AUTOMATIC DISCOVERY OF COMPLEX CAUSALITY

BY

CHEN LI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Associate Professor C. Roxana Girju, Chair
Professor Mark Hasegawa-Johnson
Professor Peter Lasersohn
Associate Professor Chi-lin Shih

Abstract

This study entails the understanding of and the development of a computational method for automatically extracting complex expressions in language that correspond to event to event sequential relations in the real world. We here develop component procedures of a system that would be capable of taking raw linguistic input (such as those from narrative writings or social network data), and find real-world semantic relations among events. Such an endeavor is applicable to many types of sequential relations, for which we use causality as a case study, both for its importance as a prominent type of sequential relation between events, as well as for its general prevalence in natural language. But we also demonstrate that the idea is also applicable in principle to other major types of event to event relations, such as reciprocity.

The study primarily focuses on those types of causalities that contain complex structures and require in-depth linguistic analyses to discover and extract. Designing an automated method for the extraction of structurally complex causal expressions entails methodologies and theories that are beyond conventional methods used in computational semantics. The classes of adjunctive causal structure, and embedded causal structure are types that are hard to access using traditional methods, but more amenable for methods developed in this study. The principal procedures employed for the extraction of these are a heavily modified form of Hidden Markov Model (HMM), which we use to deal with causal structures that have sequentially complex makeup. We also designed a highly modified Genetic Algorithm (GA) adapted for embedded context-free structures, used to rank and extract those causal structures that have deep embedding at the syntax-semantics interface. These will be reformulated, augmented, and explored in depth.

With these methods using unsupervised and semi-supervised learning, we were able to obtain reasonable results in terms of discrimination of causal pairs $\langle e_i, e_j \rangle$ pairs and some longer chains of causation from corpora. From these results, we were also able to perform additional linguistic analysis over their theoretical semantic structure, and observe aspects of each that allows us to sub-classify the relations according to standard ideas in formal logic as well as from behavioral psychology. These methods would be critical to a system for building a graph theoretic representation of a social network, from corpora produced by entities within that network, which would utilize the methods described in this project, and similar approaches can be extended to model and discover other types of complex event-relations. These types of fundamental technologies, would in turn, help us to design and build the types of on-line and mobile services that provide increased machine awareness of user behavior and to be able to target and cater to users individually.

Dedication

I would like to dedicate my dissertation work to my grandmother, Qu Shunqing, who passed away due to illness in 2013, during my thesis writing process; and to Professor Frederick Schwink of Germanics at UIUC, who also passed away in 2010, during my years of working on my thesis research, who had a profound impact on my my interest in linguistics.

Acknowledgments:

Much thanks to my advisor C. Roxana Girju for aid and guidance over the past several years, that allowed me to tackle this difficult topic without many predecessors; and to the remainder of my committee Mark Hasegawa-Johnson, Peter Lasersohn, and Chilin Shih for help and advise over numerous issues. My advisor and other committee members have been working with me through many difficult issues to get to this point, and all of them proved indispensable through this process.

Also thanks for the support and encouragement over the years from a number of faculty members from Linguistics, including Elabbas Benmamoun, Rakesh Bhatt, Jennifer Cole, Tania Ionin, Ryan Shosted, and James Yoon; and professors from other departments, including Laxmikant Kale, Steven Lavalle, Carl Niekirk, David Sanone, Gregory Whitt, and Craig Zilles. All of the above individuals have contributed to my graduate school and career that eventually led to the completion of my thesis research.

Personal Communications:

Aside from the regular advise from my committee members, I have consulted a number of other University faculty members, who provided me with valuable and informative discussions and technical suggestion over the last several years that provided helpful insights related to my research. These include:

1. Department of Mathematics:
    - Alexandr Kostochka
    - Douglas West
2. Department of Library and Information Science:
    - Les Gasser
3. Department of Computer Science:
    - Matthew Caesar
    - Julia Hockenmaier
    - Viraj Kumar
    - Manoj Prabhakaran
    - Josep Torellas
4. Department of Electrical and Computer Engineering
    - Martin Wong

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Causal expression in language is one of the most morphosyntactically as well as semantically diverse class of complex features in speech and text. Gaining a better understanding of causality in language and developing methods for extracting their various complex forms has great utility in both linguistic theory at their syntacto-semantic interface, as well as applications in information systems focused on entities and networks (mobile, social media, etc). This study is an effort to enable a better grasp of the former, and then leveraging that information to devise better methods for the extraction of complex causal structures that have been difficult in the past.

1.1 Causality

The quest to understand causality is as ancient as philosophy, the study of language, and the many branches of natural sciences. It is one of the primary ordering principles in most logical systems in the world, and at the same time provides sequence and coherence to our speech or thought. There is no generally agreed on logical formulation for causality, but can be concisely described as a relation between two events in sequence, the *cause* and the *effect*. There are numerous conceptions of the meaning of cause in philosophical thought over time, and causality is a necessary pre-requisite or outcome of various other critical concepts of describing the real-world, such as time, motion, process, probability, potentiality, etc.

In Aristotelian terms, there are several notions of causation that are mutually different and yet complementary. One may speak of some causation as an object or substance that in some way participates in the constituency of the effect, such as the cultivation and existence of coffee allows for the existence of coffee shops; this is usually termed *material cause.* One may speak of some causation as a pattern or mathematical formulation which other concepts use as a blueprint of some aspect of it, such as the concept of the *addition operator* being logically important for the formulation of a *polynomial*; this is usually termed *formal cause.* One may speak of some causation as an entity that affects the dynamical state of another entity in some way, such as a passing star close to the solar system perturbs some long-period comets so that they fall toward the inner solar system; this is usually termed *efficient cause.* One may also speak some causation as a sequence of actions with the expressed purpose and aim of resulting in some final event, such as spending time at night in local venues to eventually meet a marriage partner; this is usually termed *teleology.* Even processes that are driven entirely by nature without volition, such as evolution, maybe accorded *teleology* when there is an implicit goal that the process is directed toward, such ad adaptation in a specific ecological niche.

All of these different conceptions of cause have their own strengths in different domains of life and knowledge, such as *efficient cause* is critical to the natural sciences, and *teleology* is an essential component of social behavior. So causation, even in its most abbreviated set of conceptions, is not a monolithic logical concept, and largely relies on the perception of individuals to determine whether some causal relation between events is valid in the real-world. There are potential disagreements among individuals and between schools of philosophical thoughts in many instances. The fact that causation is this relatively fragmentary concept, and relies on individuals to judge causality between events, brings about the greatest difficulty in accurately discriminating between causal and non-causal relations, before any discussion of its actual linguistic forms.

1.2 Causal relations in language

Causation is an indispensable concept in the human understanding of events and relations in the real-world, and its use permeates clausal and discourse level expressions in human language. In their morphosyntactic forms, the expression of causality has a large variety of manifestations, ranging simple causative-inchoatives, where a single predicate presents both the causing and the caused event, to long distance causal chains that require sophisticated discourse analysis to decipher. And at a deeper semantic level, there are different properties of their real-world extensions in terms of entities, events, and relations, that can also be characterized in numerous ways, such as the previously mentioned Aristotelian causal archetypes, such as prescribed in his works φυσική ἀκρόασις (before 350 B.C. with some uncertainty) and τὰ μετὰ τὰ φυσικά (before 322 B.C.).

Some causalities are expressed through long-range discourse level mechanisms, and could only be conveyed through the use of discourse-level structure and broad real-world extra-linguistic knowledge knowledge, such as the following example about the *crisis of the third century* in Roman history:

1. ... ... at the middle of the third century the Roman Empire faced external threats from the Gallics, Palmyrenes, Vandals, Goths, and others. ... ...
2. the internal network of roads became less secure and the amount of interregional commerce decreased. ... ...
3. we saw an increased manorialism and the formation of an early form of serfdom ... ...
4. the institution of the tetrarchy by Diocletian brought more effective military command and administration to each region that are centered on Asia Minor, Dacia, Italy, and the Rhine. ... ...

Without the complete analysis of the contents in the intervening contexts, the locations of the geographical features, the identities of the warring factions and tribes, the economic model of the state, as well as some implications of technical descriptions such as manorialism and tetrarchy, it is impossible to see the causal chain through this series of descriptions. Knowing the locations and the identities of the external threats allows us to see the geographical alignment of those features. Having an understanding of the Roman economy, the structure and function of the transportation network allows us to see the connections with the external and internal conflicts as well as with concepts of manorialism and serfdom. Having an understanding of the makeup of the tetrarchy allows us to know how it relates to conflicts and administration of the economy. These and many other factors that are either extra-linguistic, or only elucidated in the complete analysis of a larger context.

Here we will focus on causal constructions; where semantically causal construction can be recognized largely through linguistic means, without the need of extensive analysis of broad contexts or a rich reservoir of knowledge of extra-linguistic facts and correlations. In this study, we focus on causalities with constituents in close proximity, within the same clause or in immediately vicinity.

a *John felled the evergreen tree*

b *John made the evergreen tree fall*

c *John had the evergreen tree felled*

d *John caused the evergreen tree to fall*

e *John caused the evergreen tree to become fallen*

f *John brought it about that the evergreen tree fell*

g *John picked up the axe, and felled the evergreen tree*

h *John drove into the hills, found an evergreen appropriate for Christmas, and felled the tree*

i *John asked Mary to purchase an axe so that Patrick may drive Rachel to the hills to fell an evergreen for Christmas*

j *John started the tradition of a holiday around winter solstice, so as to lead people to have a desire to decorate trees, in order that many trees in the forest would be felled that time of each year*

We observe that within those causal expressions in language contiguous in the same location in text, there are a variety of different lengths, lexical items, structure, and complexity among them. These range from a single predicate construction in (a); single predicate plus some TAM (tense/aspect/mode) structure in (b/c); some predication embedded in an explicitly causal clause (d/e/f); some coordinate (also adjoined) clauses which have some causal connection; to some deeply embedded series of clauses, each with its own predicate, as well as additional lexico-syntactic structures surrounding those core predications, all expressing some causal chain. We will explore most of the range of these types of causalities in language, during Section 6.1. And for the extraction phase of this project, we will focus on the more complex structures in this range, given that the simple structures are relatively straightforward to extract either through fixed patterns or some lexical resource, while the complex structures offer the significant challenges and more interesting outcomes.

1.3 Complex entity and network relations in language

This study on the complex causal constructions in language and their extraction from corpora provides a methodology for identifying a broad class of relations in a networks of entities and events in a network. Causality is a central type of relation that informs our understanding of sequential nature of sets of events in the real world, and is essential to understanding how individuals in a social network (SN) related to one another through actions and their consequences. While simple dyadically represented relations can connect individual entities in a network structure, complex relations such as causality inform us of relations among events (each of which would have entity participants), allowing for a far richer representation of the extracted SN.

The presence of complex relations allow us to seek graph-topology of a structural representation of an SN that goes beyond entity-entity relations (usually termed events), into realms where higher order relations can be efficiently represented and processed. The extracted complex and other relations could be represented with some linear representation from from target vertex using a scheme such as GLIDE (*graph linear description*) (Guigno, 2002; Guigno & Shasha, 2002; Shasha et al., 2002). Having the information about causalities in the linguistic data produced by the members of the SN allows for a far more complex topology of the SN to be constructed, especially when long chains of causation are taken into account. Thus, causality is the one relational type that would help us to understand long-distance and large-scale relational paradigms in a network structure, especially those that have a temporal aspect.

The development of methods, such as this study, in the elucidation of complex relations, that can be extracted from SN data, is an important step for cloud and mobile technology in the future to become individualized, contextually aware, and a high degree of automation in supplying relevant information and performing timely tasks on behalf of the end user. The new generation of cloud and knowledge-engineering based *"big data"* infrastructure allows future AI techniques to have much wider applications for end-users, in a way that will reduce users' information burden, and aid them in performing intellectual and daily tasks more efficiently and more temporally relevant. We will discuss a few of these potential applications in the real-world, in greater detail, in Section 10.4. Through this study, we hope that not only the techniques here would lead to better information extracted on causal constructions, but the extraction of other complex constructions that are highly relevant to an SN structure, such as *cooperation*, *explanation*, and *elaboration* would also be able to be improved by integrating these approaches.

Chapter 2

Statement of Purpose

The main purpose of this research project is to investigate fundamental but novel techniques that can be utilized to detect, rank, and extract complex event relations, such as causality, in multi-genre corpora; this is done with the ultimate aim of contributing to more knowledge rich and contextually aware cloud applications that can learn from, and serve user interactions on-line. Here, we briefly introduce a series of primary research questions to answer, and some ancillary issues to be touched on through the course of the research.

2.1 Primary research question

The primary research question in this study is *the extraction of complex causal structures from linguistic data.* This precludes the simple types of causality that can be extracted using a finite collection of patterns, and also excludes causal structures that rely on long distance discourse context to extract, or those that purely rely on pragmatics to determine its causality. This process is broken down into several discreet steps and components, which together make up a practical approach to identify semantically broad classes of causal expressions in language.

➊ We need to linguistically analyze the types of lexico-morphosyntactic structures in language that may convey causality. Those that can be identified with relatively simple lexical property or some fixed morphosyntactic patterns should be relatively straightforward to pursue, and will not require fundamentally new techniques. On the other hand, the prevalent types that are inherently complex with regard to the variability in its morphosyntax, or is contextual in nature, need to have novel methods developed in order to further pursue, these we have identified as *adjunctive causality* and *embedded causality*.

➋ A sensible representation of the constituent blocks of complex causal structures at the morphosyntax-semantics interface that can be extracted with reasonable fidelity would be necessary to provide the input for the causality identification process. This study will use a practical instantiation of *semantic frames* to represent such structures, with all of the requisite parameters for further processing, and a automata based mechanism is used for their extraction. We will also employ a representation for linguistic structure that facilitates learning through genetic algorithms, a hybrid cognitive categorization model, the *diffuse prototype*.

➌ The *adjunctive causal* constructions is one of the principal classes of complex features that express causality, this accounts for the type of adjacent frame pairs in the text that express $e_i \xrightarrow{caus} e_{i+1}$. We strive to utilize primarily the information outside/around the core frame components, basically the complement to the essential frame structure, to detect sequential structures within that convey causality. This is done through *hidden markov model* (HMM), trained through a highly-modified form of the Baum-Welch algorithm (Baum & Petrie, 1966; Welch, 2003).

➍ The *embedded causal* constructions is the other principle class of complex causal feature from our linguistic analysis, and accounts of the type of two or more frames that are form a deeply embedded structure in syntax, and tend to inform of speaker of longer chains of causalities in a semi-explicit manner. These are discovered through a new model in cognitive categorization in *diffuse prototype*, and devising a adapted and extended form of *genetic algorithm* over the set of characteristic sub-structures.

2.2 Ancillary research questions

This direction of the study falls simultaneously under several broad domains research, given its highly interdisciplinary nature. It utilized a lot of computational techniques in formal models such as automata theory and graph theory, as well as probability theory, and employs a broad range of formal and learning algorithms from computational sciences. It also requires deep linguistic analysis, especially in the areas around the morphosyntax-semantics interface, as well as a good deal of formal logic of language, in order to correctly formulate each module and its approach. The analysis and overarching goals of this study deeply ties into web-technologies and analysis of on-line social networks, thus take its input from, and has implications for psycholinguistics, sociology, as well as study of the web. This study also touches various other areas such as economics, evolutionary biology, and other areas, to be outlined next.

One set of ancillary questions if similar methodology that we have developed here would work for other complex relational features in language, specifically language that is produced in or on the topic of a community of entities. We already have a good procedure for extracting *reciprocal* relations from text (in Chapter 4), which is a type of relation that has a large intersection with causality itself. Another important type of relations that might be explored would be *cooperative*, which like reciprocity and causality, have strong implications for an representation of a relational network among entities. Causal relations form the primary serial sub-structures in such a relational representation, as causality often form chains (Section 6.2.3) that allows some entity to influence another to behave in a certain way, where the second entity may influence yet other entity down the chain. Cooperative relations, in an analogous way, form the primary parallel sub-structures within this representation, as in cooperative expressions, multiple entities often form a collective to perform the same action on an object, or individually perform the same type of parallel actions on a class of objects. So cooperatives is an essential class of relations to examine in order to form a complete representation of some social network.

We will also attempt to perform additional linguistic analysis on the discovered complex causal types, observing their semantic properties with respect to formal logic and their social properties. This is done for the purpose of eventually integrating the network-representation of these causal forms into social network frameworks and applications in the future. The types of causal structures we select are complex in their linguistic forms, which are substantially different from those sought by most traditional machine learning methods, but they also readily lend themselves to integration into the structure of a social network representation. As we will see in later sections, the adjunctive causal structures generally contribute to parallel structures in the network representation, and the embedded causal structures generally contribute to long causal chains. Together with the simply forms of dyadic frame causal structures (including subsets of forms like *causative inchoatives* in 6.1.2, or *analytic constructions* in 6.1.4), these will form an important set of relations among events contributing to the overall graph theoretic structure.

2.3 System Outline

The overall system has a simple design, with three major components that effect the extraction of complex causal expressions, we will briefly describe the overall design, and show the details of each major component in their respective sections. The raw text from each corpus is cleaned up, then tagged and parsed into preliminary context free form. The first major component, the frame extraction mechanism, takes that pre-processed input, and transforms the sets of clauses into some linear (between clauses) and hierarchical (multiple frames within a composite clause) ordering of frames, based on their estimated structures at LF (described in Section 5). The frames produced are then taken as input for the two major components that discover the adjunctive and embedded causal structures. The adjunctive causal structure performs the task in an unsupervised manner, looking at the linearly adjacent frames for signs of causality, relying on Bayesian principles, and produces a set of sequences each corresponding to some $\langle e_i, e_{i+1} \langle$ (this is a simplification, more caveats and more precise definitions will be discussed in Chapters 4, 7, and 8) that are likely to be adjunctively causal structures (in Chapter 7). The embedded causal structure performs its task in a semi-supervised manner, where a limited set of sample is pre-filtered for the types of complexity that we desire, and then labeled positive when an annotator views some matrix frame has some causal relation with one of its embedded frames. Then the training samples are run through an evolutionary process to obtain some *diffuse prototype* of the embedded causal structures within the corpus (in Chapter 8, which is then used to score other potential sets of matrix-embedded frame sets, to see which are the most likely embedded causal structures.

2.4 Summary of contributions

For this project, we were able to achieve the following: ❶ We were able to provide a procedure for locating reciprocal pairs of eventualities that are in the same text context, but not necessarily mutually adjacent, from on-line data sources. We were able to achieve a precision of 60% ❷ We were able to provide a preprocessing procedure that uses tree transformation mechanisms informed by frame structures of clauses, and provide reasonable performance (approximately 90%) as a prerequisite for deeper semantic processing for the causality modules. ❸ We were able to devise a procedure that ranks the adjacent pairs of clauses, based on a form of Hidden Markov Model (HMM), with significant modifications specific to the problem, that express a causal relation between two events, with or without any explicit cues in terms of *discourse connectives*. The top two quantiles had precisions of 85.5% and 73.0% for a BNC test data-set; and 85.4% and 79.8% for the top two quantiles from a novels test data-set. ❹ We were also able to devise a procedure that ranks complex embedded lexico-syntactic structures, with a model of representation appropriate for this type of causal structure, using a graph theoretic evolutionary computation model designed specifically for this problem. The top two quantiles had precisions of 79.6% and 67.7% for the BNC test data-set; and 80.0% and 57.4% for the novels test data-set.

Chapter 3

Background and Previous Work

This study is motivated by a need for methodologies that could discover types of causalities in linguistic data that have been difficult in previous work in computational semantics, as well as find new approaches for finding relations among events that would discriminate among structures with greater complexity. Some background and previous approaches in this specific area of extracting semantically causal relations from linguistic corpora are introduced. The starting point of this study was on a similar, but less diverse and complex set of relations to *causality*, our previous work on the extraction of *reciprocity*, which in part informed some of the approaches and strategies in our main body of work in extracting causal structures.

