts represented by the dynamics model*

	Patient tendency for result	Affector–patient concordance	Occurrence of result
CAUSE	No	No	Yes
ENABLE	Yes	Yes	Yes
PREVENT	Yes	No	No

occurs. Meanwhile, in an ENABLE situation, as in 'Vitamin B enables the body to digest food', the tendency of the body (patient) is to digest food and Vitamin B (affector) acts in concordance, assisting the tendency, and the result occurs.

The dynamics model was tested by linking it with natural language; participants were tasked to sort 23 verbs expressing causality into groups according to the verbs' similarity to each other (Wolff and Song, 2003). The result indicated that these verbs fell into the three causal categories predicted by the model.

5.3 Causal Annotation in Natural Language Text

Natural language displays a great range of devices to express causal relations, including causative verbs such as *break* or *kill*, which express some contributing factors causing some entities to become *broken* or *dead*, prepositions such as *due to*, as well as discourse connectives such as *because* or *since* (Waldmann et al., 2017). In this section I will focus on attempts to annotate causal relations between different units of discourse (e.g., clauses, verbal events). I will expand on causative verbs later in Section 5.5, because such verbs implicitly encode causal commonsense knowledge.

Existing annotation schemes for causal relations can be generally distinguished into two approaches related to different discourse units used: *text spans* and *lexical units*. With the first approach, causation in 'He hugged her tight because he felt grateful' is annotated between *He hugged her tight* and *he felt grateful*, whereas *hugged* and *felt* are causally connected with the second approach.[1] Table 5.2 summarizes available resources annotated with causal relations.

[1] With the exception of PropBank (Palmer et al., 2005) that focuses on predicate–argument relations. Text spans denoting the reasons for an action are annotated as causal arguments (ARGM-CAU) of a lexical predicate, as in 'They [*moved* $_{\text{PREDICATE}}$] to London [*because of the baby* $_{\text{ARGM-CAU}}$]'.

https://doi.org/10.1017/9781108854221.007 Published online by Cambridge University Press

Table 5.2. *Resources annotated with causal relations*

Corpus	Size (no. of documents)	Discourse unit	No. of causal pairs
PDTB (Prasad et al., 2008)	2,159	Text span	6,289
BECauSE 2.0 (Dunietz et al., 2017)	119	Text span	1,634
SemEval-2007 (Girju et al., 2007)	–	Nominal	114
Do et al. (2011)	20	ACE event	414
Bethard et al. (2008)	556	TimeML event	271
Causal-TimeBank (Mirza and Tonelli, 2014)	183	TimeML event	318
CaTeRS (Mostafazadeh et al., 2016b)	320	TimeML event	488

Text Spans: The Penn Discourse Treebank (PDTB) corpus (Prasad et al., 2008) addresses the annotation of discourse relations that hold between exactly two text spans as arguments, which can be phrases, clauses or sentences. Such semantic relations are expressed either explicitly via lexical items (e.g., *because*) or implicitly via adjacency in discourse. In PDTB, the CAUSE relation is classified as a subtype of CONTINGENCY. Out of 102 known explicit discourse markers (e.g., *and*, *in contrast*), 28 explicitly mark causal relations (e.g., *as a result*, *consequently*). In addition to explicit markers, open-ended markers such as *This may help explain why* were also considered and annotated as *AltLex* relations. Among 6,289 annotated causal pairs, 2,099 are explicit and 273 contain an AltLex, and the rest are implicit causation (Hidey and McKeown, 2016).

The BECauSE 2.0 corpus (Dunietz et al., 2017) adopts the annotation scheme of PDTB but with a focus on annotating any form of *causal language* presented in the text with explicit lexical triggers. Each causal annotation consists of *cause* and *effect* spans, as well as a *causal connective* (e.g., *because of* or *opens the way for*). In addition to the distinction between positive causation (FACILITATE) and inhibitory causation (INHIBIT), three types of causation were considered: CONSEQUENCE (e.g., '[*They moved to London* $_{effect}$] **because of** [*the baby* $_{cause}$]'); MOTIVATION (e.g., '[*Their old apartment was too small* $_{cause}$], **so** [*they moved to London* $_{effect}$]'); and PURPOSE (e.g., '[*They moved to London* $_{effect}$] **so that** [*they could have a bigger house* $_{cause}$]').

https://doi.org/10.1017/9781108854221.007 Published online by Cambridge University Press

The corpus comprises a total of 1,803 sentences expressing causation, of which 1,634 sentences contain both cause and effect arguments.