3.1 Traditional approaches in extracting causality

Here, we will briefly examine the traditional types of computational and theoretical approaches in analyzing and extracting causal relations among events, while discussing many specific elements of these in later sections, where they are relevant. Most of approaches can generally fit into one of several broad categories. Using probability theory is prevalent in looking at causal relation between some pair of events $\langle e_i, e_j \rangle$, based on the occurrences and distributions of individual event types within a corpus. Some purely parametric approaches are also widely pursued, which focuses on parameterizing a large set of features that could have some joint occurrence property with causal relations, in a way that is computationally feasible, and arrive at some linearly separable classification of the dataset. There are also formal models, mostly unimplemented computationally, that have elaborate logical structures that are used to discriminate causals from other types of relations. We will summarize each of these classes concisely, and the types of causal structures that they are likely to efficiently detect, in order to provide a contrast to methods in this study. Certain individual elements from these theories would be useful in the current study and are adopted in certain modules of our system; we will leave the discussions of these details to the later sections where each becomes specifically relevant.

3.1.1 Parametric approaches

There are a class of approaches that rely on machine learning methods over high-dimensional semantic-feature spaces. A large number of examples of this approaches exist, using a variety of machine learning methods (Abe et. al., 2008; Berthard & Martin, 2008; Riaz & Girju; Do et. al., 2011; Radinsky et. al. 2012 / 2013; Oh et. al. 2013; Hashimoto et. al., 2014; etc). A number of features relevant to causality are first identified. An extraction procedure for these features follows, usually with some type of linear sequence pattern matching procedure, each candidate form with its extracted features. The the presence and absence of these features are encoded into a feature vector for each candidate sample, and some form of multi-dimensional learning technique, often SVM when there are numerous features under consideration, is employed to find linearly or polynomial separable clusters of these feature vectors. The implementation may require some labeled positive and sometimes negative samples, if supervised; and the binary classification through ML now is able to discriminate candidate forms that are causal from non-causal.

These identify a number of distinct features in a corpus that have some ability to discriminate between causal and non-causal occurrences, which are generally engineered with some knowledge of linguistic theory. The collection of features often include classes of lexical items (i.e. relational adverbials such as *'therefore'*, *'since'*, and psycholinguistic verbs of causation, such as *'persuade'*, *'lead'*), patterns of lexico-syntactic fragments (i.e. VP containing a PP headed by a relator preposition), associations of verbal time/aspect sequences (i.e. past progressive + present perfect), temporal lexical cues (i.e. some sequence of temporal indicators: *'it was then ..., and now ...'* *'at first ..., thereafter ....'*), and many other types. With each candidate represented by a vector of feature-values as a data-point in a multi-dimensional space, where linear separability in a binary classification scheme is sought. In a generalized form of fixed polynomial, the separation of the data-points can be explained as the following:

$$\mathcal{D} = \sum_{\phi_i \in \Phi} w_i \cdot \mathcal{P}\Big(\mathcal{S}_i(x_i, \phi_i),\ d\Big) \tag{3.1}$$

Where the set Φ is some predefined set of useful features in each sample, w_i is a weight given to the relevance of ϕ_i, $\mathcal{S}_i$ is some extraction procedure associated with each feature, and $\mathcal{P}$ is a polynomial kernel with a degree of d.

Purely parametric approaches are highly adaptable, can be reformulated to work with almost any data-set, and provides a lot of flexibility in terms of the degree of fitting to a specific genre or domain of linguistic data. However, such approaches often require large labeled data-sets for a supervised approach, when the types of discriminate task is complex or is with regard to deep semantic features. A purely parametric approach also requires the experimenter to engineer specific feature sets for a specific task, which itself is not trivial, and can take up the vast majority of the time for developing the procedure, when the task or the data-set is novel; such approaches are a simulation of the scientific method, and in of itself does not contain knowledge about the current problem. Moreover, these approaches also require that the data be linearly separable in for some set of engineered features, which means that for structurally highly complex objects such as trees or graphs, the numbers of degrees of freedom would be very high for such objects. Hence these methods are not designed to be scalable, and cannot be efficiently used to develop treatment of data-object that have highly complex features, such as the types of causalities that we will encounter.

3.1.2 Knowledge base approaches

There is a class of knowledge rich methods for linguistic causality that focus on building a resource of causal pairs of event types that are used as lemmas in large computational systems, like $\langle \tau_1, \tau_2 \rangle$, where each type can be represented by a *predicate*, *predicate gerunds*, or certain types of eventive nominals (i.e. *'the rain'*). There are a number of such approaches that have been theoretically examined and pursued (Kiryakov, 2004; Hobbes, 2005; le Priol et. al., 2007; Berthar et. al., 2008; Miahila et. al., 2013; etc.). Primarily manual, or a mixture of automatic and manual methods may be used to build a data-base of these many of which are focused on some domain specific semantic relation, since that greatly eases any task in annotation. With the manual portions of these procedures, which is central to building a knowledge base, where multiple annotators are used, and some metric for inter-annotator agreement (such as Kohen's κ) is used to measure the trust in using the annotated corpus as a standard.

This class of methods almost always entails some type of annotation of a corpus, following an annotation scheme specifically designed to indicate the locations of the cues that trigger causal responses in the annotators. The annotation scheme could also contain additional classification information for sub-classes of causality, which may include *necessity*, *entailment*, *enablement*, etc, each provided with a distinct label, and each with a different underlying logical form mediating the relation between the types (some of such sub-classification are analyzed after our own extraction procedures, and presented in Chapter 9). These annotated cues could then be used to discover causal constructions in unseen data through some pattern matching algorithm, or integrated into some ML procedure down-stream.

The more human effort is pooled into developing resources in the data-base, the better this class of methods performs. But this also means that performance is highly correlated with the amount of human-intervention required, so these methods cannot be relied on to be automated and adaptable over long time-scales, since once the resources have been constructed, there is no longer the adaptability needed to change with new data, unless it is coupled with other methods. The amount of human-resource needed to develop the knowledge-base for each type and genre of linguistic input is also a critical constraint for the applicability of this class of methods for specific problems, as resource developed for one genre and a specific problem often cannot be applied to another.

3.1.3 Probabilistic causality theory

Modeling causal relations as probabilistic phenomena has always been a viable theoretical and practical route, which has a number of similar theoretical approaches that may be parts of implementations of causal discovery (Salmon, 1980; Pearl, 1999; Tian & Pearl, 2000; Spirtes et. al., 2001; Williamson, 2009; etc). We know the concept of $e_i \xrightarrow{caus} e_j$ in the real world is normally associated with some uncertainty, and $P(e_j|e_i) > P(e_j|\bar{e_i})$ is seen as a legitimate definition for causality in that light. Several basic concepts have always been important in all forms of probabilistic causality, with reference to some time-slice t, including joint probability $P(e_j^t \wedge e_i^t)$, conditional probability $P(e_j^t|e_i^t)$, and bigram probability of $P(e_j^t|e_i^{t-c})$, among others. Previously (Beamer & Girju, 2009) causal probability utilized joint and conditional probabilities between e_1, e_2:

$$\mathcal{C}(e_1, e_2) = log(\frac{P(e_2|e_1)}{P(e_2)}) + log(\frac{P(e_1 \xrightarrow{bigram} e_2)}{P(e_2 \xrightarrow{bigram} e_1)}) \tag{3.2}$$

Conditional probability and joint probability are both readily available, after finding the appropriate unit of representation for the predicate. A more principled way for computing causal probability (Pearl, 1999 / 2000; Tian & Pearl, 2000), which takes the exogeneity and monotonicity of the causal events into account, involves the addition of counterfactual probability as part of a reasonable metric. Where as Beamer & Girju defines *event* as only the main predicate, we use an alternate definition of *event* that uses multiple components of the predication structure, which is more precise for the purposes of causality. Most reasonable metrics contain the probabilist definitions of necessity P^N, and probability of sufficient P^S as factors in its terms; where e_0 is the event of causal link between e_1 and e_2 in a three event model of causality:

$$\begin{cases} P^N(e_0)\hat{=}P(\bar{e_2}_{\bar{e_1}}|e_1,e_2) \\ P^S(e_0)\hat{=}P(e_{2_{e_1}}|\bar{e_1},\bar{e_2}) \end{cases} \tag{3.3}$$

Both will require some form of computation of the counterfactual $P(e_{j_{e_i}})$, which is out impetus here. We will explore counterfactual probability and its importance in a later section (7.2.2)

Causal probability is a very useful set of concepts, and will be used in the development of some methodologies of this study, and especially participate in the extraction of *adjunctive causal* structures in Chapter 7. But the central issue with solely relying on causal probability theory is that it views events as atomic entities, or at best a loose collection of components (predicates, arguments, obliques), and lacks strong internal structures. This will prove to be highly problematic for locating certain types of causal structures (i.e. long causal chains), where different sub-structures within some linguistic unit representing the event are relevant for expressing causality. We will employ sub-atomic view of events, and even more flexible event type representations in our solutions.

3.1.4 Formal approaches

There are a variety of different formal theories on the issue of causality itself and causality that may be expressed in language. We will explore the detailed incarnations of a variety of these forms in Chapter 6, while here we will briefly outline the general approaches with some formal logical system as its underpinning (Russell, 1948; Burks, 1951; Simon 1952; Cartwright, 1979; Karimi, 2010; Schimbera & Schimbera, 2010). The basic formulation of $e_1 \xrightarrow{cause} e_2$ also entails an irreversible direction, unlike a directionality of *implication*, which may be formulated as a disjunction of $\neg p \vee q$. There is no commonly accepted orthodoxy to what sufficiently constitute causality, but we can speak of a most strict definition, the circumscribed set of relations by which would be regarded by nearly everyone as being causal, which is the following form:

$$\begin{cases} c_i \xrightarrow{caus}_U e \quad \iff \\ \\ \prod_{p_k \in \mathbb{P}(\mathcal{C}_{[-c_i]})} \mathbb{1}\Big(c_i \longleftrightarrow_U e \mid p_k\Big) \end{cases} \tag{3.4}$$

What this is saying is that when we have only a closed system U under consideration, let $\mathcal{C}$ be a set of binary features of all possible causes of events within the system, and some specific cause $c_i \in \mathcal{C}$ under consideration; for any permutation $\mathbb{P}(\mathcal{C}_{[-c_i]})$ of the vector $\mathcal{C}_{[-c_i]}$ (the set of all potentially causal features except c_i), the target causal candidate c_i if and only if e_j, then c_i causes e_j. The type of circumstances described by such a formula would defeat any objection of $\langle c_i, e_j \rangle$ of merely being in an association, and any objection in terms of the logical distance in a causal chain. The only other potential requirement would be to stipulate that $i \prec j$, which presumes a temporal order in U. Every other definition of causality is some weakening of this extraordinarily strict definition; and the above formulation would rule out much of what most people consider to be causal relations. Some may weaken the requirement of all of the other potential causes in $\mathcal{C}$, some may weaken this formulation by describing a system other than U where the implications are evaluates, some may weaken this by using some other operator than $\mathbb{P}(\cdot)$, etc; all of which allow us to extend this formal definition to include less strict forms of causality.

As we can see, a formal approach offers a good method for dealing with complex structures and logic in causation. But the lack of a consistent theory, and the lack of easy implementation for many of its theories presents a significant obstacle to employing these concepts in actual extraction or discrimination algorithms. We will use some specific pieces of ideas from these formal methods, but these cannot be relied on providing the overall framework or the bulk of the specific procedures for our tasks. Formal approaches can offer many insights into specific issues that we would encounter with causal structures of high complexity, such as those in Chapters 7 and 8 . Having a good knowledge of the logical framework behind causality allows us to augment and adapt other types of methods to arrive at solutions for extracting complex linguistic causalities, but formal logic by itself does not offer a viable application in computational semantics of causality.

3.2 Previous work on complex causal structures

Closer to the specific problem in this study, there is the same variety of approaches to the ones described above, in finding causal relations in the literature, approaches which rely mostly on machine learning methods over high-dimensional semantic-feature spaces (Abe et. al., 2008; Berthard & Martin, 2008; Riaz & Girju; Do et. al., 2011; Radinsky et. al. 2012 / 2013; Oh et. al. 2013; Hashimoto et. al., 2014; etc). Other researchers have focused on pre-identified lexico-syntactic patterns (Khoo et. al. 2001; Girju 2003) which they use to bootstrap an Expectation-Maximization procedure (Chang & Choi 2006; Paul et. al. 2009) for causality and similar semantic relations. Furthermore, these parametric and pattern recognition works are generally focused on pair-wise causal relations between event representations. For our own study, we instead focus on linguistic structures of unbounded complexity that are capable of expressing sequences of events involved in adjacent pairs and longer causal chains. Our work explores novel representations of causality, and procedures rooted in HMM and evolutionary computing in order to deal with the structural complexity of these expressions as well as retain the flexibility of parametric approaches.

Chapter 4

Reciprocal Relations

We will start with locating another closely related type of semantic relation in language. The linguistic reciprocities are a subset of 2-entity mutual interactions, specifically where their real-world occurrences have some logical link and sequence, most of the time requiring either that one event be the consequence of another, or both being the consequence of a simultaneous interaction. Common example can as follows:

1. ***Jack*** *and* ***Diane*** *collided with* ***each other*** *in the bumper-car race course*
2. ***Mary*** *back-stabbed* ***John*** *and* ***he*** *would like his vengeance on* ***her***

2-entity mutual interactions again are the most frequently exhibited subset of all complex interaction networks in the in the overall on-line social network.

4.1 Linguistic reciprocity

The set of linguistic expressions considered reciprocal have potential intersections with expressions of other complex semantic relations involving multiple entities and events, such as *causality* and *cooperativity. Reciprocity* itself is not necessarily causal, but it has a large intersection with causal relations; whenever a reciprocal relation expressed in language has some temporal ordering between e_1, e_2 such that there is a perceptual gap in time between the two in time, then $e_1 \xrightarrow{caus} e_2$ for this reciprocal pair. Thus some of the elements used in the reciprocity extraction procedure may be useful in understanding causal constructions and formulating a procedure for extracting causalities. For example, the following is a reciprocal relation that we identified from our dataset:

$Entity_x$ *confronted* $Entity_y$*,* $Entity_y$ *was interested in* $Entity_x$

Here, the two eventualities were identified by an annotator as causally linked, specifically $Entity_y$ being interested in $Entity_x$ described in the latter eventuality probably led to the confrontation recounted in the former. Other linguistic features relevant to the network structure will have their own appropriate graph representations. In this manner, we can find corresponding types of directed graph regions for many similar linguistic features at the syntax-semantic interface. Certain modifications are made to the base directed graph model to accommodate the nature of the linguistic data.

Reciprocity is a relation of mutual dependence, action or influence (cf. WordNet (Fellbaum 1998)) between two or more parties. In general terms, reciprocity refers to the response to an action with another action. Reciprocity is a well known concept that functions in multiple domains of knowledge, and can be interpreted through linguistic and extra-linguistic mechanisms. Logically, it contains a significant intersection with causal relations and explanatory relations, but is neither a proper subset or superset of either. Typical examples would be:

1. *The earth* orbits *the sun* since *the sun* gravitationally attracts *it*
2. *Mary* gives *John* a present, and *he* thanked *her* for that
3. *The thermohaline cycle* moderates *the climate* in the North Atlantic, *which* in turn perpetuates *this cycle* by increasing the surface runoff into the North Atlantic basin

As we can see, it normally (in its binary form within traditional linguistics) entails exactly two distinct entities, and exactly two events, each of which predicates over both entities. In this form, reciprocity is the most linguistically regular and the least complex in its graph theoretic representation, of the four complex linguistic features that we will utilize in this study; although more complex, and less traditionally linguistic forms will be entertained later in the study in Chapters 7 and 8.

Reciprocal expressions can refer to a an important subset of social and cultural norms which govern behavior in human society. It sometimes refers to the exchange of one economic good for another in the context of trade and commerce. In social psychology, it is an important component of contemporary exchange theory (Molm, 2010), and contributes to the control of the distribution of power and flow of benefits (Cook & Emerson, 1978). In biological sciences, it refers to a set of cooperative or symbiotic rules among organisms in the context of evolutionary dynamics. It may also refer to the mutual benefits or threats among political entities in the field of international relations. And a number of other interpretations in disparate domains of knowledge also exist. In each one of these areas, the concept of *reciprocity* denotes the causal potential of one set of actions for another as performed by distinct individuals. In each case, reciprocity itself is also an indispensable component in the wider set events that comprise all interactions in a population of entities.

4.1.1 Linguistic representation of a social phenomenon

Linguistic reciprocity is possibly the most direct and precise manifestation of reciprocal relationships in human produced data. It can manifest in various surface forms within a single languages, (Maldonado, 2011) and can denote various different types of mutual relations, such as *simultaneous*, *competitive*, *collaborative*, etc. (Nakao, 2002) A study of reciprocity in natural language provides the means for a deep analysis of social interactions. Linguistically, this mutual dependence of two entities is represented by relations on pairs of eventualities, frequently each eventuality contains an *agent* role and a *patient/goal/recipient* role, and two entities X and Y reverse their role from one eventuality to the other. Here is a simple example illustrating this pairing:

$$\textit{Entity}_x \textit{ loves Entity}_y \textit{ and Entity}_y \textit{ loves Entity}_x$$

In this case, the reciprocal pair of eventualities are clearly represented in their surface forms, and headed by the main verb *love*, and the two entities alternately play the agent and recipient roles; and these representations of entities we will term ***reciprocity template***, which we will detail in a concrete form in section 4.2.1. Here the eventualities are explicitly represented in the syntactic forms. We will term this surface form in the dataset the *canonical form* of reciprocity, represented here:

$$Entity_x\ Verb_1\ Entity_y\ CONJ\ Entity_y\ Verb_2\ Entity_y$$

The further the string edit distance from the surface representation of the reciprocal relation to the canonical form with the same semantic content, the more difficult the form would be to detect automatically.

4.1.2 Variations in linguistic representation

Some examples are syntactic transformations of some canonical form of reciprocity. The following example is semantically identical to the previous example, except with passivization and a coordinate VP structure:

$$\textit{Entity}_x \textit{ loves Entity}_y \textit{ and is beloved by Entity}_y$$

Semantic reciprocities in general, however, can take many forms on an *canonical* $\longleftrightarrow$ *latent* continuum. In many cases, the reciprocal relationship between the two entities, and sometimes the representations of the entities themselves becomes more abstract, and decoupled from the surface forms occurring in text. The following examples are examples of such, each being progressively more distant from the canonical form.

1. *Entity*$_x$ *thanked Entity*$_y$ *for Entity*$_y$ *completing the assignment with integrity.*
2. *Entity*$_x$ *regards Entity*$_y$ *as a benefactor in the current situation.*
3. *Entity*$_x$ *is hated for his reckless behavior*

In the first example, the act of *Entity*$_y$'s performing the task contains the semantic role of **theme**, which is the oblique object *'task'*. The pragmatics of the sentence dictates that the individual *Entity*$_x$ is in some way a beneficiary or otherwise related to the task performed by *Entity*$_y$. In the second example, the entire second eventuality is expressed in the form of an NP *'a benefactor in the current situation'* . This is semantically related to the surface form of *'X benefits Y'*, which takes an **agent** role, as well as **benefactive** role, which are filled by $Entity_y$ and $Entity_x$ respectively at a semantic level. In the third example, as in the second, the second eventuality is expressed as an NP oblique object *'his reckless behavior'*; but the agent of the main verb *'hate'* is also missing, due to the surface form of the first eventuality being passivized. The identity of the individual, or more likely in this case the group of individuals, playing the **agent** role in the first eventuality and the **experiencer** role in the second eventuality is not present in the local sentence, and must be recovered from the global discourse context. The difficulty of recognizing the representation of reciprocity in text rises dramatically as we drift away from the canonical form. In this study, we will primarily target those surface forms where all four occurrences of the two entities are present; and this would allow us to achieve a reasonable precision in our task. We will leave the task of targeting the more semantically opaque forms for a future study.

4.1.3 Extension into discourse context

As in the original example with four occurrences of two entities in chiastic pairing, these entities can be represented as pro-forms (pro-nouns, pro-NPs), NPs, or named entities. We can accurately detect a candidate surface form of a pro-form based reciprocity template. And such pro-forms can be exhaustively enumerated given a language. The use of pronoun templates also obviate the need for co-reference resolution, which would be necessary in mixed (pro-form and named entity) surface form candidates such as the following:

$$Entity_x \text{ loves } Entity_y \text{ and } proform_1\{ref{:}Entity_y\} \text{ loves } proform_2\{ref{:}Entity_x\}$$

But the exclusive use of pro-forms limits the distance between the surface forms of the two eventualities, as each pro-form can have a number of different referents in the discourse context. Thus templates consisting of pro-forms can detect reciprocal pairs of eventualities if they are adjacent in the text. In order to identify such pairs separated by long distance, it would require very specific template components, ideally some uniquely named entity e.g. *Sam Waterstein*, or entities bearing an identified relation to a named entity, such as *the sister of Sam Waterstein*, or *the roommate of the sister of Sam Waterstein.* This level of specificity, in turn, would require accurate named entity recognition, and consistent co-reference resolution that identify the correct pro-forms which semantically reference the named entities.

This process is in essence an abstraction away from the linear order of the content in the corpus and toward a graph based representation of the entities. This is a situation where the simple components such as **vertices** (entities) and **edges** (actions) can correspond to local context-free structures *NPs containing named entity* and *VPs with a valence of at least 2*; and where the higher order structural features in the graph, such as **multi-edge directed path**, **k-cycle**, **strongly connected component**, **maximally connected subgraph**, or **complete subgraph** as candidate regions that can identify with context-sensitive structures in the linguistic content such as a *reciprocal pair* or *reciprocal n-tuple.*

4.1.4 Extension beyond direct reciprocity

The current theories on reciprocity discussed in semantics or pragmatics (Dotlačil & Nilsen 2008; Murry 2007 / 2008; Slavcheva 2007) are not the only type of reciprocal behavior that is relevant to the group dynamics of a community that produces a linguistic dataset. While the traditional definition focuses strictly on the interaction between a pair of individuals, we need to take a broader view in order to account for complex behavior of a large group, where interactions among larger sub-groups of individuals may have similar function and effect as pair reciprocity. The individuals' influence future interactions among individuals in the same community bring some form of payoff for the original action. It is necessary to examine *indirect reciprocity*, otherwise known as *economic reciprocity* among a group of more than 2 to study these more complex, and yet reciprocity-like behavior. This phenomenon is discussed in greater detail in our technical report (Li & Girju, 2010), and will be address in a future part of our study on social networks.

4.2 Locating reciprocity candidates

The targets of the reciprocity methodology are the pairs of eventualities where the occurrence one action by an entity X can potentially be correlated with the occurrence of the corresponding action by entity Y. And this most frequently involves two individuals, each being the agent of one action, and the other being the recipient/patient/benefactive of the reciprocal action. The baseline reciprocal method first taking advantage of this fact and seeks such occurrences in adjacent pairs. And then the pairs are merged into patterns to identify the final set of reciprocity constructions.

4.2.1 Formation of templates and patterns

We refer to a linguistic construction discovered by our procedure as "pattern" (a pattern type) and to an occurrence of a pattern in the corpus as a "pattern instance". The simplest and most reliable observations are the set containing only pairs of reciprocal expressions within a single compound sentence, which has low recall but high precision. For this highly reliable case, a single template of four components used to locate reciprocity candidates.

Basic pro-form reciprocity patterns In a moderately sized data-set, the patterns that are most likely to occur in sufficient frequency, (in order to be repeatedly observed across a number of instances of reciprocities,) are the *pronoun-templates.* These are composed members within the set of pronouns $\underline{\mathfrak{P}}$ in the languages as the elementary building blocks. In the case of pronoun-templates, (which are the most abundantly observable single-sentence templates,) the observable sequence is of the form $\bullet P1 \bullet P2 \bullet P3 \bullet P4 \bullet$, where P1 and P4 have the same number and person if pro-forms, and being identical or within the same class if entities; and P2 and P3 having a similar relationship. (The • represents any intervening material between any two entities, or between an entity and one of the clausal boundaries.) Since pronouns' syntactic cases are readily discernible, the reciprocity pairs observed through the use of pronoun-templates must adhere to noun case constraint that specifies *Nominative-Accusative-Nominative-Accusative* sequence. Third-person forms also adhere to gender constraint.

"[Part1] **I** [Part2] **him** [Part3] **he** [Part4] **me** [Part5]" and

"[Part1] **they** [Part2] **us** [Part3] **we** [Part4] **them** [Part5]",

4.2.2 Named entity based templates in text

The other naturally abundant building blocks in these corpora are the recognized named entities, the mechanism of which is specifically designed for on-line data of this genre of web-forum having discussion topics focused on human relationship. The process of their extraction uses a *linear bounded automaton* based process, that locations all of the expressions that can identify an entity based on some expressed chain relationships ultimately with some user of the forum, such as *'the co-worker of the sister of his best friend'*.

The entities that correspond to forum users are only a subset of the identifiable entities in the data, other entities bearing familial and other close relationships with the forum users can also be accurately identified. First, the NP candidates containing a user-entity or a pro-form identified with a user (some of these are newly inserted during co-reference resolution), such as *his {ref: USER_GRKSCORP} brother* are identified, and these chunks of texts are shallow parsed to reveal a simple structure of its components, which in this simple case would be (DP(DT: his)(NN: brother)). We can define the set of entities inductively, with $\mathbb{U}$ as the identifiable types among forum users, $\mathbb{R}$ as the set of unique relationships, and $\mathbb{S}$ as the set of non-unique relationships. Here is the set of recognizable entities $\mathcal{N}$ inductively defined.

$$\mathcal{N} := \begin{cases} \varepsilon_i \in \mathcal{N} \mid \varepsilon_i \in \mathbb{U} \\ \varepsilon_i \in \mathcal{N} \mid \varepsilon_i = \varrho_k(\varepsilon_j), \\ \qquad \varepsilon_j \in \mathcal{N}, \quad \varrho_k \in \mathbb{R} \\ \varepsilon_i \in \mathcal{N} \mid \varepsilon_i \in \varsigma_l(\varepsilon_j), \\ \qquad \varepsilon_j \in \mathcal{N}, \quad \varsigma_l \in \mathbb{S} \end{cases} \tag{4.1}$$

The occurrences fit into one of number of predictable patterns. Each pattern is consistent with only one type of production from the above inductive definition of relations, the most frequent one being *[Entity Possessive-Morpheme [Modifier]* Noun]*, e.g.*USER_GRKSCORP 's brother* (originally his {ref: USER_GRKSCORP} brother), or *USER_PRIMO 's baseball boyfriend* (originally *her {ref: USER_PRIMO} baseball boyfriend*). These forms correspond to a simple state machine; in the example, our desired output would also be a regular expression in the form of *[Determiner [Modifier]* Noun 'of' Entity]* (with the '*' representing the transitive closure of the content within the brackets). Other slightly more complex regular expressions exist, such as *[Entity*$_1$ *'and' Entity*$_2$ *Poss-Morpheme [Modifier]* Noun]*, with the corresponding form: *[Determiner [Modifier]* Noun 'of' Entity*$_1$ *'and' Entity*$_2$*]*; or the even more complex form: *[Determiner Noun*$_2$ *Rel-Pronoun VB Entity Poss-Morpheme [Modifier]* Noun]*, with the corresponding output as *[Determiner Noun*$_2$ *Rel-Pronoun VB Noun*$_1$ *'of' Entity]*, and others. As expected, the simplest form predominates.

There is also a co-reference resolution module that was designed to specifically deal with the co-reference chains that exist in web-forums of this type of format, utilizing the structures of posts and threads to its advantage. Both the NER and the co-reference modules provide the necessary entity information for the reciprocal procedure, but will not be discussed in detail due to space. The extraction of entity information for the later causal module will be part of the frame-structure extraction module instead, described in Chapter 5

The named entities extracted is composed of two subsets. One set is the representations of forum users in the set $\underline{\mathfrak{U}}$, formed from the base case of the definition in Formula 4.1; and other is the representations of related entities in the set $\underline{\mathfrak{R}}$ formed from inductive cases of the definition in Formula 4.1. These patterns are much rarer for any single type, but includes a large number of distinct types. This property of $\mathfrak{E}$ and $\mathfrak{R}$ also implies that template constructed from these large number of rare and distinct types would provide very high specificity, where a single template type made up of named entities is not likely to occur more than once in the entire corpus.

Given this advantage of using named entities in template formation, these can also be used in identifying reciprocal patterns in a more flexible way. Since the occurrence of some representation of any one distinct named entity is very rare in the corpus, it is likely that two occurrence of the same entity representation are referring to the exact same individual. So when two actions of $Entity_1 - event_1 - > Entity_2$ and $Entity_2 - event_2 - > Entity_1$ in two separate locations in the corpus, this can also represent a potentially reciprocal pattern, without the two actions canonically occurring adjacent in the text. The intervening portion between these two events, analogous to the [Part3] of the previous patterns, is much less accurate than if we simply concatenated the two events. But the balance of the pattern is reasonably likely to be homologous to adjacent pairs of reciprocals. We refer to these as *loosely adjacent reciprocal patterns*, which are less accurate, but can potentially greatly increase the number of reciprocity candidates to be found in the corpus, and supplement the system's knowledge of blind spots within pattern discovery process. In the following, example 3 in each set is the fusion the two preceding.

1. *"[Part1]* **USER_X** *[Part2]* **the_mother_of_USER_X**
2. "**the_mother_of_USER_X** *[Part4]* **USER_X** *[Part5]" and*
3. *"[Part1]* **USER_X** *[Part2]* **the_mother_of_USER_X** *[Part3]* **the_mother_of_USER_X** *[Part4]* **USER_X** *[Part5]" and*

1. *"[Part1]* **the_brother_of_USER_Y** *[Part2]* **the_niece_of_USER_Y**
2. "**the_niece_of_USER_Y** *[Part4]* **the_brother_of_USER_Y** *[Part5]"*
3. *"[Part1]* **the_brother_of_USER_Y** *[Part2]* **the_niece_of_USER_Y** *[Part3]* **the_niece_of_USER_Y** *[Part4]* **the_brother_of_USER_Y** *[Part5]"*

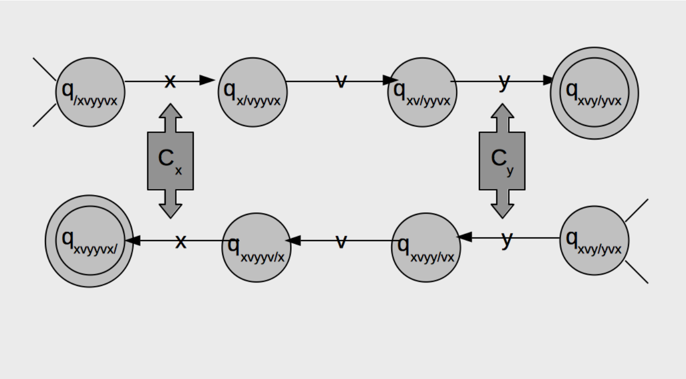
In addition, it is possible to include templates which are a mixture of a pair of pronouns and a pair of occurrences of a specific entity. These templates can also be used to identify the more latent non-adjacent patterns; but these have much higher probability of occurring than patterns composed entirely of entities, hence will have lower precision. We elected to include those mixed non-adjacent patterns where the second and third constituents are named entities, to be among the potential templates in the interest of recall, as illustrated by the example below:

1. *"[Part1]* **She** *[Part2]* **the_friend_of_USERZ**
2. "**the_friend_of_USERZ** *[Part4]* **her** *[Part5]" and*
3. *"[Part1]* **She** *[Part2]* **the_friend_of_USERZ** *[Part3]* **the_friend_of_USERZ** *[Part4]* **her** *[Part5]" and*
   ... ...

To sum it up briefly, the possible templates $\vec{T}$, is composed of four components, each being a named entity or a pronoun. Each named entity is specified by an user identity i_m, an entity type, τ_n, and an ordered set of relations r_l. For example, in the entity name

$the\_niece\_of\_a\_friend\_of\_the\_mother\_of\_USER\_Adalia$, i_m is $USER\_Adalia$, and the type τ_n is $\mathbb{R}(\mathbb{S}\{\mathbb{R}(\cdot)\})$, with the vector specifying the specific relations $r_l =< Niece,\ Friend,\ Mother >$, applied to τ_m to form $niece(friend\{mother(\cdot)\})$, which in turn is applied to i_m to form the non-unique entity $niece(friend\{mother(USER\_Adalia)\})$.

Figure 4.1: An illustration of how the two DFAs and the relevant constraints that are placed on them, before they are combined into a single DFA for pattern recognition



4.2.3 Representation of reciprocal relationships in the text

Our algorithm takes into account both canonical and latent reciprocities. For those reciprocal actions occurring adjacently, they are identifiable by an unified pattern of adjacent clauses. This occurs in This occurs in a single compound sentence, or in two adjacent sentences (as delimited by punctuation). The distinction here is not significant, since the English conventions of punctuation are not often followed in these forums. The following two examples are from the pre-processed corpus, the first is two clauses within a sentence, while the second is adjacent clauses separated by punctuation:

if/IN she/PRP approached/VBD USER_LOSTINCONFUSION/NNP

USER_LOSTINCONFUSION/NNP would/MD be/VB automatically/RB more/RBR

interested/JJ in/IN her/PRP

she/PRP through/IN USER_SHHEADZ/NNP under/IN the/DT bus/NN ,/,

USER_SHHEADZ/NNP stills/VBZ miss/VB her/PRP like/IN crazy/JJ

The non-adjacent pairs of eventualities are indirectly inferred mutual relationship, where two entities or two classes of entities perform a pair of actions with the counter-party as a patient or recipient, with the reciprocants performing action at different times. This second type by definition is much more numerous than the first, but relies on the specificity of the named elements within the templates for any reasonable reliability. Even though the second type does not require adjacency, the scale of distance within the text between the two eventualities in the pair has an impact on the reliability of such identifications, this we will deal with slightly later. The following are a few examples of non-adjacent occurrences with varying amount of intervening material:

1. *she/PRP wanted/VBD to/TO come/VB with/IN USER_SHHEADZ/NNP* ··· [two intervening sentences] ··· *USER_SHHEADZ/NNP has/VBZ deleted/VBN her/PRP number/NN*
2. *USER_YGGDRASIL/NNP has/VBZ hurt/VBN him/PRP* ··· [five intervening sentences] ··· *he/PRP has/VBZ left/VBN USER_YGGDRASIL/NNP*
3. *USER_VASHTI/NNP sees/VBZ them/PRP out/RP* ··· [ten intervening sentences] ··· *they/PRP approach/VBP USER_VASHTI/NNP*

4.2.4 Procedure for pattern discovery

After additional preprocessing steps for syntax (University of Tokyo, Sagae & Tsujii 2007), NER with our own module supplemented by results from Stanford NLP's NER module (Manning et. al. 2014), and filters, the data is scanned for all potential components of templates, the pronouns and identified entities. A set of potential templates are formed as defined in the preceding section's formulation. The explicit patterns can be found by building simple DFAs each accepting a formal language defined according to one of the potential templates. The standard form of these state machines has the representation:

$$DFA_i = \begin{cases} \Sigma = \{\sigma_a, \sigma_b, ...\sigma_z\} \\ Q = \{q_i \mid 0 \leq i \leq |Q| - 1\} \cup \{q_f, \} \\ q_0 \in Q \\ \delta \subseteq Q^2 \times \Sigma \\ F \subseteq Q \end{cases} \tag{4.2}$$

The regular patterns that we want to target in the text corpus can be recognized by building state machines that correspond to some form of the general template of $< Entity_x, Entity_y, Entity_y, Entity_x >$. Hence, in the case of two reciprocal clauses each with a monotransitive VP, the algorithm can be described by the form of the following state machine (here, σ_0 represents any symbol that does not represent an entity, and v_t represents any transitive verb surface form):

$$\binom{DFA}{p_{tr}^{xyyx}} \begin{cases} \Sigma = \{\sigma_{Entity^x}, \sigma_{Entity^y}, \sigma_{v^{tr}}, \sigma_0\} \\ Q = \{q_{/xv_tyyv_tx},\ q_{x/v_tyyv_tx}, \\ \qquad q_{xv_t/yyv_tx},\ q_{xv_ty/yv_tx}, \\ \qquad q_{xv_tyy/v_tx},\ q_{xv_tyyv_t/x}, \\ \qquad q_{xv_tyyv_tx/}, q_{trap}\} \\ q_0 = q_{/xv_tyyv_tx} \\ \delta = \delta_w \bigcup \delta_0 \bigcup \delta_{trap} \\ \begin{cases} \delta_w = \Big\{ \\ (q_{/xv_tyyv_tx}, q_{x/v_tyyv_tx}, \sigma_{Entity^x}), \\ (q_{x/v_tyyv_tx}, q_{xv_t/yyv_tx}, \sigma_{v^{tr}}), \\ (q_{xv_t/yyv_tx}, q_{xv_ty/yv_tx}, \sigma_{Entity^y}), \\ (q_{xv_ty/yv_tx}, q_{xv_tyy/v_tx}, \sigma_{Entity^y}), \\ (q_{xv_tyy/v_tx}, q_{xv_tyyv_t/x}, \sigma_{v^{tr}}), \\ (q_{xv_tyyv_t/x}, q_{xv_tyyv_tx/}, \sigma_{Entity^x}), \\ \qquad \Big\} \\ \delta_0 = \Big\{(q_{w^r/w^s}, q_{w^r/w^s}, \sigma_0) \mid \\ \qquad w^r.w^s = w^{xv_tyyv_tx}\Big\} \\ \delta_{trap} = \bigcup\limits_{\sigma_i}^{\sigma_i \in \Sigma \backslash \{\sigma_0\}} \\ \qquad \Big\{(q_{r./\sigma_j.s}, q_{trap}, \sigma_i) \ \Big| \\ \qquad w^r.\sigma_i.w^s = w^{xv_tyyv_tx}, \\ \qquad \sigma_j \neq \sigma_i \ \Big\} \ \bigcup \\ \quad \Big\{(q_{trap}, q_{trap}, \sigma_n) \Big| \sigma_n \in \Sigma\Big\} \end{cases} \\ F = \{q_{xv_tyyv_tx/},\} \end{cases} \Bigg| \begin{cases} .q_{w^u/w^v} \\ .w_u = \\ \quad p^{seen} \\ .w_v = \\ p^{unseen} \\ .w_0 = \\ \quad (\sigma_0)^m \end{cases} \tag{4.3}$$

We may also want to allow the two events to contain ditransitive VPs. In this case, a modified DFA would accept the appropriate form expected from a pair of ditransitive VPs. In the following, $Entity_\alpha$ is a wildcard entity that can represent any entity other than $Entity_x$ or $Entity_y$, and can intervene in multiple positions within the reciprocity pattern. And the resulting DFA would have a modified Q and δ to account for this change; the δ_α in the generalized definition below contain the necessary additional transitions (here, v_d represents any ditransitive verb surface form). This is algorithmically represented as:

$$
\binom{DFA}{p_{ditr}^{xyyx}}
\left\{
\begin{array}{l}
\Sigma = \{\sigma_{Entity^x}, \sigma_{Entity^y}, \sigma_{v_ditr}, \\
\qquad \sigma_{Entity^\alpha}, \sigma_0\} \\
Q = \{q_{/xv_dyyv_dx},\ q_{x/v_dyyv_dx}, \\
\qquad q_{xv_d/yyv_dx},\ q_{xv_dy/yv_dx}, \\
\qquad q_{xv_dyy/v_dx},\ q_{xv_dyyv_d/x}, \\
\qquad q_{xv_dyyv_dx/}, q_{trap}\} \\
q_0 = q_{/xv_dyyv_dx} \\
\delta = \delta_w \bigcup \delta_0 \bigcup \delta_{trap} \bigcup \delta_\alpha \\
\left\{
\begin{array}{l}
\delta_w = \Big\{ \\
(q_{/xv_dyyv_dx}, q_{x/v_dyyv_dx}, \sigma_{Entity^x}), \\
(q_{x/v_dyyv_dx}, q_{xv_d/yyv_dx}, \sigma_{v_ditr}), \\
(q_{xv_d/yyv_dx}, q_{xv_dy/yv_dx}, \sigma_{Entity^y}), \\
(q_{xv_dy/yv_dx}, q_{xv_dyy/v_dx}, \sigma_{Entity^y}), \\
(q_{xv_dyy/v_dx}, q_{xv_dyyv_d/x}, \sigma_{v_ditr}), \\
(q_{xv_dyyv_d/x}, q_{xv_dyyv_dx/}, \sigma_{Entity^x}), \\
\qquad \Big\} \\
\delta_0 = \Big\{(q_{w^r/w^s}, q_{w^r/w^s}, \sigma_0) \mid \\
\qquad w^r.w^s = w^{xv_dyyv_dx}\Big\} \\
\delta_\alpha = \Big\{ \ (q_{w^r/w^s}, q_z, \sigma_{Entity^\alpha}) \ \Big| \\
\ q_z = q_{w^r/w^s} \leftarrow q_{w^r/w^s} \in Q_\alpha \ , \\
\ q_z = q_{trap} \leftarrow q_{w^r/w^s} \notin Q_\alpha \ , \\
\ Q_\alpha = \{q_{xv_d/yyv_dx}, q_{xv_dy/yv_dx}, \\
\ q_{xv_dyyv_d/x}, q_{xv_dyyv_dx/}\} \\
\qquad \Big\} \\
\delta_{trap} = \quad \bigcup_{\sigma_i}^{\sigma_i \in \Sigma \setminus \{\sigma_0\}} \\
\qquad \Big\{(q_{r./\sigma_j.s}, q_{trap}, \sigma_i) \ \Big| \\
\qquad w^r.\sigma_i.w^s = w^{xv_dyyv_dx}, \\
\qquad \sigma_j \neq \sigma_i \ \Big\} \quad \bigcup \\
\quad \Big\{(q_{trap}, q_{trap}, \sigma_n) \ \Big| \ \sigma_n \in \Sigma\Big\}
\end{array}
\right. \\
F = \{q_{xv_dyyv_dx/}, \}
\end{array}
\right.
\qquad \Bigg|
\begin{cases}
\cdot q_{w^u/w^v} \\
.w_u = \\
\quad p^{seen} \\
.w_v = \\
p^{unseen} \\
.w_0 = \\
\quad (\sigma_0)^m
\end{cases}
\tag{4.4}
$$

4.2.4.1 Altering the Adjacency Restriction

During the previous study that we conducted (Paul et. al., 2009) several sources of data were used, including BNC, which yielded the best results for reciprocities through our algorithm with simple pronoun templates, restricted to adjacent pairs. Here, the user-generated web data has a very different composition from BNC, and resulted in very low recall of reciprocities if we replicated the previous technique. This is due to the fact that most of the reciprocities represented in the text do not reside in a single clause. So if we only look at the adjacent cases of reciprocities, even when including templates composed of both pro-forms and named entities, there is very low rate of recall. On the other hand, if we refrain from imposing any limit on distance, there would be a very large number of potential such combinations, $\mathcal{O}(length(input)^2)$ of them. This leads to excessive pollution of the useful data for the ranking stage, and in turn leads to very low precision of final output. We attempted to utilize pure forms of both of these approaches, and have confirmed that the first leads to excessively low recall, while the second leads to excessively low precision.