Lexical Units: The *SemEval-2007* shared task on *Classification of Semantic Relations between Nominals* (Girju et al., 2007) gives access to a corpus containing nominal causal relations, among other semantic relations considered in the task. A nominal is defined as a noun or a noun phrase, excluding named entities and complex noun phrases (e.g., *the engine of the lawn mower*). Textual data was collected via wild-card search patterns; for instance, with '* cause *' as a query. The CAUSE–EFFECT relation is then annotated as true between e_1 and e_2 in sentences such as 'Happiness and [*laughter* $_{e_1}$] can cause [*wrinkles* $_{e_2}$]'.

Several works focus on annotating causal relations between *events* triggered by lexical units, following either *TimeML* (Pustejovsky et al., 2003) or *ACE* (Linguistic Data Consortium, 2005) annotation schemes for modelling events. Both TimeML and ACE define an event as *something that happens/occurs* or *a state that holds true*, which can be expressed by a verb, a noun, an adjective, as well as a nominalization either from verbs or from adjectives. However, both event models are designed for different purposes, hence resulting in different annotation of events. In addition to basic features of events existing in both models (tense, aspect, polarity and modality), ACE events have more complex structures involving *event arguments*, which can be either *event participants* or *event attributes* (location and time).[2] Though all events in TimeML are annotated, only 'interesting' events falling into a set of particular types and subtypes are annotated in ACE.

ACE Events: The event model adopted in Do et al. (2011) for identifying event causality is a simplification of the ACE events, in which only the subjects and objects of a *predicate* – a word triggering the presence of an event – are taken into account as *event arguments*, if any. Both verbal and nominal predicates were considered, allowing the detection of causality between *verb–verb*, *verb–noun* and *noun–noun* triggered event pairs. Event arguments were automatically extracted via dependency parsing. In order to evaluate their approach, they developed an evaluation corpus by collecting 20 news articles from CNN and annotating the causal pairs using two simple notions for

[2] Similar predicate–argument structures are found in FrameNet (Baker et al., 1998), which systematically describes semantic frames. FrameNet also captures relationships between frames, including, among others, *Precedes* and *Is Preceded by*, capturing temporal ordering, and *Is Causative of*, expressing causality.

https://doi.org/10.1017/9781108854221.007 Published online by Cambridge University Press

causality (C): (1) the *cause* should temporally precede the *effect*, and (2) the *effect* occurs because the *cause* occurs. In the case where the existence of causation is debatable, they annotate the pairs with *relatedness* (R) indication. Evaluated on 10 documents, the annotators agreed on 67% of 248 distinct event pairs having $C+R$ relations. However, they only agreed on 58% event pairs for having C relation, highlighting the difficulty of distinguishing causally related events. In total, 492 $C + R$ and 414 C relation annotations were obtained.

TimeML Events: Bethard et al. (2008) utilized an existing event extraction system for identifying TimeML events in the Penn Treebank corpus (Marcus et al., 1993) and then annotated 1,000 pairs of events conjoined with the conjunction *and*. The extracted event pairs were annotated with both temporal and causal relations in parallel. For causal relations, the authors investigated two annotation guidelines, with the latter shown to be more robust: (1) identifying NECESSARY and SUFFICIENT events based on the counterfactual model and (2) choosing between CAUSAL vs NO-REL by paraphrasing the word *and* with, for instance, *and as a result* vs *and independently*, respectively. The second guideline is relatively simple for annotators, but agreement is only moderate (kappa of 0.556), in part because there are both causal and noncausal readings of such connective phrases.

Mirza and Tonelli (2014) augmented the TimeBank corpus (Pustejovsky et al., 2006) taken from the TempEval-3 shared task (UzZaman et al., 2013) – containing gold annotated TimeML events, temporal information and ordering – with causal information. The annotators relied on explicit *causal signals* (marked as CSIGNAL) to identify *causal relations* (annotated with CLINK, analogous to TLINK for temporal relations in TimeML) between events. A CLINK was established between two events when one of the considered causal constructions was identified, including, among others (Mirza and Tonelli, 2014; Mirza et al., 2014), expressions containing CSIGNALS such as 'Its shipments [*declined* $_{\text{EFFECT}}$] **as a result of** a [*reduction* $_{\text{CAUSE}}$] in inventories' and periphrastic constructions involving verbs associated with Cause, Enable and Prevent relations listed in Wolff and Song (2003), as exemplified in 'The [*blast* $_{\text{CAUSE}}$] **caused** the boat to [*heel* $_{\text{EFFECT}}$] violently'. The total number of annotated CLINKS in the resulting Causal-TimeBank corpus (Mirza and Tonelli, 2014) is 318, with Dice's coefficient of 0.73 on a subset of five documents.