The solution here is to adopt a more flexible, graded definition for *adjacency*, including the previously mentioned *loosely adjacent* candidates. We applied a coefficient to scale the scores from potential reciprocities, which is the inverse of the distance between the two clauses. The maximum distance is limited to an experimentally determined value that provides the best combination of recall and time complexity; determined to be a distance of 4 for the Family and Marriage Relationship Forum, 12 for LoveForums. Within this limit, we assume that a pronoun (with the identical gender if 3rd sing) within two candidate eventualities would very likely to refer to the same individual. The identical named entities are already guaranteed to refer to the same individual, with a few exceptions such as *USER_ANONYMOUS* that refer to multiple individuals due to forum administration.

$$Adjacency\_Scale = \begin{cases} \frac{1}{(Distance(e1,e2)-Min\_Distance)}, \\ \quad \text{if } n \leq \text{ Max\_Distance} \\ 0.0, \text{ if } n > \text{ Max\_Distance} \end{cases} \tag{4.5}$$

The maximum allowed distance is k, as determined through the experimental results for each set of data. A potential template in the form of $[\cdot Entity_1 \cdot Entity_2 \cdot Entity_2 \cdot Entity_1 \cdot]$

would split into two halves of [$\cdot Entity_1 \cdot Entity_2 \cdot$] and [$\cdot Entity_2 \cdot Entity_1 \cdot$], each being represented by a simple DFA, resulting in the pair DFA_0 and DFA_1 for each reciprocity. And the two DFAs below are used in place of the template DFA described in Figure 4.3 earlier. These are based on the version of the template DFA where both reciprocities employ monotransitive verbs, ones with ditransitive verbs, or a mixture of verbs with several adicities would have a similar structure. The following is the DFA representation of the mechanism in the $Entity_x \Rightarrow Entity_y$ direction:

$$DFA_0 = \begin{cases} \Sigma = \{\sigma_{Enity^x}, \sigma_{Entity^y}, \sigma_{v^{tr}}, \sigma_0\} \\ Q = \{q_{/xv_ty},\ q_{x/v_ty},\ q_{xv_t/y},\ q_{xv_ty/}, q_{trap}\} \\ q_0 = q_{/xv_ty} \\ \delta = \delta_w \bigcup \delta_0 \bigcup \delta_f \\ \begin{cases} \delta_w = \Big\{(q_{/xv_ty}, q_{x/v_ty}, \sigma_{Entity^x}), \\ \qquad (q_{x/v_ty}, q_{xv_t/y}, \sigma_{v^{tr}}), \\ \qquad (q_{xv_t/y}, q_{xv_ty/}, \sigma_{Entity^y})\Big\} \\ \delta_0 = \{(q_{w^u/w^v}, q_{w^u/w^v}, \sigma_0) \\ \quad \mid w^u + w^v = w^{xv_ty}\} \\ \delta_{trap} = \bigcup\limits_{\sigma_i}^{\sigma_i \in \Sigma \setminus \{\sigma_0\}} \\ \qquad \Big\{(q_{r./\sigma_j.s}, q_{trap}, \sigma_i) \ \Big| \\ \qquad w^r.\sigma_i.w^s = w^{yv_tx}, \\ \qquad \sigma_j \neq \sigma_i \ \Big\} \ \bigcup \\ \quad \Big\{(q_{trap}, q_{trap}, \sigma_n) \Big| \sigma_n \in \Sigma\Big\} \end{cases} \\ F = \{q_{xv_ty/},\} \end{cases} \tag{4.6}$$

And the following is the mechanism in the reversed $Entity_y \Rightarrow Entity_x$ direction, where both together would produce a complete pattern that identifies a likely reciprocal eventuality pair:

$$DFA_1 = \begin{cases} \Sigma = \{\sigma_{Enity^x}, \sigma_{Entity^y}, \sigma_{v_tr}, \sigma_0\} \\ Q = \{q_{/yv_tx},\ q_{y/v_tx},\ q_{yv_t/x},\ q_{yv_tx/}, q_{trap}\} \\ q_0 = q_{/yv_tx} \\ \delta = \delta_w \bigcup \delta_0 \bigcup \delta_f \\ \begin{cases} \delta_w = \Big\{(q_{/yv_tx}, q_{y/v_tx}, \sigma_{Entity^y}), \\ \qquad (q_{y/v_tx}, q_{yv_t/x}, \sigma_{v_tr}), \\ \qquad (q_{yv_t/x}, q_{yv_tx/}, \sigma_{Entity^x})\Big\} \\ \delta_0 = \{(q_{w^u/w^v}, q_{w^u/w^v}, \sigma_0) \\ \quad \mid w^u + w^v = w^{yv_tx}\} \\ \delta_{trap} = \overset{\sigma_i \in \Sigma \setminus \{\sigma_0\}}{\underset{\sigma_i}{\bigcup}} \\ \qquad \Big\{(q_{r./\sigma_j.s}, q_{trap}, \sigma_i) \ \Big| \\ \qquad w^r.\sigma_i.w^s = w^{yv_tx}, \\ \qquad \sigma_j \neq \sigma_i \ \Big\} \quad \bigcup \\ \quad \Big\{(q_{trap}, q_{trap}, \sigma_n) \Big| \sigma_n \in \Sigma\Big\} \end{cases} \\ F = \{q_{yv_tx/},\} \end{cases} \tag{4.7}$$

While the text is scanned, if DFA_0 is a match when the scanning process hits clause n, the text from clause $n+1$ to $n+k$ are scanned by DFA_1. And any given match by DFA_1 is returned as a potential eventuality pair with clause n. The pair must satisfy the constraints on the entity pairs that were part of the NER process. In this case, according to the five PARTs system, the pattern within the middle partition (PART3), is rendered largely useless, since it is a concatenation of parts of two independent clauses, and should be discounted. The performance of this module will be evaluated independently at a later time.

4.3 Scoring function

The baseline method of computing the prominence of the discovered patterns would be to simply consider the frequency of the five partitions of a pattern individually. However, as our preliminary experiments shows, most of the patterns, especially the longer ones, seldom occur more than once in the entire corpus. Thus merely using the token frequencies in histogram on these highly infrequent pattern types would produce very poor ranking. Only when the individual parts of the patterns are very short, would there be sufficient number of instances of these types for there to be a nice frequency distribution in order to rank them by frequency.

So we developed an alternative scoring system in lieu of mere token frequencies. As defined here, a sub-pattern is simply a substring of the original complete surface pattern; such as for the pattern *he did not accept the offer* , its next level sub-patterns (length of $n-1$) would be *he did not accept the* , and *did not accept the offer.* Taking into account the frequencies of the sub-patterns occurring within instances of each partition, we devise a inductive definition of this score: For any $n > 1$ (where n is the length of the sub-pattern, and SEQ^n is the set of sub-patterns of length n), with a discount factor to scale the overall score that we will explain shortly:

$$Score(seq_i) = \begin{cases} Disc(freq(seq_i)) + \displaystyle\sum_{seq_j \in SEQ^{n-1}} seq_j \quad \cdot & \\ \quad Disc(Score(seq(n-1))), & \text{if } n > 1 \\ freq(seq(n)), & \text{if } n = 1 \end{cases} \tag{4.8}$$

$$where \;\; seq_i \in SEQ^n$$

In addition, in order to insure a fair ranking over the extracted patterns with different lengths for each partition, we need to normalize the scores obtained for the five PARTs. In other words, we need to scale the scores obtained for each partition to discount the scores of longer partitions, so that the maximum possible score would remain the same regardless of the length of the partition. The discount factor gives fractionally less weight to the next set of sub-patterns with each recursive call. So we use the following formula to discount the score for each of PARTs, where n is the length of the subsequence:

$$Disc = \begin{cases} (1.0 * fraction) * \frac{fraction^{m-n-1}}{m-n+1}, & \text{if } n > 1 \\ \frac{fraction^{m-n}}{m-n+1}, & \text{if } n = 1 \end{cases} \tag{4.9}$$

Fraction is some predetermined parameter that has been empirically tested to give reasonable results, in our case we set it to 0.5 for the purposes of this experiment. The variable m is the length of the relevant partitions combined. This allows not only the frequency of the exact pattern to contribute to the score, but also occurrences of similar patterns would contribute to the score to a lesser extent. We compute the score to be ranked as $\prod_{\alpha \in PARTs} Score(\alpha)$. The implementation of this recursive scoring system is done through dynamic programming, and is efficient in use of time and space. The system is designed to be a ranking procedure, and the binary classification is done using an experimentally determined fraction of top-ranked patterns. The performance of this module will be evaluated independently at a later time.

4.4 Reciprocity post-processing

After the results are obtained, it is put through some post-processing with several purposes. These are fairly compute intensive stages, but since the dataset has been narrowed down to 0.1% of the original size, the tasks are very manageable in terms of time complexity at this point.

4.4.1 Grammar and Spelling Amelioration

The first and most mundane of these tasks is to filter out some of the most common grammatical and spelling mistakes that we have observed along the previous several steps of processing. This is important since the original data often does not observe conventional English grammar, or filled with slang that are not part of standard English; this also catches a few of grammatical errors introduced through earlier processes. This is done solely to make the final output text more readable to any annotator, so the discovery and ranking of patterns are done on the text with the original grammatical and spelling mistakes. The most significant part of the grammatical correction process is finding and replacing the incorrect verbs with the correct form in terms of tense, mood, transitivity, and finiteness. The text is partially parsed to reveal the position of closed classes of words. The earlier mentioned morphological analyzer is employed again to provide the necessary information about English verbs. Information from WordNet (Fellbaum, 1998) is also used to detect certain lexical ambiguities. For example, the error of *[Entity1] tried to been [Entity2]'s friend* is corrected to *[Entity1] tried to be [Entity2]'s friend* based on the rule of in.

4.4.2 Inclusion of semantically relevant portions of Part 3 and Part 5

Even though the main verb(s) of each candidate clause provide the key piece of meaning for the eventuality, often in English, as in most Germanic languages, contain separable particles that also carry a part of the meaning, providing additional specification for the main verb. An example would be *he* ***turns*** *the light* ***off***, where the particle comes sequentially after the direct object, which would be the boundary between *Part2* and *Part3* as well as between *Part3* and *Part5* in the scheme used here. So we need to selectively include portions of these in order to complete the construction of each eventuality, such as in the case of $ENTITY_X$ ***turned*** $ENTITY_Y$ ***on***. There are less frequent content from Part3 and Part5 that also need to be included, as when the second entity in the eventuality heads an embedded clause, as in $ENTITY_Y$ *had some qualities* ***that*** $ENTITY_X$ ***admires*** or $ENTITY_X$ *told* $ENTITY_Y$ *how USER_GIGABITCH felt*. The cues in the text in the form of the complementizer *that* and the finite verb *admires* that immediately follows the second entity tells us that the second content verb should be included as part of the eventuality, in order for it to make sense to any annotator. A number of other similar sets of cues let us know if and what from *Part3* and *Part5* of the patterns should be included as essential parts of the semantic content of the eventualities. We constructed mechanisms to examine the partially parsed candidate clauses to see if such cues exist, and add the appropriate portions of Parts 3 and 5.

4.4.3 Identifying core meanings

The third component of the post-processing steps is the identification of the core of each eventuality. Majority of the time, the main content verb would suffice as the representative of the core meaning. At other times, such as the aforementioned verbs with separable particles, the extra component is necessary; these are identified in the same way as in the aforementioned step.

There are also more complex constructions with more than one content verb. The verbs and their tense, aspect, voice are identified through the combination of word-net and our own English verb morphological analyzer. These include many infinitival constructions such as $ENTITY_X$ ***wants*** *to* ***help*** $ENTITY_Y$, and constructions containing small clauses such as $ENTITY_X$ ***stopped contacting*** $ENTITY_Y$. In these cases, the two or more of the content verbs in the main as well as the subordinate clauses are necessary to carry to full meaning of the eventuality. There are also instances where the second entity is an oblique object but a fitting argument can still be made for the eventuality, such as in the case of $ENTITY_X$ ***felt*** *stronger* ***towards*** $ENTITY_Y$, or in the case of $ENTITY_X$ *has since* ***ran into*** $ENTITY_Y$. In these cases, the inclusion of the lexical head of the PP (preposition) and the head of the subordinate NP (usually noun or named entity) would also be necessary to convey the accurate eventuality relation between the two entities.

This is an artifact of the transitivity of the particular lexical entry, as we can easily substitute these with *loved more* and *encountered* to achieve eventualities with nearly identical meaning with transitive verbs; so semantically they are similar to the simplest eventualities. Another frequently observed complex eventuality is where the second entity being the subject of an embedded relative clause/gerund that serves the actual patient, recipient, or goal. An example would be the case of $ENTITY_Y$ *may hate what* $ENTITY_X$ *is saying*, where the *hate sayings of* $Entity_X$ would be the accurate representation, so the content verbs of matrix and relative clause form the core. There are also a number of other types that are more scarce, which also are identified and processed. The mechanism to recognize this is also in place during this post-processing stage. The components that make a significant semantic contribution are marked in the output form. These mechanisms that perform the extraction of specific components of each eventuality is a primitive fore-runner for the later far better developed frame extraction mechanism in Chapter 5. This ad hoc mechanism here would be replaced by a systematic approach to detection and movements using automata, and instead of a few key elements within the eventuality, a more comprehensive detection for frame components will be in place.

4.5 Finding optimal adjacency limit

Annotation of the result relies heavily on human judgment. In order to find the optimal adjacency limitation on the discovery of these patterns. We set aside a portion of the results consisting of pairs of eventualities for annotation by human. The annotation is also done with access to the raw text from the original web-crawl as reference when needed. The utility of accessing the reference material is limited, since the amount of effort that goes into verifying whether a pair of eventuality is indeed semantically reciprocal, when the eventuality pair itself does not provide enough information, is very high. And this additional resource is seldom used by the annotator.

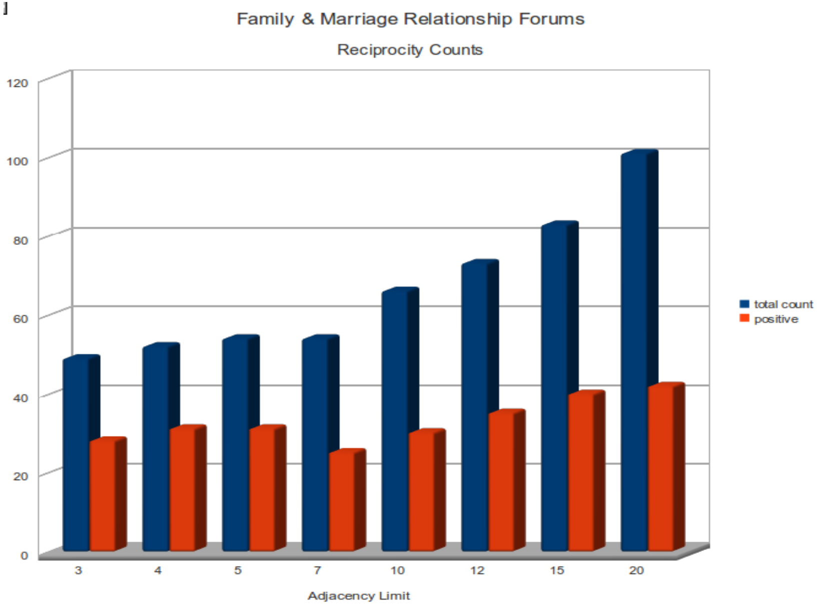
The size of the relevant context can become very large, depending on the **adjacency limit** that we choose for a particular run of the system. For the ease of annotation, we did not require the annotators to read all of the intervening material between a pair of eventualities. So for the annotation task, we permitted a third category of **indeterminate** in addition to **positive** and **negative**, which is used whenever a decision by the annotator of whether an eventuality pair is reciprocal cannot be made without detailed reading of the entire thread.

4.5.1 Optimal distance limit in Family and Marriage Relationship Forums reciprocities

Finding the optimal limit on the distance between an eventuality pair is essential, because the set of characteristics of the data and forum structure is distinct for each web-site. Many factors would influence the amount of leeway we can place between the pair, depending on the language use in the forum, the nature of the problems being discussed, the intimacy of knowledge of forum members in each other's lives, as well as the average number of clauses in threads and posts. This compels us to find the **adjacency limit** that provide the best results in terms of agreeing with the judgment of the annotators.

The performance for the *Family and Marriage Forums* site taking the annotator as the gold standard, the peak precision is 0.60, when conservatively grouping the **indeterminate** labels with the **negatives**. The peak performance occurs at the Adjacency limit of 4; so we can surmise that the vast majority of the valid reciprocities have the two components occurring within the distance of 4 or less.