The CaTeRS annotation scheme (Mostafazadeh et al., 2016b) looked at causality between events more from a commonsense reasoning standpoint rather than linguistic markers. The authors annotated 320 stories from the ROCStories Corpus (Mostafazadeh et al., 2016a) with events and semantic

https://doi.org/10.1017/9781108854221.007 Published online by Cambridge University Press

relations, both temporal and causal. The semantic relation annotation was initiated with deciding whether one of nine causal relation types (combinations of CAUSE, ENABLE, PREVENT and CAUSE_TO_END relations with valid temporal relations such as BEFORE and OVERLAP) occur, followed by choosing one of existing temporal relation types (BEFORE, OVERLAP, CONTAIN, IDENTITY) when no causality was detected. For instance, given a sentence 'It was [*raining* $_{e_1}$] so hard that it prevented me from [*going* $_{e_2}$] to school', a PREVENT–OVERLAP relation is established between e_1 and e_2.[3] There are overall 488 causal links, with CAUSE–BEFORE being the most frequent type, and moderate agreement of the semantic link annotation in general (kappa of 0.49).

5.4 Extracting Event Causality

Understanding discourse relations that are expressed in natural language texts is important for downstream natural language processing applications such as question-answering and decision-making support systems. Recognizing causal relations in particular is necessary to reconstruct a causal chain of events as part of story or narrative understanding. Existing annotated resources for causal relations allow us to consider automatic approaches for identifying causation in natural language texts, both as training data to build machine learning models and as evaluation benchmarks to measure the performance of the automatic methods.

Even though there exists a body of work on the extraction of causal relations between nominals (e.g., Dasgupta et al., 2018), as well as work leveraging the PDTB corpus for the automatic detection of discourse relations in general (e.g., Pitler and Nenkova, 2009; Liu and Li, 2016) or causal relations specifically (e.g., Hidey and McKeown, 2016), in this section I will focus only on automatic extraction systems developed for identifying causality between *events* triggered by lexical items, for each event model previously discussed in Section 5.3, using the corresponding annotated data.

ACE Event Causality: Do et al. (2011) explored a minimally supervised approach by devising a causality measure in terms of *cause–effect association (CEA)* between two events e_i and e_j, which takes into account the *pointwise mutual information* (PMI) between (1) event predicates, (2) the predicate of

[3] This example illustrates the temporal implications of causation, in which the *cause* start before the *effect*; however, there is no restriction on their relative ending, allowing the OVERLAP relation to happen.

https://doi.org/10.1017/9781108854221.007 Published online by Cambridge University Press

an event and the arguments of the other event, and (3) event arguments. Furthermore, they leveraged the interactions between event causality and discourse relation predictions through a global inference procedure, which can be formalized via an integer linear programming (ILP) framework as a constraint optimization problem (Roth and Yih, 2004). One introduced constraint is related to the types of predicted discourse connectives (e.g., CAUSE, CONDITION) that are allowed to enclose a causal event pair. Evaluated on the evaluation corpus they built, mentioned in Section 5.3, the system achieves a performance of 41.7% F1-score on extracting causality between events.

TimeML Event Causality: Bethard and Martin (2008) built a classification model for extracting causal relations using a subset of the annotated corpus of parallel temporal and causal relations described in Section 5.3. They used 697 out of 1,000 event pairs to train a support vector machine (SVM) classifier using both syntactic and semantic features and used the rest for evaluating the system, resulting in a 37.1% F1-score. The performance of the causal relation classifier is significantly boosted to a 52.4% F1-score by additionally exploiting gold-standard temporal labels as features. Rink et al. (2010) performed textual graph classification using the same event causality corpus built by Bethard and Martin (2008), leveraging syntactic features as well as the hypernym chain for the senses resulting from word sense disambiguation. Following Bethard and Martin (2008), they also made use of manually annotated temporal relation types as a feature to build the classification model. This results in a 57.9% F1-score, a 15 percentage point increase compared with the system without the additional feature of temporal relations. However, the system proposed by Bethard and Martin (2008) for automatically extracting temporal links yields only a 49% F1-score. Hence, it is highly likely that leveraging automatically extracted temporal labels as features will show no benefit and propagate errors instead.

Mirza and Tonelli (2016) proposed a hybrid approach for the extraction and classification of both temporal and causal relations between events from English documents, which are pre-annotated with TimeML events and temporal expressions. They developed CATENA (CAusal and TEmporal relation extraction from NAtural language texts), which consists of two main modules, one for temporal and the other for causal; both modules rely on a *sieve-based architecture* introduced in Chambers et al. (2014) for event ordering, in which the remaining unlabelled pairs – after running a rule-based component and/or a transitive reasoner – are fed into a supervised classifier.