Figure 4.2: Counts of marriage and relationship forums



4.5.2 Optimal distance limit in LoveForums reciprocities

The Performance for the second dataset has a peak of 0.62 agreement between the system and annotator, when grouping the **indeterminate** labels with **negatives**. This set has a peak with much higher **adjacency limit**, at 12; this shows that the reciprocity pairs in this dataset has the potential to be placed much further apart in the text. This is consistent with

Figure 4.3: Precisions of marriage and relationship forums

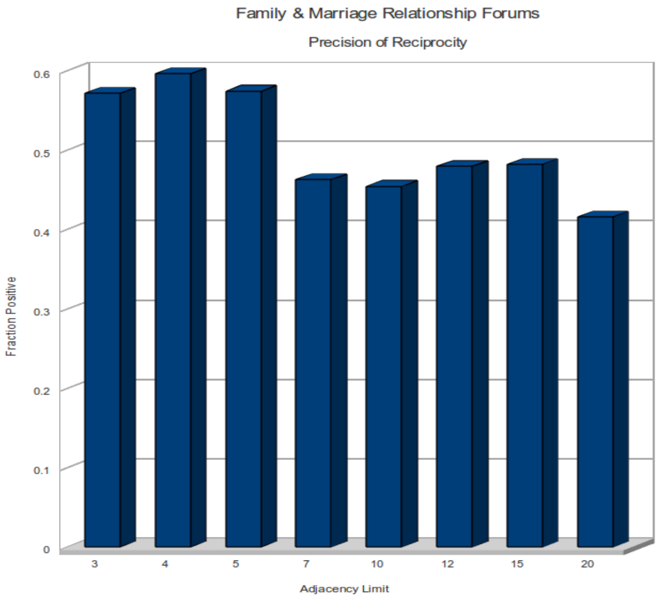
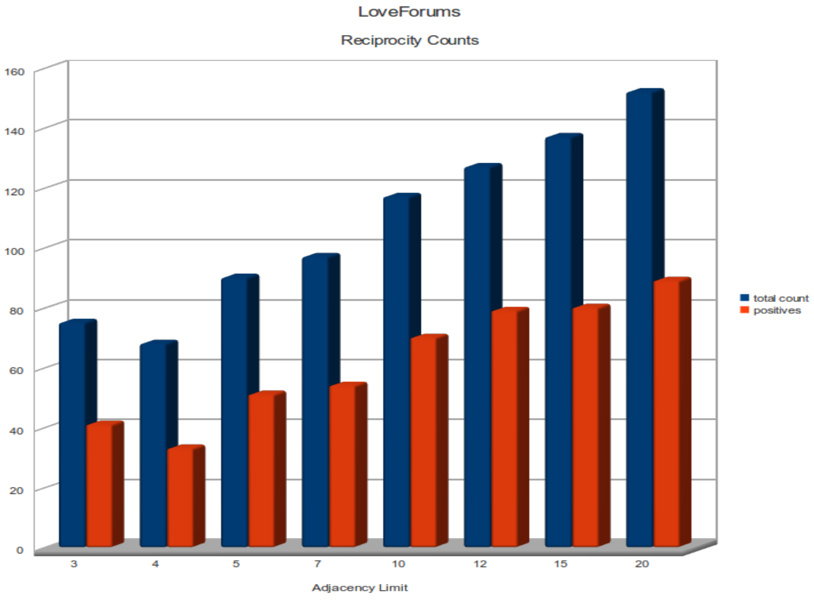
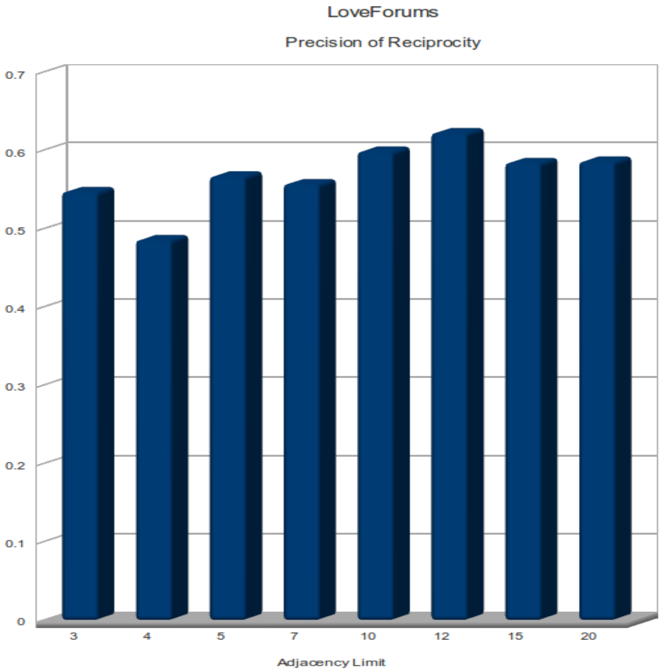


Figure 4.4: Counts of loveforums



the fact that the second dataset has much longer threads with an average of nearly 10 posts per thread; whereas the first dataset has an average of less than three posts per thread. So the potential for long distance reciprocity, which depends on the discourse context, should be

Figure 4.5: Precisions of loveforums



much higher here as well. This fact also indicates that the custom named entity identification and co-reference resolution did provide enough specificity to link two distant eventualities in most cases. This will in turn benefit the social networking model's ability to identify relationships between users and other entities present in the forum community.

4.6 Results and evaluation

There are two sets of results from each of the web forums, depending on which portions of the reciprocity patterns are being included in the final ranking of the results. The general form of the pattern is:

$$\textbf{\textit{Part1}}\ Entity_x\ \textbf{\textit{Part2}}\ Entity_y\ \textbf{\textit{Part3}}\ Entity_y\ \textbf{\textit{Part4}}\ Entity_x\ \textbf{\textit{Part5}}$$

The semantic cores of the two eventualities most often occupy **Part2** and **Part4** of the patterns; occasionally, the semantic cores will extend into **Part3** and **Part5** as we will discuss in detail later; **Part1** is seldom included in the final eventuality forms, as adverbial phrases, conjunctions, as well as appositives most often occupy that partition of each pattern. For the full patterns composed of the five partitions, the vast majority occur only once, given the minute probability of two exact forms of reciprocity pairs in the corpus. The classification of whether a pattern is causal is taking the top 0.01 fraction of the patterns in the ranking, which is close to that of our previous work (Paul et. al. 2009). Since that study had a strict adjacency requirement (patterns only contain eventualities that are immediately adjacent), so in this study, there are potentially a much larger number of possible patterns, even given a smaller data-set, depending on the eventual adjacency chosen.

4.6.1 Top ranked examples

Given that we have permitted a more generalized definition of reciprocal pairs, those examples with the two eventualities coming from the same sentence (distance 0), or from two adjacent sentences (distance 1), are the pairs that are adjacent according to the strict definition of adjacency, and this corresponds to the types of reciprocities that were targeted in the previous study (Paul et. al., 2009). These are the most accurate set of reciprocities extracted, because of the high probability that the pair is semantically related, and since these types of pairs are tested to be accurate in the previous study. The pairs that are fused together with *distance* > 1 have varying degrees of plausibility in terms of grammaticality and semantically well-formedness. For these pairs that are not strictly adjacent, there is often conjunctions in **Part3** of the formation (the partition of first eventuality clause after $Entity_y$ and the partition of the second eventuality preceding $Entity_x$) which are not valid conjunctions between the pair. These conjunctions would not semantically make sense for a reciprocal relation, such as the **Part3** conjunction in the following:

so $Entity_x$ *learned to accept* $Entity_y$ ***if*** $Entity_y$ *tells* $Entity_x$

4.6.1.1 Top Examples from family and relationship forums

For the resulting pairs of clauses in their surface forms, the entire pattern of five parts are displayed, for the readability for annotation. These are either the more canonical strictly adjacent reciprocal pairs, or some constructed representations from more latent forms, through the concatenation of two parts with varying amount of distance in between. We also need to keep in mind that some of the examples in the original text may not be entirely grammatical, which is often the case with forum data, especially in these cases where many users are not native speakers of English.

The following are the top 20 patterns from the first dataset extracted at the optimal adjacency limitation. In the results from the first forum, most of the top ranked examples can clearly be interpreted as reciprocal by the annotator, even without any additional contextual information. There is a sub-group of examples constructed from interrogative predicates with embedded questions, particularly involving the verb *ask*. This is not a semantically reciprocal relation between the action in the embedded clause and the act of inquiring, but it is similar to reciprocity in the sense that the reciprocal action of **asking** is the answer about the **action referenced in the embedded clause**. So even though it is not actually reciprocal, the relationship between these pairs is pragmatically related to reciprocity. There is one example that is labeled by the annotator as **indeterminate**:

Entity$_x$ ***loves*** *Entity*$_y$ *a lot* *Entity*$_y$ *is* ***changing*** *Entity*$_x$

Another notable type is not a directly reciprocal relation, but involves a more complex interaction among three individuals, which is in this example:

but *Entity*$_x$ ***knows*** *if USER_ANONYMOUS* ***tells*** *Entity*$_y$ *Entity*$_y$ ***would leave*** *Entity*$_x$

This is a special case of the type where the first eventuality is not actually between two individuals, but rather between an individual and an event. We will talk about the graph theoretic treatment of such cases in later sections of formulation the graph representation. This is not direct reciprocity, but is also pragmatically related to reciprocity in the sense that there is also a strong causal relationship between the actions performed in the two eventualities.

4.6.1.2 Top results from loveforums

Table 4.2 contains the top 20 of total of 216 examples at peak performance adjacency limitation (12) from the second source of data. Most top examples are clearly reciprocal,

Table 4.1: Marriage and relationship forum table

$Entity_x$ ACCEPTED $Entity_y$ and $Entity_y$ ACCEPTED $Entity_x$
STILL $Entity_x$ WANTS to STAYED with $Entity_y$ because $Entity_y$ MAKES $Entity_x$ happy
$Entity_x$ LOVES $Entity_y$ and $Entity_y$ LOVES $Entity_x$ too
$Entity_x$ had to LOVED $Entity_y$ if $Entity_y$ LOVES $Entity_x$
$Entity_x$ ACCEPTED $Entity_y$ and $Entity_y$ ACCEPTED $Entity_x$ he is nice
$Entity_x$ ASKED $Entity_y$ if $Entity_y$ could BE faithful to $Entity_x$
$Entity_x$ NOTICED $Entity_y$ and $Entity_y$ NOTICED $Entity_x$ as they WALKED in
$Entity_x$ ACCEPTED $Entity_y$ and $Entity_y$ ACCEPTED $Entity_x$ and UNDERSTANDING
$Entity_x$ LOVES $Entity_y$ a lot $Entity_y$ is CHANGING $Entity_x$
if $Entity_x$ LIKED $Entity_y$ inside $Entity_y$ BECAME more STIMULATING to $Entity_x$ sexually
$Entity_x$ NOTICED $Entity_y$ and $Entity_y$ NOTICED $Entity_x$ as they WALKED in polite ...
when $Entity_x$ ASKED $Entity_y$ a week ago how $Entity_y$ FELT about $Entity_x$
$Entity_x$ WANTS to STAYED with $Entity_y$ because $Entity_y$ MAKES $Entity_x$ happy
USER_ANONYMOUS KNOWN $Entity_x$ LOVES $Entity_y$ and $Entity_y$ LOVED $Entity_x$ too
$Entity_x$ EXCHANGED EMAILS with $Entity_y$ at WROUGHT $Entity_y$ EMAILED $Entity_x$'s home ACCOUNT
$Entity_x$ ASKED $Entity_y$ if $Entity_y$ had SEEN someone that FLOORED $Entity_x$
PLEASED REPLIED $Entity_x$ LOVES $Entity_y$ and $Entity_y$ LOVES $Entity_x$ too
USER_ANONYMOUS KNOWN $Entity_x$ LOVES $Entity_y$ and $Entity_y$ LOVED $Entity_x$ too marriage ...
but $Entity_x$ KNOWS if USER_ANONYMOUS TELLS $Entity_y$ $Entity_y$ WOULD LEAVED $Entity_x$
when $Entity_x$ TRIES to TALKED to $Entity_y$ $Entity_y$ SAYS its in $Entity_x$ HEAD
$Entity_x$ TOLD $Entity_y$ $Entity_y$ TOLD $Entity_x$ that is not faire

and recognized so by a human annotator. There is one instance with *ask* as in the other dataset. One example is obviously non-reciprocal:

*but $Entity_x$ was brilliant $Entity_y$ **wishes** $Entity_y$ had more friends **like** $Entity_x$*

Another example is a comparative in (1), also not reciprocal. The first eventuality here can be classified as a copula predicate syntactically, and is roughly equivalent to the reconstructed pair (2). Semantically, it is actually two eventualities, both from $Entity_y$ to $Entity_x$.

1. *obviously $Entity_x$ is less important to $Entity_y$ than $Entity_y$ is **telling** $Entity_x$*
2. *$Entity_y$ **considers** $Entity_x$ not as **important** as $Entity_x$ is **TOLD by** $Entity_y$*

Table 4.2: Loveforums table

so $Entity_x$ DID LIKED $Entity_y$ and $Entity_y$ LIKES $Entity_x$
since $Entity_x$ LIKED $Entity_y$ and $Entity_y$ LIKES $Entity_x$
since $Entity_x$ LIKED $Entity_y$ and $Entity_y$ LIKES $Entity_x$ hey
so $Entity_x$ LEARNED to ACCEPT $Entity_y$ if $Entity_y$ TELLS $Entity_x$
only $Entity_x$ TURNED $Entity_y$ on $Entity_y$ LOVES $Entity_x$
if $Entity_x$ ASKS $Entity_y$ if $Entity_y$ LOVES $Entity_x$
$Entity_x$ could SEEN $Entity_y$ online $Entity_y$ could SEEN $Entity_x$ online
DID $Entity_x$ BREAK UP with $Entity_y$ or DID $Entity_y$ BREAK UP with $Entity_x$
obviously $Entity_x$ is less important to $Entity_y$ than $Entity_y$ is TELLING $Entity_x$
$Entity_x$ TOLD $Entity_y$ $Entity_y$ LIED to $Entity_x$ USER_LOVEADMIN said
if $Entity_x$ DIGGED into $Entity_y$ past $Entity_y$ is in LOVED with $Entity_x$
since $Entity_x$ LIKED $Entity_y$ and $Entity_y$ LIKES $Entity_x$ could USER_FULLOFSIGHS KISSED you
if $Entity_x$ FEELS the same way about $Entity_y$ as $Entity_y$ DID about $Entity_x$
since $Entity_x$ LIKED $Entity_y$ and $Entity_y$ LIKES $Entity_x$ not much EXPERIENCED here
since $Entity_x$ LIKED $Entity_y$ and $Entity_y$ LIKES $Entity_x$ she ALLOWS that
should $Entity_x$ STILLED TRIED to BE friends with $Entity_y$ if $Entity_y$ REJECTS $Entity_x$
but $Entity_x$ was brilliant $Entity_y$ WISHES $Entity_y$ had more friends LIKE $Entity_x$
$Entity_x$ TOLD $Entity_y$ how USER_GIGABITCH FELT and $Entity_y$ TOLD $Entity_x$
since $Entity_x$ LIKED $Entity_y$ and $Entity_y$ LIKES $Entity_x$ haha again
why is $Entity_x$ CALLING $Entity_y$ if $Entity_y$ said he WOULD HAVE CALLED $Entity_x$

4.6.1.3 Characterization of the extracted examples

The top ranked examples extracted from each data source are presented in Tables 4.1 and 4.2. The majority of the top-ranked reciprocity pairs are clauses that describe nearly identical set of actions that $Entity_x$ and $Entity_y$ mutually perform on each other. This is expected, as the nature of a reciprocal relationship often dictates that similar benefit, detriment, or a change of state is conferred on each of the pairs of reciprocal entities on the other, such as:

1. since $Entity_x$ LIKED $Entity_y$ and $Entity_y$ LIKES $Entity_x$ hey
2. why is $Entity_x$ CALLING $Entity_y$ if $Entity_y$ said he WOULD HAVE CALLED $Entity_x$
3. $Entity_x$ NOTICED $Entity_y$ and $Entity_y$ NOTICED $Entity_x$ as they WALKED in

Some example reflects a different type of reciprocity, representing a type of connection between the reciprocal pair, which is most aptly described as an *exclusive-or* (or $\bigoplus$ for short). The semantic link between the two is that the occurrence of either one of the actions would obviate the need or the precondition for performing the other, as in the following:

1. DID $Entity_x$ BREAK UP with $Entity_y$ or DID $Entity_y$ BREAK UP with $Entity_x$
2. why is $Entity_x$ CALLING $Entity_y$ if $Entity_y$ said he WOULD HAVE CALLED $Entity_x$

There are a few example where the reciprocal pair does not fit into the narrow definition of a reciprocal pair. At least one of the two actions is actually a composite action involving a third entity. This type of examples would be a better fit if we include the broader definition of the reciprocal pair if more latent forms of reciprocity are permitted.

1. $Entity_x$ ASKED $Entity_y$ if $Entity_y$ had SEEN someone that FLOORED $Entity_x$
2. but $Entity_x$ was brilliant $Entity_y$ WISHES $Entity_y$ had more friends LIKE $Entity_x$

4.6.1.4 Reciprocity and other linguistic features

We cannot have complete certainty of the existence of reciprocity in some example by looking only at syntactic or lexical cues. Many of our examples lack the unambiguous signs present in some reciprocities that are easy to detect from the superficial features in the text. Such examples would be marked by some indicative reciprocal adverbial or a *bipartite quantifier*, such as these below:

- $Entity_x$ gave $Entity_y$ the amulet and $Entity_y$ gave it RIGHT BACK to $Entity_x$
- $Entity_x$ and $Entity_y$ congratulated EACH OTHER

This is due to the sparseness of such occurrences, as well as our template based approach, that we used to identify these forms. The types of reciprocities that we seek require human judgment to reach certainty, since even in the case where these are adjacent in the original text, the forms may not differ from other non-reciprocal pairs of expressions, such as in the following examples:

- $Entity_x$ shared his good news with $Entity_y$ and $Entity_y$ congratulated $Entity_x$ that day

- *Entity*$_x$ shared his sumptuous meal with *Entity*$_y$ and *Entity*$_y$ congratulated *Entity*$_x$ that day
- *Entity*$_x$ shared his incessant frustrations with *Entity*$_y$ and *Entity*$_y$ congratulated *Entity*$_x$ that day

Although not all annotators may agree on these, the first example is much more likely to be annotated as reciprocity, the second example can only be considered by a few as marginally reciprocal constructions, and it would be surprising if anyone considers the third example as reciprocity. This occurs despite no syntactic distinction exists among the three examples, and the lexical difference occurs within only one NP in the first eventuality. This underscores the importance of taking into account all of the cues in the different partitions of each pattern, especially those that occur within the core portions of the eventualities themselves, *Part2* and *Part4* within our patterns.

A few of the example contain strong syntactic and lexical cues for reciprocity, such as in the example of *if ENTITY*$_1$ *FEELS the same way about ENTITY*$_2$ *as ENTITY*$_2$ ***did*** *about ENTITY*$_1$, where *Entity*$_1$ *[VP]* **the same way ...** *Entity*$_2$ **as** *Entity*$_2$ **[VP]** *Entity*$_1$ in itself contains indicative elements that leads to a reciprocal interpretation by the annotator. Some eventuality pairs contain nearly identical VP or sharing the same main verb for each of the two eventuality clauses, such as *Entity*$_1$ ***liked*** *Entity*$_2$ *and Entity*$_2$ ***likes*** *Entity*$_1$ *and Entity*$_1$ ***calling*** *Entity*$_2$ *if Entity*$_2$ *said he* ***would call*** *Entity*$_1$. These can also be extracted with a number of simple rules and a morphological analyzer. This type of example basically shows that similar surface forms of the VP or similar main verbs is a strong indicator of potential reciprocity between a pair of clauses.

Certain other top ranked examples also demonstrate deeper semantic similarity between the core VPs of the two eventuality clauses, while lacking overt cues or any superficial similarity at the syntactic or lexical level. One example of such is *if Entity*$_1$ ***liked*** *Entity*$_2$*('s) inside Entity*$_2$ ***became*** *more* ***stimulating to*** *Entity*$_1$ *sexually.* This particular instance of potential reciprocity is easy for a human annotator to observe, but it is very difficult to identify with mechanical rules. Thus the use of the basic template as well as three of the pattern parts *liked*, *became more stimulating to* ***someone*** *sexually* are essential to identifying and ranking this reciprocity. The similarity and relation between these two clauses lies in deeper levels of semantics. A similar example is the form *only* $Entity_1$ ***turned*** $Entity_2$ ***on,*** $Entity_2$ ***loves*** $Entity_1$, which was in fact not observed as an adjacent pair in the source, is found by the algorithm and ranked very highly.

There are other examples in the extracted results that were actually not deliberate by the algorithm's design, but was nevertheless semantically reciprocal. One such example would be the pair *but* $Entity_1$ *was brilliant* $Entity_2$ ***wishes*** $Entity_2$ *had more friends* ***like*** $Entity_1$, which does not conform to the intended form of $[_{Clause_1}$ $[Entity_1]$ V_1 $[Entity_2]]$ *[Conjunction]* $[_{Clause_2}$ $[Entity_2]$ V_2 $[Entity_1]]$ by any of the patterned DFAs, where each of the clauses is bracketed by the appropriate pair of entities. So this serves as an incidental true positive, in terms of the original template. Although the way the five part patterns were extracted may have something to do why this is highly ranked.