In the causal module, the rule-based (RB) sieve is responsible for identifying causal constructions involving *affect* and *link* verbs presented in

https://doi.org/10.1017/9781108854221.007 Published online by Cambridge University Press

Mirza and Tonelli (2014), as well as *periphrastic causative* verbs associated with Cause, Enable and Prevent listed in Wolff and Song (2003), which are further expanded using the Paraphrase Database (PPDB) by Ganitkevitch et al. (2013). The machine-learned (ML) sieve is a linear SVM classifier leveraging various features including syntactic and semantic features, event attributes and features related to temporal and causal signals, which are also extended with PPDB phrases.

The authors made use of the Causal-TimeBank corpus (see Section 5.3) to train the supervised classifier and additionally annotated 20 TempEval-3-platinum documents (UzZaman et al., 2013) with causal links as the evaluation set, following the same annotation schemes and guidelines. Whereas the data-driven ML sieve achieves only an 18.2% F1-score, the combination of RB and ML sieves yields a 62.2% F1-score, a 4.3 percentage point increase compared with the standalone RB sieve, with the ML sieve contributing to boosting the recall of the highly precise RB sieve (recall moves up from 42.3% to 53.8%).

The low performance of the ML module is mostly due to dependency parsing mistakes, especially for long sentences. For instance, a causal link was established between *acquire* and *respond* in 'StatesWest Airlines, Phoenix, Ariz., said it [*withdrew* $_{\text{EFFECT}}$] its offer to [*acquire*] Mesa Airlines **because** the Farmington, N.M., carrier didn't [*respond* $_{\text{CAUSE}}$] to its offer . . . ' instead of between *withdrew* and *respond*, as a result.

Explicit vs Implicit Causal Signals: Another issue of the ML sieve in CATENA is the difficulty in disambiguating signals such as *and* and *since*. As previously mentioned above, Bethard and Martin (2008) and Rink et al. (2010) tackled this specific problem of classifying event pairs conjoined with the conjuction *and*, resulting in only 52.4% and 57.9% F1-scores respectively with the help of gold-annotated temporal labels. One possible explanation for the low performance is that a data-driven approach for this particular task would require sufficient training data of causal event pairs in order to uncover causal common sense; for example, that feeling grateful may be one of the reasons for hugging someone. Discovering such causal common sense will also be beneficial for identifying causal relations when there are no markers at all, which are more common in natural language texts (e.g., 62% of causal pairs in the PDTB corpus are of implicit nature). I will discuss existing efforts in this line of research of discovering causal common sense in the following section.

5.5 Causal Commonsense Discovery

Discovering causal common sense is crucial for recognising latent causation that is implicitly present in natural language text; for instance, to notice the causal relations between ‘helped’ and ‘hugged’ events in ‘She helped him a lot. He hugged her gratefully’. As humans we understand that we would feel thankful if someone helped us, and feeling thankful often leads to hugging the person who helped us. Furthermore, it is important to note that we can only hug someone when that person is in close proximity to us. Such background knowledge is a necessity for an automatic method to be able to reconstruct and to reason about causal chains among events, taking into account triggering and enabling factors that persist in the causal complex for the causation to hold. In this section, I will discuss several attempts on building a repository of causal common sense to serve its purposes.

Causative Verbs: Causative verbs were briefly mentioned in Section 5.3 and were defined as verbs that implicitly encode states resulting from some events and contributing factors (Neeleman and van de Koot, 2012; Waldmann et al., 2017). For example, the verb ‘break’ in ‘Mary broke the vase’ may express ‘some event involving Mary was the cause of the vase becoming broken’. Such linguistic representation of causative verbs can be characterized as $[\ldots x \ldots]_{e_1}$ CAUSE $[[\ldots y \ldots]_s]_{e_2}$, which denotes ‘event e_1 involving entity x was the cause of event e_2 of entity y becoming in state s’. Note that there is no universal agreement among linguists regarding the representation of causative verbs (see, for instance, the discussion in Neeleman and van de Koot, 2012). Further accounts on causative verbs related to (in)direct causation and resultative contructions (as in ‘John hammered the metal flat’) were discussed in Waldmann et al. (2017).

In VerbNet (Kipper et al., 2008), causative verbs may be identified via semantic frames of some verb classes (e.g., BREAK-45.1), in which semantic predicates CAUSE(AGENT, E) and *state*(RESULT(E), PATIENT) are present; *state* denotes semantic predicates in VerbNet related to the resulting states, such as DEGRADATION_MATERIAL_INTEGRITY for BREAK-45.1. In FrameNet (Baker et al., 1998), we may assume that verbs belonging to CAUSE_* frames, such as CAUSE_TO_FRAGMENT or CAUSE_TEMPERATURE_CHANGE, are causative verbs.

Event Causality: A few resources on word/frame semantics such as WordNet (Fellbaum, 1998), FrameNet (Baker et al., 1998) and VerbOcean (Chklovski and Pantel, 2004) provide information about causal relations between lexical units that can be realized as events. However, they suffer from limited coverage due to having been manually constructed, or due to the semi-automatic approach adopted on highly implicit nature of language.

Riaz and Girju (2013) focused on the identification of causal relations in verb pairs. For example, *kill–arrest* has a higher likelihood of encoding causation than *build–maintain*. They relied on the unambiguous discourse markers *because* and *but* to automatically collect training instances of causal and non-causal event pairs, respectively. By utilising a set of metrics capturing causal associations, which exploit information available from a large number of unlabeled verb pairs, the result is a knowledge base of causal associations of verbs (KB_c). In KB_c, roughy 10K verb pairs are categorized as *Strongly Causal* (S_c), *Ambiguous* (A_c) and *Strongly Non-causal* ($S_{\neg c}$), based on the likelihood of causality encoded by the pairs, assuming uniform distribution across three categories.

One limitation of prior work in event causality is that it has primarily focused on newswire, limiting the causal understanding to newsworthy events, instead of everyday events such as human activities in general. Hu et al. (2017) explored unsupervised methods for modelling causality to learn event relations from blogs (topically focused on *camping*) and movie scene descriptions (excluding dialogues), resulting in high-quality causal pairs (e.g., ⟨*person, pack up*⟩ → ⟨*person, go home*⟩). Over 80% of pairs were indeed judged as causal by human annotators. The authors relied on the *Causal Potential* (CP) measure, based on Suppes' probabilistic account of causation (Suppes, 1970), to assess the causal relation between events. Subsequent work by Hu and Walker (2017) further explored four different types of *narrative causality* presenting in film scene descriptions, including PHYSICAL (A physically causes B), MOTIVATIONAL (A happens with B as a motivation), PYSCHOLOGICAL (A brings about emotions expressed in B) and ENABLING (A enables B).

ATOMIC[4] (Sap et al., 2019) introduced a commonsense repository for everyday events, causes and effects. Recall our example event earlier 'He hugged her tight'. By querying ATOMIC with 'PersonX hugs PersonY tight' we obtained the following as results: (1) *Causes of PersonX* (e.g., 'Because PersonX wanted to show their love and level of care'), (2) *Attributes of PersonX* (e.g., 'PersonX is seen as emotional'), (3) *Effects on PersonX* (e.g., 'As a result, PersonX feels happy') and (4) *Effects on others* (e.g., 'As a result, others feel loved').

[4] https://mosaickg.apps.allenai.org/

The ATOMIC repository was built through answering questions about an event with regards to *If–Then* knowledge, collected via crowdsourcing. The scheme distinguishes nine different relations belonging to three types of *If–Then* relations: *if-Event-then-Mental-State*, *if-Event-then-Event* and *if-Event-then-Persona*. Around 24K common *event phrases* were extracted for the annotation from various corpora, including stories, books, Google Ngrams and Wiktionary idioms. Considered as events are verb phrases with a verb predicate and its argument like ‘drinks coffee in the morning’. For a more general representation of events, tokens referring to people were replaced with a `Person` variable; e.g., ‘PersonX drinks coffee in the morning’. The resulting knowledge graph of ATOMIC contains over 300K nodes.

Recent progress in training deep contextualized language models using large-scale text corpora (e.g., BERT by Devlin et al., 2019) gives rise to explorations beyond extractive methods as an avenue for discovering commonsense knowledge. COMmonsEnse Transformers (COMET; Bosselut et al., 2019) was proposed to learn to generate novel commonsense tuples by leveraging such language models, given existing tuples in ATOMIC as a seed set of knowledge. Evaluated on 100 sampled events, this approach yields 56.5% precision (at 10) of generated inferences that are scored by human judges as being plausible or not. Anecdotal examples of generated causal pairs include ‘As a result, PersonX gets fat’ for ‘PersonX eats red meat’. Though this is an interesting approach for discovering causal commonsense pairs, the plausibility of the generated tuples depends a lot on the textual corpus used to train the language model, which could suffer from the reporting bias.

Event Circumstantiality. Commonsense knowledge about causes and effects of strongly causal links is not sufficient to reconstruct coherent causal chains as we need to take into account all triggering and enabling factors in the causal complex of a given narrative. For instance, in order to *hug* someone, that person must *be in close proximity* as a precondition. Vossen et al. (see Chapter 6) adopt a broader class of event relations comprising causality, enablement, prevention and entailment – termed *circumstantiality*. The Circumstantial Event Ontology (CEO; Segers et al., 2018) was proposed to capture such a wide range of circumstantial relations. Events are circumstantially related if there exist shared properties in the pre, during, and post situations. As an example, the property ‘fire exist true’ could tie a circumstantial relation from *ceo:Arson* to *ceo:ExtinguishingFire* because the presence of fire is a post situation of an arson event and a pre situation of a fire-extinguishing event. The ontology consists of 223 event classes of which 189 are fully modeled with pre, during and post situations.