4.6.1.5 Possible false negatives by annotators

There are some instances where the system may have identified true reciprocities that are marked **negative** by an annotator, the following are some examples. A few of the examples are potentially reciprocal, but the logical ordering between the eventuality pair may have been reversed by the algorithm. The reason such exist is the inherent lack of linearity in the organization of the forum data (same issue faced in NER), and thus there is some probability of some distant reciprocal pair being reversed during the algorithm. The following example can be more easily interpreted as reciprocal and logically connected by reversing the order of the eventualities:

- $Entity_x$ *confronted* $Entity_y$ *CONJ* $Entity_y$ *pushed* $Entity_x$ *away emotionally*

Certain other examples were not identified to be reciprocal by the annotator due to the lack of any clear causal or entailment relationship between the eventualities. But these often exhibit the strong likelihood of a common cause, as in the following, where both eventualities follow from a close long term friendship between the entities:

- *$Entity_x$ tried to be $Entity_y$('s) friend CONJ $Entity_y$ misses $Entity_x$*

These are borderline cases in human judgment, where the pair of actions are clearly related, but does not fit the traditional definition of semantic reciprocity.

There are additional cases where the eventuality pair express not reciprocity, but the irony in the lack of expected reciprocal action. In the following, the eventuality pair expresses incommensurate pair of actions where $Entity_y$ reciprocates evil for the good $Entity_x$ does:

- *$Entity_x$ gave $Entity_y$ all he had ... ... $Entity_y$ giving $Entity_x$ treachery*

Or in the following, the eventuality pair expresses the desire for a relationship in one direction, but the total lack of interest in the other:

- *$Entity_x$ texts $Entity_y$ CONJ all $Entity_y$ did was ignore $Entity_x$*

This type is also not direct reciprocity, but implies that an expectation of reciprocity exists from the perspective of the writer. All of the above are cases where the algorithm found features that cannot be reciprocally interpreted directly by an annotator, but implies some form of underlying reciprocity recognized by the community.

4.6.2 Extracted eventuality forms

The final eventualities are compiled from the two sets of core meanings from each pair of the eventualities. These give a succinct but reasonable portrayals of the semantic content of the eventualities without any extraneous information. The following are the top ranked samples from the loveforums dataset:

Many of the samples from loveforums are symmetrically reciprocal, such as

$$see(Entity_x, Entity_y)$$

$$see(Entity_y, Entity_x)$$

, with the same action in both eventualities. For those samples where the eventuality pair are not simultaneous or symmetric, there is a very high proportion of pairs that

Table 4.3: Top ranked reciprocities from loveforums

Eventuality pairs extracted from	loveforums.com
$learned - to - accept(Entity_x, Entity_y)$	$tells(Entity_y, Entity_x)$
$like(Entity_x, Entity_y)$	$likes(Entity_y, Entity_x)$
$said(Entity_x, Entity_y)$	$met(Entity_y, Entity_x)$
$sees - out(Entity_x, Entity_y)$	$approached(Entity_y, Entity_x)$
$liked(Entity_x, Entity_y)$	$likes(Entity_y, Entity_x)$
$see(Entity_x, Entity_y)$	$see(Entity_y, Entity_x)$
$breakedupwith(Entity_x, Entity_y)$	$break - up - with(Entity_y, Entity_x)$
$to(Entity_x, Entity_y)$	$telling(Entity_y, Entity_x)$
$feels - about(Entity_x, Entity_y)$	$about(Entity_y, Entity_x)$
$still - tried - be - with(Entity_x, Entity_y)$	$rejects(Entity_y, Entity_x)$
$wishes(Entity_x, Entity_y)$	$liked(Entity_y, Entity_x)$
$calling(Entity_x, Entity_y)$	$call(Entity_y, Entity_x)$
$hurts(Entity_x, Entity_y)$	$hurt(Entity_y, Entity_x)$
$told(Entity_x, Entity_y)$	$lied - to(Entity_y, Entity_x)$
$digged - into(Entity_x, Entity_y)$	$love - with(Entity_y, Entity_x)$
$made(Entity_x, Entity_y)$	$knows - best(Entity_y, Entity_x)$
$told - up(Entity_x, Entity_y)$	$arent(Entity_y, Entity_x)$
$promises(Entity_x, Entity_y)$	$hate - saying(Entity_y, Entity_x)$
$into(Entity_x, Entity_y)$	$professed - loved(Entity_y, Entity_x)$
$tried - to - be(Entity_x, Entity_y)$	$misses(Entity_y, Entity_x)$
$turned - on(Entity_x, Entity_y)$	$loves(Entity_y, Entity_x)$
$is - who - is(Entity_x, Entity_y)$	$wanted - to - date(Entity_y, Entity_x)$
$hurted(Entity_x, Entity_y)$	$try - to - support(Entity_y, Entity_x)$
$respected(Entity_x, Entity_y)$	$chased - after(Entity_y, Entity_x)$
$call(Entity_x, Entity_y)$	$call(Entity_y, Entity_x)$

also have some causal meaning, such as $still - tried - be - with(Entity_x, Entity_y)$ and $rejects(Entity_y, Entity_x)$.

Chapter 5

Data and Preprocessing for Causal Modules

Here, we will discuss the data sources for the causal modules, their preparation, their annotation, and preprocessing steps, especially an important step in preprocessing necessary for complex causality ($\theta-$structure informed transformations).

5.1 Data and annotation for complex causal structures

For computing complex causal structures, as input to both the *adjunctive causal structure* module (Chapter 7), as well as to the *embedded causal structure module*

5.1.1 Selected data sets

The data for the complex causal detection and extraction modules consist of multiple parts of the BNC, which is an admixture of various genera, including the likes of news reports, parliamentary proceedings, magazine articles, memoranda, fiction, etc. Due to the computational complexity of the procedure, only a small part of the (2M clauses) is used as training data, of which only a small proportion exhibit counterfactual characteristics. We also used several novels with simple prose language (mostly from the Gutenberg collection), with works such as *The Great Gatsby*, *Pride and Prejudice*, *Little Women*, *Emma*, and *Lily of the Nile*, which are from the romantic fiction and historical novella genera. One part of BNC (due to its literary diversity) is used for training purposes; another part of BNC, and several novels are used for testing purposes. The original input was raw text that was approximately one line per sentence (not necessarily a clause). The part of BNC use for training had very different forms for *adjunctive* and *embedded* causal modules; but the testing data sources were the same, BNC-testing data contained 196314 lines, and novels set 129695 lines.

5.1.1.1 Data for adjoined structures

The adjoined causal structure extraction procedure utilizes the data-sets in unannotated form, after the preprocessing steps are complete. A portion (approximately 25% for the available BNC data) was used in training purposes, since it is a large, mixed genre document set that should contain all different structural types of causality in language. The testing is done on both another separate portion (also approximately one quarter) of BNC, as well as on the collection of novels. The length of the training and testing sets were selected to accommodate the practical memory footprint available on the machine for the experiment. Both the training as well as testing sets were pre-processed through the same steps for the adjunctive causal module. The training data was unannotated part of BNC, that contained 218440 lines (approximately that many sentences).

5.1.1.2 Data for embedded structures

Embedded causals are a highly specific type of constructions of high complexity, which will be detailed in Chapter 8, which occurs very infrequently in most genres, and vary from genre to genre in their prevalence. It also much more conducive for the embedded causals to occur, when the topic of discussion is highly logical and involves some type of argumentation (as opposed to casual conversation). And the complexity of the training data is important in determining whether we obtain sufficient number and variety of relevant substructures for the diffuse prototype; so this consideration for training data is paramount. The BNC corpus affords a good chance of finding significant number of them, since it is a mixed corpus with certain official sources such as parliamentary proceedings and news article, and should contain a significant portion that is well formed and structurally complex, and more logically based. So it is the best candidate to be training and testing data; we also tested the built diffuse prototype against testing data from the novels corpus for comparison.

The data used in building this embedded causal detection and extraction module consist of multiple parts of the British National Corpus (BNC) with an admixture of various genera; and several novels with simple prose language. For one, we would like to have a procedure that works for non-domain specific text; so the mixed genre corpus such as BNC is a good choice in that regard. We also need select a small number of training samples from texts that tend to contain relatively high frequency of them. Several genres in BNC (news articles, journals and other periodicals, academic texts, etc) would be highly formal and likely to contain complex logical arguments resulting in causal chains; also certain other genres that are more informal but also may contain complex arguments (parliamentary proceedings). So the BNC would be a good source to select structures that describe complex causal chains, especially when we focus on sections of certain genres.

The training data for *embedded causal* are 500 samples selected from The labeling of the training data is necessarily on heavily pre-processed data. These must be tagged and parsed, some tree transformations detected and reconstructed, including those that result from gapping and extraposition, and separated into individual semantic frames. This is due to how embedded causative structures are defined, largely on structures present at LF. Each sample of the training data's annotation was very simple, only needing to label the inner-most embedded clause with an extra *'E'* on the tag of its topmost syntactic node, forming ES/ESBAR/ESINV/etc; the reasoning behind this will be discussed when embedded causal structures are defined and characterized in Section 8.1.1. Thus this requires considerable less effort in annotation than the eventual annotation of testing results; and a total of 500 samples are used from BNC as training for embedded causality.

5.1.2 Data annotation

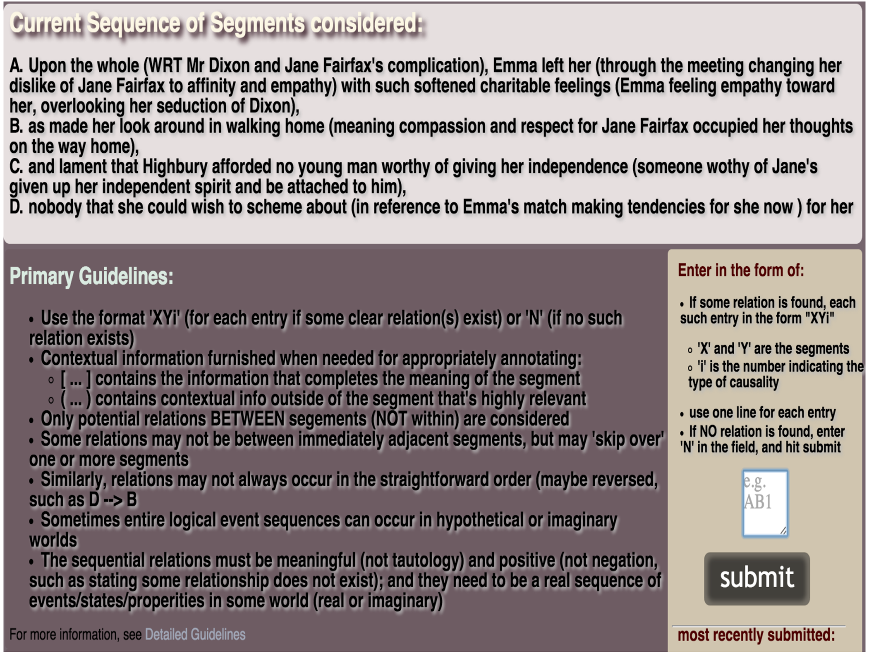
Both adjunctive and embedded causal modules are fundamentally procedure that produces a relative score for each candidate pair (adjunctive) and sequence (embedded) of clauses. For each of these modules, since complex causal structures are a small proportion of a large data-set, if the ranking is successful, the true positive samples would be preferentially concentrated near the top of the ranking.

Since complex causal structures are scarce forms in a large corpus, randomly selecting samples to annotate pre-ranking is not practically feasible. Any samples that are randomly selected adjacent pairs, or embedded clausal-complexes, would be highly unlikely to be genuinely causal, and unlikely to be ranked highly by either procedure. Additionally, since these are deep semantic concepts that are difficult and time consuming to annotate, annotation of the entire corpus or even a significant portion of it (very little of which would be causal), would be impractical. This would be the case when locating samples with a property where said samples are very sparsely distributed; or when dealing with a ranking of the property that has a long-tailed distribution.

For the primary instructions of the annotation scheme, a set of strict guidelines were formulated on marking each sample as complex causality or not. The annotation is presented as a pair, or a sequence of several utterances from the data-source, and the annotator is also provided with any contextual information that might be relevant. This contextual information is gleaned by reading the adjacent several paragraphs of text in both directions of the textual context, and occasionally search for certain events, entities, unclear terms and associations, in the remainder of the corpus (BNC) or in the remainder of the document (for each novel). And occasionally, especially for the BNC testing corpora, some terminology or event (for news stories or parliamentary proceedings) further consultations with sources external to the data-corpora were needed.

The original annotation consumes 2-30 minutes per sample, and between 150 and 200 hours per data-set per module; for further annotations thereafter, the original annotator has noted down the important contextual information, so it may be less time consuming. Given the contextual nature of determining deeply semantic causal relations, it would not be possible to determine all samples presented to the annotators with high confidence. So the annotators has the option to annotate *'Y'* or *'N'* for each sample where he/she has reasonable confidence, and provide an *indeterminable* annotation *'U'* when the confidence is very low. We have a sample page in Figure 5.1 as the basic format that is seen by the annotator, with some contextual information (in parentheses) already recorded by the original annotator.

Figure 5.1: basic annotation format



Current Sequence of Segments considered:

A. Upon the whole (WRT Mr Dixon and Jane Fairfax's complication), Emma left her (through the meeting changing her dislike of Jane Fairfax to affinity and empathy) with such softened charitable feelings (Emma feeling empathy toward her, overlooking her seduction of Dixon),
B. as made her look around in walking home (meaning compassion and respect for Jane Fairfax occupied her thoughts on the way home),
C. and lament that Highbury afforded no young man worthy of giving her independence (someone wothy of Jane's given up her independent spirit and be attached to him),
D. nobody that she could wish to scheme about (in reference to Emma's match making tendencies for she now) for her

Primary Guidelines:

- Use the format 'XYi' (for each entry if some clear relation(s) exist) or 'N' (if no such relation exists)
- Contextual information furnished when needed for appropriately annotating:
  - [...] contains the information that completes the meaning of the segment
  - (...) contains contextual info outside of the segment that's highly relevant
- Only potential relations BETWEEN segements (NOT within) are considered
- Some relations may not be between immediately adjacent segments, but may 'skip over' one or more segments
- Similarly, relations may not always occur in the straightforward order (maybe reversed, such as D --> B
- Sometimes entire logical event sequences can occur in hypothetical or imaginary worlds
- The sequential relations must be meaningful (not tautology) and positive (not negation, such as stating some relationship does not exist); and they need to be a real sequence of events/states/properties in some world (real or imaginary)

For more information, see Detailed Guidelines

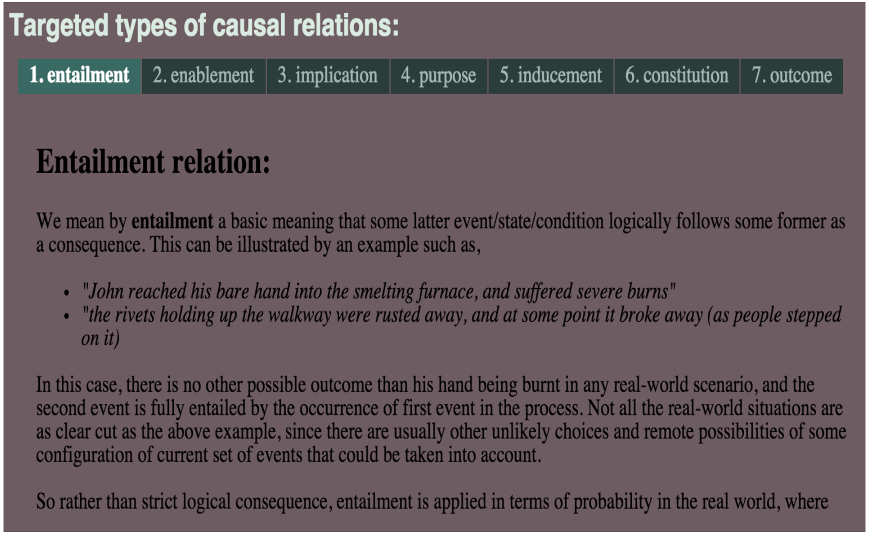
Enter in the form of:

- If some relation is found, each such entry in the form "XYi"
  - 'X' and 'Y' are the segments
  - 'i' is the number indicating the type of causality
- use one line for each entry
- If NO relation is found, enter 'N' in the field, and hit submit

most recently submitted:

Causality is a concept that has many disagreements among academics as well as individual speakers alike. We also, as part of the guideline, in order to further enforce a somewhat uniform idea of what consists causality, provided a categorization scheme, described in detail in Section 9.4.1 in order to help the annotator to understand what we mean by causal structures; the figure 5.2 would be the format that the annotator uses as a prompt. This is also a part of the annotation guidelines that are followed during the process.

Figure 5.2: class based prompts



Targeted types of causal relations:

Entailment relation:

We mean by **entailment** a basic meaning that some latter event/state/condition logically follows some former as a consequence. This can be illustrated by an example such as,

- *"John reached his bare hand into the smelting furnace, and suffered severe burns"*
- *"the rivets holding up the walkway were rusted away, and at some point it broke away (as people stepped on it)*

In this case, there is no other possible outcome than his hand being burnt in any real-world scenario, and the second event is fully entailed by the occurrence of first event in the process. Not all the real-world situations are as clear cut as the above example, since there are usually other unlikely choices and remote possibilities of some configuration of current set of events that could be taken into account.

So rather than strict logical consequence, entailment is applied in terms of probability in the real world, where

The detailed instructions that we utilized in the annotation scheme may be found in the Appendix in Chapter 10.4 (Appendix A).

5.1.3 Annotation Assessment

The majority of studies in computational semantics classification tasks use an evaluation system that looks at the precision and recall of an entire testing data set. In binary classification, this works well when there is a significant amount of both positive and negative samples, where given a very large data-set, some random selection of samples would produce a smaller set of relatively similar level of positive and negative samples. Most binary classification problems in computational linguistics also generally assumes a Zipfian distribution of the property investigated in the data-set (Zipf 1949; Powers 1998; Tullo & Hurford 2003).

The algorithms that we propose are appropriately used as ranking procedures over testing samples. For both adjoined causal and embedded causal structures, since these are both highly complex structures of highly specific semantic class. The adjunctive causals are some implicit causal relation between events corresponding to two clauses expressed by locating them adjacently in speech; and embedded causality are usually utilized to express long chains of causation that are otherwise difficult to detect. Given their sparsity in each dataset, a randomly drawn small subset of testing samples is not likely to contain a significant amount of either semantic type.

Both adjunctive and embedded causal structures also cannot be described with a single or a few features in a parametric model. And given their complexities in terms of feature spaces that would describe them (detailed in sections 7.1 and 8.1), it is not likely that we would be able to find all of the discriminating features with our first study; thus a less clean long-tailed distribution, where a significant portion of positive samples are not highly ranked. So given these factors, computing recall of all testing samples, or a randomly selected subset, is practically infeasible for this class of problems with complex structures and multi-modal semantics; so we will adopt a quantile based evaluation procedure for the rankings in these two complex causal modules.

5.1.4 Morphosyntactic preprocessing

The morphosyntactic preprocessing are done using standard toolkits. The basic syntactic tagging and parsing are performed using the shift-reduce dependency parser from University of Tokyo NLP's Tsujii Lab (described in Sagae & Tsujii 2007). There is also a stage of named entity recognition, for the following frame structure-informed post-syntactic processing is performed taking NER information into account as well; that is a module from the toolkit of the Stanford NLP. (a study of the Stanford NLP group, described in Manning et. al. 2014)

5.2 Semantic frame theory

Extraction of semantic frames is an essential preprocessing step for latter components of the system. The transformations in this step enables us to provide structurally similar samples for samples that have disparate surface orders but similar deep semantic structures to be treated as largely identical samples. This process is by and large based on formal linguistic observations of how semantic frame structures are related to surface constituent orders.

5.2.1 $\Theta-$structure in language

Thematic structure, and its successor, $\theta-$theory, is developed in connection with Government and Binding as well as lexical semantics, especially that of predicate verbs. It is designed to serve as a intermediary for processing information between the deepest layer of syntactic representation (often LF, depending on morphosyntactic framework) and compositional semantics. The concept of $\theta-$*criterion* allows the semantics of a frame to inform the structure in syntax (Chomsky, 1981), such that $\forall A_i \in A, \exists!\ \rho_j \in \rho^\theta\ [\varsigma(\rho_j) = A_i]$, meaning that there is a exact one to one correspondence between the set of $\theta-$roles ρ^θ and the set of arguments in syntax A that symbolize them (with provisions for some instances of under-specifications, which could be viewed as lexical variability for predicates).

Information about the semantic structure of the clause, within this framework, originates in the predicate, and proceeds as scaffold for the remaining elements in the syntactic structure. A set of $\theta-$roles are specified according to the lexical information from the predicate verb, and some anaphora binding occurs between one of the roles in the argument list of the predicate antecedent, to the syntactic positions in SPEC-IP or within the VP, through coindexation (Williams, 1989 / 1994). Every $\theta-$role furnished in the argument structure by the lexical predicate must be assigned, so as to provide the required structure for compositional semantics.

The motivation is to provide a fixed set of $\theta-$role labels for structures of $\theta-$frames of all predicate-types. The fixed set is not generally agreed on by linguists, but usually contains a core set that almost everyone uses, such as *agent*, *cause*, *source*, *experiencer*, *theme*, *patient*, *goal*, *percept/phenomenon*, *recipient*, *instrument*, *location*, *time*; these are augmented with additional ones depending on the emphasis of granularity of different parts of the semantic-space for arguments in each classification system. Ideally the classification system should contain categories that are each relevant to both semantic notions as well as to some distinction that they make at LF.

Alternatively, the classification can be effected by modeling each $\theta-$role as a feature vector, as a more paradigmatic approach. There, a relatively limited set of binary/ternary features exists for each role termed a feature cluster (Rozwadowska, 1987; Reinhart, 2000

/ 2002), and the values of *+ / - / ?* (*'?' is don't care that is necessary in some systems*) at each of these feature positions combinatorially determines the role's identity. The most frequently utilized features in these cases are [$\pm$*cause*], and [$\pm$*mental/sentience*] (animacy or ability for volition) in Reinhart's system, and viewing each as ternary features makes 9 combinatorial possibilities. For instance, the *agent* $\theta-$role is regarded as having a [$+cause, +mental$], having a role in precipitating the event, where as the *experiencer* $\theta-$role has [$-cause, +mental$], still sentient but without such a direct role, or the *subject matter* $\theta-$role has [$?cause, -mental$], which is irrelevant to any causal relation, and cannot really behave as a sentient or volitional being in the relation. Other features that could be used to define $\theta-$roles include [$\pm$*change*], [$\pm$*immediacy*], etc; these are useful with increased granularity in roles, such as [$\pm$*immediacy*] differentiating between *instrument* and *force* in their actual or logical distance to the caused event.

This theoretical concept serves as a prototype for the further formulation of semantic frames on which computational work is often based, and this system informs the cognitive linguistic concept of *semantic frame*s, (which differs from the concept of the same name in *categorial grammar*). The system of classification for $\theta-$roles can be highly systematic and theory driven; the cognitive semantic frame system derives many of the same classifications, but often with far finer distinctions. The organization of all roles are arranged in a hierarchical manner, with the top level of the hierarchy close resembling the roles espoused in $\theta-$theory, but with the leaf levels having far smaller granularity of semantic classes, such as roles in *judgment*, *hearer*, *adornment*, etc. (Fillmore & Baker, 2009) These normally are represented as a type of structure, each instance with some fixed number of slots, and a set of constraints associated with each slot as to the subset of all arguments that can fill it.

5.2.2 Matter of representation

For most logical representations of this class of propositions about events, termed subatomic semantics (Parsons, 1985 / 1990), a logical representation containing the predicate, several types of arguments, other components, along with an event variable that specifies the real occurrence of e_i. The most prevalent form of this representation is a conjunctive formula, with the predicate and the obligatory arguments in one factor, and the rest separated individually: $\phi(A_0,\ A_1,\ ...\ A_k, e_i) \wedge X_1(e_i) \wedge X_2(e_i) \wedge\ \cdots$, given the predicate $\phi()$, the obligatory arguments $A_0, ...A_k$, and the rest of the components $X_1,\ ....$ In this logical form, only the predicate and the obligatory arguments must be present in the logical form to complete the meaning of the event type, the rest may be present when e_i is instantiated as a specific occurrence. The most likely logical structure can be extrapolated for each clause, given the parsed surface form, the location of the main content V, as well as a list of candidate adicity structures for that V form; consider the following:

1. $[_{NP}$ *Ian* $]$ ***hated*** $[_{NP}$ *Jane* $]$ $[_{SBAR}$ *for being a tattle tale* $]$
2. $[_{NP}$ *Ian* $]$ ***hated*** $[_{SBAR}$ *that Jane is a tattle tale* $]$ $[_{PP}$ *on Tuesday* $]$

In the first example, we find the content V *'hate'* with a selection of adicity structures {SV, SVO, SVS}. Since the surface structure indicates that it likely has at least a direct object, the monadic SV structure is unlikely; and given the order of the candidate arguments *'Jane'* and *'for being a tattle tale'*, and given *'hate'* is at most dyadic, the NP *'Jane'* is the most likely the second candidate, and adicity is likely SVO. This can be represented as $\phi_{hate}(ARG0_{Ian}, ARG1_{Jane}, e_i) \wedge for\_being\_tattle\_tale(e_i)\ |\ e_i \in \mathcal{E}$. In the second example, SV adicity can be similarly eliminated. But SVO is no longer feasible, since there is no NP complement, so SVS is selected. This can be represented as $\phi_{hate}(ARG0_{Ian}, ARG1_{that\_Jane\_....})$. Here, by $ARGn|n \in \mathbb{N}$, we mean their D-structure positions. There is no easy way to merge non-obligatory components into $\phi(A_0,\ A_1,\ ....)$, at least not without resorting to polymorphism of $\phi(\cdot)$. This indicates that obligatory and non-obligatory arguments have fundamentally different logical functions in a frame, and should be treated differently when used to distinguish an event token. So it is reasonable, according to semantic theory, to include the obligatory arguments, but leave out the remaining components in measuring causality between events.

5.3 Transformation mechanism

We hand built a system according to high-frequency rules of transformation, rooted in formal linguistic principles at the syntax-semantic interface. Although this process would be effective for the relatively common cases, but not for all possibilities of surface orders. This provides the preprocessing for the latter modules to enhance the abilities of adjunctive causality (in Chapter 7) and embedded causality (in Chapter 8) recognize sequential and structural patterns indicative of causality. For any further analysis of complex semantic events in the data, it is highly beneficial to have a transformed version of the lexico-syntactic structure that correspond well to the forms that are closest to a fully semantic representation. And this process works well when informed by the representations at a $\theta-$level.

5.3.1 Purpose of frame-informed transformation

The transformations performed in this module allows us to put forms with superficially differing orders but deep semantically similar content into the same form, which then can be more readily be used by causality modules to locate common patterns within. A simple example would be the difference in surface order between a L-topicalized and a R-extraposed form of the same semantic frame, such as:

1. With Tuesday came the agreeable prospect of seeing him again, and for a longer time than hitherto;
2. The agreeable prospect of seeing him again came, and for a longer time than hitherto, on Tuesday

Using simple rule-based transformations of that takes into account of both topicalization and extraposition rules, we can see that the above two semantic frames are equivalent in a frame-theoretic sense, such that $\phi_{come}(ARG_{propsect\_of\_seeing\_him}, ARG_{for\_a\_longer\_time}, ARG_{on\_Tuesday})$ Another simple example would be the completion of an expression with ellipses. In these cases, some pair of full and elliptical expressions are adjacent in the same coordinated clause, where the full expression can be used to complete the semantic frame of the elliptical one, such as:

1. For Mrs. Weston there was nothing to be done;

2. For Harriet [there was] every thing [to be done]

Using rules surrounding ellipses in language, we can observe that the second frame above, when occurring immediately following the first, would have an existential meaning, as well as a similar small clause *'to be done'* shared with the first; thus this would allow us to exploit an otherwise unusable frame. These and a number of other types of necessary common transformations will be discussed in detail in the following sections. In addition, as we see in Chapters 8 and 7, it is also helpful to distinguish frame essential and non-essential components in a surface form, and it is necessary to preserve some context free or DAG structure for those discrimination tasks.

5.3.2 Transformation and automata

For any unique logical proposition, there exists a number of possible surface strings, where any single form has very small probability of being replicated in moderate sized dataset; To obtain any meaningful measure of frequency, only frame components that have a bearing on the SN structure should be considered when providing a specific ordering of arguments, the remaining, largely adjunctive constituents can be ordered in an unspecified order after the essential components. The most important frame components include the *predicate* which is always a part of the semantic core; zero or one *external argument* which take $\theta-$roles such as *cause*, *agent*, *origin*; zero or more *internal arguments* that is capable of taking on a variety of semantic roles; zero or more *oblique arguments* that fulfill optional roles such as *locative*, *path*, *temporal*; as well as adverbials, sentential connectives, etc. In assessing the probability of each event, the relevant components are restricted to the predicate and its obligatory arguments.

5.3.2.1 Branching pushdown automata

Given the scope and variety of transformational rules that must be accounted for, it is far more efficient to have a unified algorithmic-framework that can be deployed to simulate a variety of different syntacto-semantic movements. We can employ a single computational model, with a common set of associated algorithms to effect all movement types, which can be parameterized and further modified to accommodate each class of transformation. For this we adapted an existing formal model in automata theory, with additional features and mechanism to accommodate multiple types of transformations. Each component is extracted using an algorithm designed to extract distinct nodes from a context free structure, This mechanism is able to record distinct machine states depending on path reached from the tree root, and at the same time could perform some quantification of properties at each branching point.

The ability to reason on ∃-quantification over some property of a subtree (T_s) given some precondition being met are especially important, where $dsc(T) :=$ is a descendent function, and $dsc^d(T) :=$ includes descendents down to $d-$levels below the node T. $\varsigma(T_s) :=$ provides the relevant symbol for a subtree, and $\mathcal{M}(\sigma_i) :=$ provides the consequence for the premise σ_i such that:

$$\begin{cases} \forall T_s \in dsc(T) \ [\ \exists T_t \in dsc(T_s) \wedge \varsigma(T_t)\ ] \\ \quad \mid \mathfrak{M}(\varsigma(T_s)) = \varsigma(T_t) \end{cases} \tag{5.1}$$

This type of property can readily be seen in example such as the observation of constituent entails the observation of its head, or that the observation of a predicate (of a certain arity) entails the observation of an internal argument. Every context free structure, when T_s is at a phrasal constituent (XP) level $\exists \varsigma(T_s) \in \mathfrak{M}$ such that it can be tested for one or more $\varsigma(T_t)|dcs(T_s)$; thus $\langle \varsigma(T_s), \varsigma(T_t)$ and be viewed as a pair of properties, where the first can be detected, prompting a *push* action of some *requirement*, which at a later point can be undone with a *pop* action, with the detection of the $\varsigma(T_t)$ property.

Just due to checking the constituency type (XP), these properties are prevalent within any context free structure, and should be built into any detection mechanism at a default level. There are also additional, specialized properties that would need to be detected for locating movements; these special properties and their implementations will be discussed in detail in later sections. Incidentally, the design is similar to a *branching pushdown tree automaton* (BPTA) (Schimpf & Gallier, 1985; Alur & Choudhuri, 2006) performing DFS over a defined region of the tree of e_i. Thus we will use its formalism, for a different purpose than BPTA's original design, with appropriate modifications, to describe our system. The module searches for each of these components in a defined sequence, such that the locations of the previously found elements are used as delimiters for the subsequent DFSs. BPTA is a modification of *pushdown tree automaton* (PTA) with increased expressiveness, and can efficiently process structures that require both branching and pushdown properties.

In other words, BPTA provides simultaneous confirmation of two types of properties for a context-free structure T: one is for all paths from the root r^T of T to each leaf, some existential property with regard to node-label holds, so that

$$\forall P_i = \langle r^T \; ... \; v_i \rangle \, \exists v_j \in P_i[\varsigma(v_j) \in \Sigma_s]$$

where v_i is a leaf node, $\varsigma(v_j)$ gives the label at v_j, and $\Sigma_s \subseteq \Sigma$ is the defined property for that machine; and the other is for the entire set of tree nodes of some subtree T of T, some constant count property of node labels holds, such that:

$$\sum_{v_i \in V(T_u)} \mathbb{1}(\varsigma(v_i) \in \Sigma_s)$$

5.3.2.2 Basic BPTA operations

Trees here have a maximal arity of 4, determined by max branch factor. We give a description of the algorithms for trees with an arity of 2, but can easily be extended to higher arities. The set trees $\mathcal{T}$ is:

$$T \in \mathcal{T} \begin{cases} T^0 = () & \textit{the empty tree} \\ a(T_1, T_2) \;\Big|\; T_1, T_2 \in \mathcal{T} & \end{cases} \tag{5.2}$$

The default PTA A is defined as a seven-tuple. (Guessarian, 1983; Schimpf & Gallier, 1985) Given that the set of states Q of the finite control, the set of input symbols Σ, the stack symbols Γ, some initial configuration$\langle q_0, Z \rangle$, the current stack w, then A is defined:

$$A = \begin{cases} Q \\ \Sigma = \{\sigma_a, \sigma_b, \ ....\} \\ \Gamma = \{\gamma_1, \gamma_2, \ ....\} \\ q_0 \in Q \\ Z \in \Gamma^* \ \Big| \ A_0 = \langle q_0, Z\rangle \\ \delta = \delta_{push} \cup \delta_{pop} \cup \delta_\varepsilon \cup \delta_\sigma \\ F \subseteq Q \end{cases} \tag{5.3}$$

$$\delta = \begin{cases} \delta_{push} \subseteq \Big\{\langle q, w\rangle \longrightarrow \langle q', \gamma.w\rangle \Big| q, q' \in Q, \gamma \in \Gamma\Big\} \\ \delta_{pop} \subseteq \Big\{\langle q, \gamma.w\rangle \longrightarrow \langle q', w\rangle \Big| q, q' \in Q, \gamma \in \Gamma\Big\} \\ \delta_\varepsilon \subseteq \Big\{q \xrightarrow{\epsilon} q' \Big| q, q' \in Q\Big\} \\ \delta_\sigma \subseteq \Big\{q \xrightarrow{\sigma} q_1.q_2 \Big| q, q_1, q_2 \in Q, \sigma \in \Sigma\Big\} \end{cases} \tag{5.4}$$

When the input tree T is accepted, $\exists$ a run of the machine resulting in ϵ as the remaining input and the machine reaches some configuration $\langle q_f, Z\rangle | q_f \in F$. The machine can also start at configurations other than $\langle q_0, Z\rangle$; machines with alternate start configuration is denoted as $A_{\langle q,w\rangle}$; for clarity, we will denote the default starting configuration as $A_{\langle q_0,Z\rangle}$. $\mathcal{F}(A_{\langle q_i,w_l\rangle}, T)$ occurs when the input T is accepted with A starting in state q_i and with w_l on the stack. The algorithm that produces the semantics of a PTA is as follows, with the input T, if there exists a run at $A_{\langle g_0,Z\rangle}$ such that:

$$\mathcal{F}(A_{\langle q,w\rangle}, T) \begin{cases} \langle q, Z\rangle \Big| q \in F, T = T^0 \\ (q \xrightarrow{\epsilon} q') \in \delta \wedge \mathcal{F}(A_{\langle q',w\rangle}, T) \\ T = a(T_1, T_2) \ \Big| \ q \xrightarrow{\sigma} (q_1, q_2), \\ \qquad \mathcal{F}(A_{\langle q_1,w\rangle}, T_1), \mathcal{F}(A_{\langle q_2,w\rangle}, T_2) \\ q \longrightarrow (q', push(u)) \ \Big| \ \mathcal{F}(A_{\langle q',u.w\rangle}, T) \\ q \longrightarrow (q', pop(u)) \ \Big| \ w = u.v, \ \mathcal{F}(A_{\langle q',v\rangle}, T) \end{cases} \tag{5.5}$$

5.3.2.3 Modification and extension of BPTA mechanisms

The branching-PTA B offers greater expressive power for the discrimination of tree structures by making a pair of modifications to $\delta-$function, and allow us to examine the properties of the subtrees of T in a richer way; such as stipulating that for some T^s subtree it must have at minimum a k leaves of a certain label to be classified into a certain semantic class; or such as stipulating that T^s must contain a previously classified subtree $T^t \in dsc(T^s)$ in addition to properties that can be elucidated by a PTA. Given a series (constant number) of these machines, potentially tests both $\exists$ and $\forall$ type properties can be recognized at different points in $V(T)$. The BPTA has a slightly augmented notion of execution, where a successful run is denoted by $\mathcal{F}'(B_{\langle q,Z\rangle}, \alpha_\omega, T)$, where B is the machine, and $\alpha_\omega \in Q*$ is the ordered set of leaf nodes of T with some possible DF traversal order ω of T.

One such mechanism involves the use of a set of count constraints over such α_ω. In Alur & Chaudhury (2006), the implementation of a constraint is a single multiset $\mathbb{U}$, where $\mathbb{U}$ contains both the desired number of each $q_i \in \alpha_\omega$ as well as the indicators $\{q^*|q \in Q\}$ of whether each is an equivalence or non-equivalence ($\geq$) relation. While typologically simpler and potentially saving space when q types are sparse, this is difficult to implement and less efficient to process. We implemented each constraint as a double $\chi = \{\langle \upsilon_i, \rho_i\rangle \ |q_i \in Q\}|\chi| = |Q|$, each υ_i is the count of the i^{th} type $\in Q$, and each $\rho_j = 0$ when the relation in the constraint for q_j is equivalence, and $\rho_j = 1$ when the relation is *'at least'*. We also have a lookup for the symbol's position in $\mathcal{M}$, such that the symbol represented by parameters $\langle \upsilon_j, \rho_j\rangle$ is $\mathcal{M}_j$. This allows the machine, at every node, to reason about the findings in the node's subtree with a precise numerical equivalence of the value of some feature (such as number of labels in the subtree), or with a $\geq$ relation to the value, including the semantics of $\exists$.

Distinct constraints can be related to one another with a semi-lattice structure with $\preccurlyeq$ and $\not\succcurlyeq$. We will use $\mu(\chi, q_i)$ to denote the count of q_i within the constraint χ, and use $Q(\chi) = \{q_i|\langle \upsilon_i, \rho_i\rangle \in \chi\}$. For any pair of constraints $\chi = \{\langle \upsilon_i, \rho_i\rangle|q_i \in Q\}, \chi' = \{\langle \upsilon'_i, \rho'_i\rangle|q_i \in Q\}$, $\chi \preccurlyeq_Q \chi' \iff \Big(\forall q_i \in Q[\rho_i \rightarrow \upsilon_i \leq \upsilon'_i]\Big) \bigwedge \Big(\forall q_i \in Q[\neg\rho_i \rightarrow \upsilon'_i \leq \upsilon'_i]\Big)$. In terms of the summation operation at tree nodes, in the original implementation of the BPTA, the entire $\chi = Q_\chi \cup Q^*_\chi$ (the set of all q and q^* elements in χ) such that Q_χ is a multi-set where $\forall q_i \in Q_\chi \rightarrow q_i \in Q$, and the number of q_i present $\in Q_\chi$ determines $\mu(\chi, q_i)$; and Q^*_χ is a uni-set, and $\forall q_i \in Q, q^*_i \rightarrow q_i \in Q_\chi$, where the presence of q^*_i determines that the count constraint requires and inequality $\geq$ instead of $=$. Summation in that case is simply a union of χ, χ', where $sum(\chi, \chi') = (Q_\chi \cup_{multi} Q_{\chi'}) \cup_{multi} (Q^*_\chi \cup_{uni} Q^*_{\chi'})$ Similarly, we can also define a subtraction operation on $sub(\chi, \chi')$, where $sub(\chi, \chi') = (Q_\chi \setminus_{multi} Q_{\chi'}) \setminus_{multi} (Q^*_\chi \setminus_{uni} Q^*_{\chi'})$ Using the alternative implementation of $\chi = \{\langle \upsilon_i, \rho_i \rangle \; | q_i \in Q\}$, we can define the following to be the summation and subtraction:

$$\begin{cases} sum(\chi, \chi') = \Big\{ \langle \upsilon_i + \upsilon'_i, \rho_i \vee \rho'_i \rangle | q_i \in Q \Big\} \\ sub(\chi, \chi') = \Big\{ \langle max(\upsilon_i - \upsilon'_i, 0), \rho_i \wedge \rho'_i \rangle \; | q_i \in Q \Big\} \end{cases} \tag{5.6}$$

The other modification is the capability of *substitution* during run, given that $\alpha_{[i]}$ denotes the i^{th} element of the string α, $\alpha_{[i]}(\alpha')$ denotes that the new string α' substitutes i^{th} element in α. and that $T \circ_i T'$ indicates that the i^{th} leaf of T is substituted with the tree T'. In a run of BPTA, if $\mathcal{F}'(B_{\langle q_0, Z \rangle}, \alpha, T)$ and $\mathcal{F}'(B_{\langle q'_0, Z \rangle}, \alpha', T')$, as well as $\alpha_{[i]} = q'_0$, then $\mathcal{F}'(B_{\langle q_0, Z \rangle}, \alpha_{[i]}(\alpha'), T \circ_i T')$. This capability allows certain limited transformations on the original tree, without rearranging the underlying data structure.