5.6 Conclusions

In this chapter, I have discussed research endeavours in identifying causal chains, or *causelines*, in order to understand and reconstruct coherent stories in narratives. Such causal chains can be constructed from identifying causal links between events presented in a narrative. Causal expressions in natural language texts may contain explicit causal markers, which are identifiable by automated extraction systems. I have reviewed existing work on causal extraction systems, which leverage language resources annotated with causal relations. I have listed several such corpora using different annotation schemes, guided by existing causal models in psychology and psycholinguistic that I have summarized briefly in the beginning of this chapter.

However, causation in language is more commonly expressed implicitly via adjacency in discourse. In this case, existing data-driven causal relation extraction systems are lacking in discovering the underlying causal links, due to limited training data available. Furthermore, circumstantial events surrounding the causal complex are rarely expressed with language as they are part of human common sense. Hence, discovering causal common sense is crucial to fill the gaps in the causal chains and the causal complex. I have discussed existing work in this line of research of causal commonsense discovery.

References

Baker, Collin F., Fillmore, Charles J., and Lowe, John B. 1998. The Berkeley FrameNet Project. Pages 86–90 of: *Proceedings of the 17th International Conference on Computational Linguistics: Volume 1*. Montreal, QC, Canada: Association for Computational Linguistics.

Bethard, Steven, and Martin, James H. 2008. Learning Semantic Links from a Corpus of Parallel Temporal and Causal Relations. Pages 177–180 of: Moore, Johanna D., Teufel, Simone, Allan, James, and Furui, Sadaoki (eds.), *Proceedings of ACL'08: HLT, Short Papers*. Columbus, OH: Association for Computational Linguistics.

Bethard, Steven, Corvey, William, Klingenstein, Sara, and Martin, James H. 2008. Building a Corpus of Temporal-Causal Structure. In: Calzolari, Nicoletta, Choukri, Khalid, Maegaard, Bente, et al. (eds.), *Proceedings of LREC'08*. Marrakech, Morocco: European Language Resources Association (ELRA).

Bosselut, Antoine, Rashkin, Hannah, Sap, Maarten, Malaviya, Chaitanya, Celikyilmaz, Asli, and Choi, Yejin. 2019. COMET: Commonsense Transformers for Automatic Knowledge Graph Construction. Pages 4762–4779 of: Korhonen, Anna, Traum, David, and Màrquez, Lluís (eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics.

Chambers, Nathanael, Cassidy, Taylor, McDowell, Bill, and Bethard, Steven. 2014. Dense Event Ordering with a Multi-pass Architecture. *Transactions of the Association for Computational Linguistics*, **2**, 273–284.

Cheng, Patricia W., and Novick, Laura R. 1991. Causes versus Enabling Conditions. *Cognition*, **40**(1–2), 83–120.

Chklovski, Timothy, and Pantel, Patrick. 2004. VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations. Pages 33–40 of: Lin, Dekang, and Wu, Dekai (eds.), *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. Barcelona, Spain: Association for Computational Linguistics.

Dasgupta, Tirthankar, Saha, Rupsa, Dey, Lipika, and Naskar, Abir. 2018. Automatic Extraction of Causal Relations from Text Using Linguistically Informed Deep Neural Networks. Pages 306–316 of: Komatani, Kazunori, Litman, Diane, Yu, Kai, et al. (eds.), *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. Melbourne, Australia: Association for Computational Linguistics.

Devlin, Jacob, Chang, Ming-Wei, Lee, Kenton, and Toutanova, Kristina. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Pages 4171–4186 of: Burstein, Jill, Doran, Christy, and Solorio, Thamar (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, MN: Association for Computational Linguistics.

Do, Quang, Chan, Yee Seng, and Roth, Dan. 2011. Minimally Supervised Event Causality Identification. Pages 294–303 of: Barzilay, Regina, and Johnson, Mark (eds.), *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Edinburgh: Association for Computational Linguistics.

Dunietz, Jesse, Levin, Lori, and Carbonell, Jaime. 2017. The BECauSE Corpus 2.0: Annotating Causality and Overlapping Relations. Pages 95–104 of: Schneider, Nathan, and Xue, Nianwen (eds.), *Proceedings of the 11th Linguistic Annotation Workshop*. Valencia, Spain: Association for Computational Linguistics.

Fellbaum, Christiane (ed.). 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Ganitkevitch, Juri, Van Durme, Benjamin, and Callison-Burch, Chris. 2013. PPDB: The Paraphrase Database. Pages 758–764 of: Vanderwende, Lucy, Daumé III, Hal, and Kirchhoff, Katrin (eds.), *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta, GA: Association for Computational Linguistics.

Girju, Roxana, Nakov, Preslav, Nastase, Vivi, Szpakowicz, Stan, Turney, Peter, and Yuret, Deniz. 2007. SemEval-2007 Task 04: Classification of Semantic Relations between Nominals. Pages 13–18 of: Agirre, Eneko, Màrquez, Lluís, and Wicentowski, Richard (eds.), *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*. Prague, Czech Republic: Association for Computational Linguistics.

Hidey, Christopher, and McKeown, Kathy. 2016. Identifying Causal Relations Using Parallel Wikipedia Articles. Pages 1424–1433 of: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Berlin: Association for Computational Linguistics.

Hitchcock, Christopher. 2018. Probabilistic Causation. In: Zalta, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Hobbs, Jerry R. 2005. Toward a Useful Concept of Causality for Lexical Semantics. *Journal of Semantics*, **22**, 181–209.

Hu, Zhichao, and Walker, Marilyn. 2017. Inferring Narrative Causality between Event Pairs in Films. Pages 342–351 of: Jokinen, Kristiina, Stede, Manfred, DeVault, David, and Louis, Annie (eds.), *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*. Saarbrücken, Germany: Association for Computational Linguistics.

Hu, Zhichao, Rahimtoroghi, Elahe, and Walker, Marilyn. 2017. Inference of Fine-Grained Event Causality from Blogs and Films. Pages 52–58 of: Caselli, Tommaso, Miller, Ben, van Erp, Marieke, et al. (eds.), *Proceedings of the Events and Stories in the News Workshop*. Vancouver: Association for Computational Linguistics.

Kipper, Karin, Korhonen, Anna, Ryant, Neville, and Palmer, Martha. 2008. A Large-Scale Classification of English Verbs. *Language Resources and Evaluation*, **42**(1), 21–40.

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

Linguistic Data Consortium. 2005. *ACE (Automatic Content Extraction) English Annotation Guidelines for Events*.

Liu, Yang, and Li, Sujian. 2016. Recognizing Implicit Discourse Relations via Repeated Reading: Neural Networks with Multi-Level Attention. Pages 1224–1233 of: Su, Jian, Duh, Kevin, and Carreras, Xavier (eds.), *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, TX: Association for Computational Linguistics.

Marcus, Mitchell P., Marcinkiewicz, Mary Ann, and Santorini, Beatrice. 1993. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, **19**(2), 313–330.

Mirza, Paramita, and Tonelli, Sara. 2014. An Analysis of Causality between Events and Its Relation to Temporal Information. Pages 2097–2106 of: Tsujii, Junichi, and Hajic, Jan (eds.), *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin, Ireland: Dublin City University and Association for Computational Linguistics.

Mirza, Paramita, and Tonelli, Sara. 2016. CATENA: CAusal and TEmporal Relation Extraction from NAtural Language Texts. Pages 64–75 of: Matsumoto, Yuji, and Prasad, Rashmi (eds.), *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. Osaka, Japan: The COLING 2016 Organizing Committee.

Mirza, Paramita, Sprugnoli, Rachele, Tonelli, Sara, and Speranza, Manuela. 2014. Annotating Causality in the TempEval-3 Corpus. Pages 10–19 of: Kolomiyets, Oleksandr, Moens, Marie-Francine, Palmer, Martha, Pustejovsky, James, and Bethard, Steven (eds.), *Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL)*. Gothenburg, Sweden: Association for Computational Linguistics.

Mostafazadeh, Nasrin, Grealish, Alyson, Chambers, Nathanael, Allen, James, and Vanderwende, Lucy. 2016b. CaTeRS: Causal and Temporal Relation Scheme for Semantic Annotation of Event Structures. Pages 51–61 of: Palmer, Martha, Hovy, Ed, Mitamura, Teruko, and O'Gorman, Tim (eds.), *Proceedings of the Fourth Workshop on Events*. San Diego, California: Association for Computational Linguistics.

Mostafazadeh, Nasrin, Chambers, Nathanael, He, Xiaodong, Parikh, Devi, Batra, Dhruv, Vanderwende, Lucy, Kohli, Pushmeet, and Allen, James. 2016a. A Corpus and Cloze Evaluation for Deeper Understanding of Commonsense Stories. Pages 839–849 of: Knight, Kevin, Nenkova, Ani, and Rambow, Owen (eds.), *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, CA: Association for Computational Linguistics.

Neeleman, Ad, and van de Koot, Hans. 2012. The Linguistic Expression of Causation. Pages 20–51 of: *The Theta System: Argument Structure at the Interface*. Oxford: Oxford University Press.

Palmer, Martha, Gildea, Daniel, and Kingsbury, Paul. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, **31**(1), 71–106.

Pitler, Emily, and Nenkova, Ani. 2009. Using Syntax to Disambiguate Explicit Discourse Connectives in Text. Pages 13–16 of: Su, Keh-Yih, Su, Jian, Wiebe, Janyce, and Li, Haizhou (eds.), *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*. Suntec, Singapore: Association for Computational Linguistics.

Prasad, Rashmi, Dinesh, Nikhil, Lee, Alan, Miltsakaki, Eleni, Robaldo, Livio, Joshi, Aravind, and Webber, Bonnie. 2008. The Penn Discourse TreeBank 2.0. In: Calzolari, Nicoletta, Choukri, Khalid, Maegaard, Bente, et al. (eds.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco: European Language Resources Association.

Pustejovsky, James, Castao, José, Ingria, Robert, et al. 2003. TimeML: Robust Specification of Event and Temporal Expressions in Text. Pages 28–34 of: Maybury, Mark T. (ed.), *New Directions in Question Answering*. AAAI Press.

Pustejovsky, James, Littman, Jessica, Saurí, Roser, and Verhagen, Marc. 2006. *TimeBank 1.2 Documentation*. Technical Report, Brandeis University.

Reichenbach, Hans. 1956. *The Direction of Time*. Dover Publications.

Riaz, Mehwish, and Girju, Roxana. 2013. Toward a Better Understanding of Causality between Verbal Events: Extraction and Analysis of the Causal Power of Verb–Verb Associations. Pages 21–30 of: Eskenazi, Maxine, Strube, Michael, Di Eugenio, Barbara, and Williams, Jason D. (eds.), *Proceedings of the SIGDIAL 2013 Conference*. Metz, France: Association for Computational Linguistics.

Rink, Bryan, Bejan, Cosmin Adrian, and Harabagiu, Sanda M. 2010. Learning Textual Graph Patterns to Detect Causal Event Relations. In: *FLAIRS Conference*.

Roth, Dan, and Yih, Wen-tau. 2004. A Linear Programming Formulation for Global Inference in Natural Language Tasks. Pages 1–8 of: *Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004) at HLT-NAACL 2004*. Boston: Association for Computational Linguistics.

Sap, Maarten, LeBras, Ronan, Allaway, Emily, et al. 2019. ATOMIC: An Atlas of Machine Commonsense for If–Then Reasoning. In: *Proceedings of AAAI'19*.

Segers, Roxane, Caselli, Tommaso, and Vossen, Piek. 2018. The Circumstantial Event Ontology (CEO) and ECB+/CEO: An Ontology and Corpus for Implicit Causal Relations between Events. In: Calzolari, Nicoletta, Choukri, Khalid, Cieri, Christopher, et al. (eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association.

Sloman, Steven. 2005. *Causal Models: How People Think about the World and Its Alternatives*.

Suppes, Patrick. 1970. *A Probabilistic Theory of Causality*. Amsterdam: North-Holland Publishing Co.

Talmy, Leonard. 1988. Force Dynamics in Language and Cognition. *Cognitive Science*, **12**(1), 49–100.

Tenenbaum, Joshua B., and Griffiths, Thomas L. 2001. Structure Learning in Human Causal Induction. Pages 59–65 of: Leen, T., Dietterich, T., and Tresp, V. (eds.), *Structure Learning in Human Causal Induction*, Vol. 13. MIT Press.

UzZaman, Naushad, Llorens, Hector, Derczynski, Leon, et al. 2013. SemEval-2013 Task 1: TempEval-3: Evaluating Time Expressions, Events, and Temporal Relations. Pages 1–9 of: Manandhar, Suresh, and Yuret, Deniz (eds.), *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. Atlanta, GA: Association for Computational Linguistics.

Vossen, Piek, Caselli, Tommaso, and Segers, Roxane. 2021. A Narratology-based Framework for Storyline Extraction. Pages 130–147 of: Caselli, Tommaso, Palmer, Martha, Hovy, Eduard, and Vossen, Piek (eds), *Computational Analysis of Storylines: Making Sense of Events*. Cambridge University Press.

Waldmann, Michael R., Solstad, Torgrim, and Bott, Oliver. 2017. *Causality and Causal Reasoning in Natural Language*.

Wolff, Phillip. 2007. Representing Causation. *Journal of Experimental Psychology: General*, **136**(1), 82–111.

Wolff, Phillip, Klettke, Bianca, Ventura, Tatyana, and Song, Grace. 2005. Expressing Causation in English and Other Languages. Pages 29–48 of: *Categorization Inside and Outside the Laboratory: Essays in Honor of Douglas L. Medin. APA Decade of Behavior Series*. Washington, DC: American Psychological Association.

Wolff, Phillip, and Song, Grace. 2003. Models of Causation and the Semantics of Causal Verbs. *Cognitive Psychology*, **47**(3), 276–332.