5.3.2.4 Additions to symbol equivalence function

The symbol equivalence function of the BPTA requires some basic features, such as being able to selectively equate either tag or token to the reference set Σ_s, as selected by a parameter $\in \Psi_s$, a control parameter set associated with each Σ_s; to be able to equate to a lemmatized version of any token, determined by another parameter $\in \Psi_s$. The equivalence function of can also be designed to evaluate against additional conditions. These would be useful when some complex properties of some region of the tree can be detected using one or more base-line BPTAs, but more efficiency can be achieved through minor modifications of the equivalence function. We will use the notation of $\cdot^{+\Psi_s}$ to denote these control parameters in the execution $\mathcal{F}'(A_{\langle q_0, Z \rangle}, \alpha^0, \Sigma_s^{+\Psi_s}, T)$, where we allow the initial string in the machine to be $\alpha^0 = [\;]$.

Some modifications also will be more convenient when we are able to directly exclude some subtrees from further consideration, those that have root with some $\gamma_i \in \Gamma_s$ as its symbol, in the process of detecting $\sigma_j \in \Sigma_s$. This can be inefficiently simulated by using two stages of BPTA, first locating all of the subtrees with the set $\Sigma \setminus (\Gamma_s \cup \Sigma_s)$ with the execution $\mathcal{F}'(A_{\langle q_0, Z\rangle}, \zeta = (\Sigma \setminus (\Gamma_s \cup \Sigma_s))^{+\Psi_s}, T)$, and at each root r^{T_s} of such a subtree T_s, the execution is continued with different Σ_s and different control parameters, $\mathcal{F}'(A_{\langle q_0, Z\rangle}, \alpha^0, \zeta' = \Sigma_s^{+\Psi_s}, T^s)$. So each execution would also be augmented in this way as $\mathcal{F}'(A_{\langle q_0, Z\rangle}, \alpha^0, \zeta, \eta, T)$ where η is the set of exclusions and associated control parameters Φ_s.

The most prominently used additional features include the following, each being available to be selected, or deselected depending on the value of the corresponding parameter within Ψ_s. The notation $N^{+/-}(\cdot)$ is the standard neighbor function in a DAG. ❶ A modified BPTA can be designed to exclude subtrees with roots having symbol $\sigma_i \in \Sigma_s$ instead of detecting them ($|\Sigma|$ is constant). We will express execution with this modified equivalence function as $\mathcal{F}^{\mathcal{X}}(A_{\langle q_0, Z\rangle}, \alpha^0, \zeta, \eta, T)$. This is equivalent to the unmodified equivalence function using $\zeta = (\Sigma \setminus \Sigma_s)^{+\Psi_s}$

❷ A modified BPTA can exclude a subset of symbols on the path to detecting Σ_s, thus locating all $\varsigma(v_i) \in \Sigma_s$ having excluded certain subtrees, which is the aforementioned $\eta = \Gamma_s^{+\Phi_s}$. Since this mode is frequently required, we will assume the baseline $\mathcal{F}'$ has this functionality available. And thus the morphology of the function becomes $\mathcal{F}'(A_{\langle q_0, Z\rangle}, \alpha^0, \Sigma_s^{+\Psi_s}, \Gamma_s^{+\Phi_s}, T)$.

❸ Some modified BPTA detects $v_i \in V(T), \varsigma(v_i) \in \Sigma_s$, $\forall v_j \in N_T^+(v_i)[\varsigma(v_j) \notin \Sigma_s]$, which detects with the additional condition that the detected node v_i disagrees with the symbols of all of its children. This would be useful in instances of a chain of a certain label (say VP), where we need to locate the deepest node with that symbol on the chain of symbols of all members of that same set on path from r^T to some leaf. We express this modification as $\mathcal{F}^{\checkmark}(A_{\langle q_0, Z\rangle}, \alpha^0, \zeta, \eta, T)$ which can easily be simulated with a two stage execution of $q \;\Big|\; \mathrm{o}\Big(\mathcal{F}'(A_{\langle q_0, Z\rangle}, \alpha^0, \Sigma_2^{+\Psi_2}, \Gamma_2^{+\Phi_2}, T_q \in \mathcal{F}'(A_{\langle q_0, Z\rangle}, \alpha^0, \Sigma_1^{+\Psi_1}, \Gamma_1^{+\Phi_1}, T))\Big)$, where symbol sets have the relations $\Sigma_1 = \Sigma_2 = \Sigma \setminus \Gamma_2$. (we will use $\imath(\mathcal{F}'(\cdot))$ to indicate that the return from $\mathcal{F}'(\cdot)$ contains a set of at least 1 member, and $\mathrm{o}(\mathcal{F}'(\cdot))$ to indicate an empty set in return.)

❹ Having a mechanism to detect $\varsigma(v_i) \in \Sigma_s$, but to pass on sisters $\{v_j | v_i, v_i \in N_T^+(v_h), v_j \neq v_i\}$ to the next BPTA. This is used for a command relation, and allows the continued execution of the $m+1^{th}$ stage of the BPTA on the sister nodes of the detected node during stage m. We express this modification as $\mathcal{F}^{\int}(A_{\langle q_0,Z\rangle}, \alpha^0, \zeta, \eta, T)$, which can be simulated with a three stage execution of $\mathcal{F}'(A_{\langle q_0,A\rangle}, \Sigma_3^{+\Psi_3}, \varnothing^{+\Phi_3}, T_s \in \imath\Big(\mathcal{F}'(A_{\langle q_0,Z\rangle}, \alpha^0, \Sigma_2^{+\Psi_2}, \Gamma_2^{+\Phi_2}, T_q \in \mathcal{F}'(A_{\langle q_0,Z\rangle}, \alpha^0, \Sigma_1^{+\Psi_1}, \Gamma_1^{+\Phi_1}, T))\Big)$, where symbol sets have the relations $\Sigma_2 = \Sigma \setminus \Gamma_2$ and $\Sigma_3 = \Sigma \setminus \Sigma_2$.

❺ is similar to the previous one, instead of detecting any $\varsigma(v_i) \in \Sigma_s$, it detects its terminal children's, $v_j \in N_T^+(v_i) \wedge |N_T^+(v_j)| = 0$. We express this modification as $\mathcal{F}^{\mathcal{N}}(A_{\langle q_0,Z\rangle}, \alpha^0, \zeta, \eta, T)$, which would be useful in cases such as: a long continuous path of VPs, needing to examine the terminal child of each VP in chain. This can be simulated with a three stage execution of $\mathcal{F}^{\int}(A_{\langle q_0,A\rangle}, \Sigma_2^{+\Psi_2}, \varnothing^{+\Phi_2}, T_s \in (N_T^+(r^T) \cap \mathcal{F}'(A_{\langle q_0,Z\rangle}, \alpha^0, \Sigma_1^{+\Psi_1}, \Gamma_1^{+\Phi_1}, T))\ )$,

❻ there are certain types of transformational-grammar structures that have explicit directionality built into their definition, and preserving the distinction of such directionality has advantages in frame semantics, down the line in the system, for future social network analysis. Some common types of such structures include L-topicalization and R-extraposition. Note that movement such as topicalization and extraposition often have an extra-frame semantic effect on the interpretation of the form, especially at a discourse level, and in pragmatics; but we will ignore such additional semantics effects, and leave any of their impact on causality to be explore by the specific *adjunctive* and *embedded causal* extraction procedures. Even given a forest of $k-$ary trees, it is only necessary to specify only the left-most and right-most children of a node in such directional processes. We express this modification as $\mathcal{F}'_{\blacktriangleleft}(A_{\langle q_0,Z\rangle}, \alpha^0, \zeta, \eta, T)$, and $\mathcal{F}'_{\blacktriangleright}(A_{\langle q_0,Z\rangle}, \alpha^0, \zeta, \eta, T)$. Each of these execution proceeds as other execution modes, except that the return set, in which a node is only included in the return set if it is the leftmost (or rightmost) node of its parent.

5.4 Frame preprocessing algorithm

In this section, we will describe the representation in data-structure and algorithms used to perform frame extraction. The focus of this part shifts from the theoretical to the practical, and aligns closely with the actual coding for the frame extraction component of the system.

5.4.1 Reformulation of semantic frame

The various conceptions of semantic frame theory provide a foundation for understanding processes at the syntax-semantic interface, which allows us to find an operational definition for semantic frame that must also be easy to extract an manipulate computationally. The structure in this study will largely be based on subatomic semantics in 5.2.2, with various augmentations that provide the extensibility to include information necessary for the following modules (including future planned parts of the project). The semantic frame instance is implemented as a single object containing all the necessary parameters to describe its structure, and its relations to other frames. It is also treated as relational database, such that any subset of frames that has a certain value for a specific parameter can be queries by a later module in the system.

5.4.1.1 Basic components

The basic components will be designed according to the specifications of subatomic semantics, which has the form

$$\phi_p(A_0,\ A_1,\ \ldots\ A_k, e_i) \wedge X_1(e_i) \wedge X_2(e_i) \wedge\ \cdots$$

. Each event token will have a unique identity in e_i, whose presence will be used to unify the essential structure within ϕ_p headed by the predicate p, which has associated with in a set of one or more permissible adicity structures drawn from a data-base; auxiliary elements such as separable particles are also regarded as part of the predicate string, as in *'pick ... up'*. Each of the A_i would an argument that can be placed in one to one correspondence to the predicate's lexical information. This process simulates the binding of one of the roles within the argument list of the predicate (acting as antecedent here) to the syntactic positions within the context free structure (acting as anaphora) as in Section 5.2.1. Those syntactic components that are the remainder from that process will be considered non-essential arguments X_j. It may be difficult to differentiate between oblique arguments versus pure adverbials, and they do not play a central role in the remainder of the processing, but are only important when they themselves represent events that have some association to the target event of the current frame.

5.4.1.2 Logical-functional parameters

Many additional parameters are there to allow the frame function correctly in terms of its semantics, as well as in temporal/sequential relation to other frames. The majority of these functional morphemes of the language reside in small, closed categories, and thus the extraction of these are relatively straightforward to incorporate into the automata-based procedure. Some of the basic logical parameters are extracted, such as *negation*, *polarity items*, and *quantifiers* which are essential for accurately analyzing the semantics of multi-frame constructions such as reciprocals conditionals, causals, cooperatives, and counterfactuals. For each corpus, there is a linear sequence among all of the extracted frames, allowing us to assign a baseline sequential index to each extracted frame. There are a few exceptions, such as coordinate structures that indicate parallelism, but the semantic of these are only important in a future *cooperative* module. This is the first, and simplest of the types of relations among events that we can extraction.

The linear sequence is also augmented by the full elucidation of the time and aspect of each frame's INFL elements, all of the necessary elements necessary for a Reichenbach (1947) style analysis of temporal sequence among event thought the analysis of TAM elements in clauses. The automaton-based extraction mechanism described in the following sections will extract all of the necessary tense and aspect information necessary for a future module on temporal inference, that places each frame-event into some permissible interval of time relative to the rest (equivalent to temporal sequence DAG). The modal information is also extracted by this mechanism, that allows for future analysis of each frame, and its categorization into various classes under *realis*, *deontic*, *epistemic* modalities.

5.4.1.3 Inductive forms

The baseline arrangement of the set of frames would be a linear sequence, but language affords mechanisms for extensive hierarchical arrangement of frames, where the temporal and logical connection of some frames are subordinate to that of other frames in the corpus. The main way this occurs is through the derivation of some frame-baring CP to become an argument of another frame, an adjunct of another frame, or adjunct of some argument in the arg-list of another frame. We have made allowance for a recursive mechanism to deal with these types of embedding when necessary. The arguments and adjuncts of each frame extracted has some syntactic subtree associated with each, and the subtree is recursively examined for the presence and structure of a contained frame, until we reach some base-subtree that no longer contained any element that can serve as a predicate of the frame. The extracted embedded frame at each level is then back-associated to its parent, which adds a level of structure in the relational graph of the event-frames of the data-set. Thus in accordance with type inference, each embedded event-frame can be made and argument (denoted as $a(\cdot)$), or made into an adjunct $c(\cdot)$, such that we can expand the definition of the set of frames $\mathbb{F}$ as:

$$\mathbb{F} := \begin{cases} \phi_p(A_0^p,\ A_1^p,\ ...\ A_k^p, e_i) \wedge X_1(e_i) \wedge X_2(e_i) \wedge\ ... \\ \phi_q(\ ...,\ A_j^q = a(\phi^x)\ ...\ ) \wedge X_1(e_i) \wedge .... \\ \phi_r(\ ...,\ ...) \wedge\ ...\ \wedge\ X_k(e_i) = c(\phi^x)(e_i)\ \wedge ... \\ \quad \mid\ \phi^x \in \mathbb{F} \end{cases} \tag{5.7}$$

where A_i^p is the i^{th} argument of the p^{th} frame within the linearly ordered set, and ϕ^x is some embedded frame in the construction, and $c(\phi^x)(\cdot)$ has same type as $X_j(\cdot)$.

5.4.1.4 Discourse relations

The sentential adverbs and connectives are not treated as part of the ϕ-structure or its adjuncts. They do not play a role in the frame structure or its compositional semantics, but some connectives are critical in elucidating bi-frame relations such as conditionals and counterfactuals. These connectives are either lexical adverbs that occur outside IP-environment, or small lexico-syntactic constructions such as *'in case'*, *'in conclusion'*, which are a small closed functionally defined class of elements. A few of the pairings of modal verbs between two adjacent clauses also contribute to these bi-clausal semantics. Some examples include *'... should ...., ... would ....'*, or *'... had ...., ... must ....*, these will be discussed in greater detail in Section 7.2.2 . There is a specially designed module to detect these connectives, and pairs of these special modals, by examining each pair of $\langle e_i, e_{i+1}\rangle$ in the linear sequence of the corpus.

5.4.2 Generic detection mechanism

The extraction of each element from T that corresponds to the necessary components of the semantic frame involves a run of B, with some set of constraints $\zeta = \{\langle \upsilon_i, \rho_i\rangle | i \in \mathcal{M}\}$. The baseline mechanisms include those that check constituency completeness on all levels using BPTA mechanisms. The mechanism manipulates constraints by pushing and popping individual requirements within the constraint, using the $sum(\chi, \chi')$ operation each time when a new requirement need to be added, when a push operation at $T_s \in dsc(T)$ occurs given the operation of the machine. Some basic properties of a tree that is syntactically well formed, and can be transformed to an LF-like structure, could be readily checked given the design of the mechanism.

5.4.2.1 Basic properties

One such property is observing $\varsigma(T_s) = XP, T_s \in dsc(T)$, and consequently needing to make an observation $T_t \in dsc(T_s | \varsigma(T_t) = X, X = head(XP)$. For the standard BPTA, $\mathcal{F}(q_0, \alpha, T_s)$ requires the stack reaching each one of the leaves to be empty, such that

$$\because q_1 \rightarrow (q_2, \gamma_i.u), Accept((q_2, \gamma_i.u), T_s), |T_s \in dsc(T)$$

$$\therefore \forall P = \langle T_s,\ ...,\ T_l\rangle[\exists T_t \in P \wedge q_3 \rightarrow (q_4, pop(\gamma_i)), \wedge Accept((q_4, u), T_t)]|T_l \in leaf(T)$$

; meaning that the given some $push(\gamma_i)$ at T_s, all paths that leads from T_s to a leaf T_l must entail the run of the machine to perform a corresponding $pop(\gamma_i)$. This requirement of $\forall P = \langle T_s, ..., \rangle$ is undesirable in our case, and would lead to more complex planning of the execution runs.

So we added the ability for this to be changed to an existential quantifier for some subtree, $\exists P = \langle T_s, ..., T_l\rangle[\exists T_t \in P \wedge q_3 \rightarrow (q_4, pop(\gamma_i)) \wedge Accept((q_4, u), T_t)]$; which then allow us easily use the machine to check many properties, like XP $\longrightarrow$ XP-head; or e.g. PP+prep $\longrightarrow$ Nominal. Many of these are necessary as precondition to see whether the S-structure of the frame-candidate, or candidate constituents of frame components are well formed, in order for transformations and frame extraction to occur. Here on, we will assume that a set of such basic mechanisms exist, and discuss more specific mechanisms for individual transformations and frame extraction. For a majority of the subtrees of the structures, the context free property of the root-leaf path in fact need to be existential, thus we make the default form $\mathcal{F}(A_{\langle q_0, Z\rangle}, \alpha, \zeta, \eta, T)$ to have the existential path-property; and when the universal path-property is necessary, we will change the denotation to $\mathcal{F}_\forall(A_{\langle q_0, Z\rangle}, \alpha, \zeta, \eta, T)$

5.4.2.2 Frame structure

For the extraction of some frame component, the target element is considered found iff

$$\mathcal{F}'(B_{\langle q_0', Z\rangle}, \alpha^0, \zeta, Z, T_s)|Z = \langle\rangle$$

where T_s is a root of a subtree that is determined by some prior run of the extraction of some other element, and β is the ordered set of requirements that must be met in order to locate the current element in T^s. Some relatively simple operations, such as locating any content V, can be performed using a single set of BPTA requirements; such as $\mathcal{F}'(A_{\langle q_0, Z\rangle}, \alpha^0, \zeta = \{VB, VBZ, VBD\}^{+\Psi_s}, T)$. This specific case has Σ_s = {VB, VBD, VBZ, ... }, Ψ_s containing the parameter specifying tags, and Γ_s = {is, has, ..., may, shall, ...}, and Φ_s containing the parameter specifying tokens.

More complex detections have multiple requirements which must be met in a particular sequence on a path from root to leaf, that need to applied with multiple BPTA executions. Each execution potentially has its own $\zeta = \Sigma_s^{+\Psi_s}$ and $\eta = \Gamma_s^{+\Phi_s}$. Multiple such $\langle\alpha, \beta\rangle$ pairs are applied, each in its own defined execution run from some subtree T_s within T. Given execution stages $[1, ..., l]$, the m^{th} stage begins its execution as $\mathcal{F}'(A_{\langle q_0, Z\rangle}, \alpha^0, \zeta_m, \eta_m, T^n_{m-1})$, where T^n_{m-1} is some node that is located by execution during stage $m-1$, serving as the root of the new execution. In this conception, one can view each $m-1$ stage execution as spawning one or more executions of m^{th} stage (zero would imply the entire detection for that particular $m-1^{th}$ execution failed); and each m^{th} returns a binary variable whether it was independently successful.

5.4.2.3 Generalize examples

One may conceive the structure as, starting off with a stack data-structure for each complex detection

$$(\langle\zeta_1,\eta_1\rangle,\langle\zeta_2,\eta_2\rangle,\ ....,\ \langle\zeta_m,\eta_m\rangle)$$

, for each branch of each execution of the machine, whenever a valid location of a node is made, the top of the stack is popped, and the next pair is used in execution. Any valid detection mechanism, for computational reasons, must have a small constant size for this stack. As a relatively straightforward abstract example, if we have tree-nodes of the type with symbol Y_1, Y_2 which are descendent from some node with symbol X_1 or X_2, but not in any subtree of X_1 with the labels U, V. This necessitates a two stage execution; in which $\Sigma_1 = \{X_1, X_2\}$, $\Gamma_1 = \varnothing$; and $\Sigma_2 = \{Y_1, Y_2\}$, $\Gamma_2 = \{U, V\}$. The generation execution can be defined as

$$\mathcal{F}'(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_2^{+\Psi_2},\Gamma_2^{+\Phi_2},T_i\in\mathcal{F}'(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_1^{+\Psi_1},\Gamma_1^{+\Phi_1},T))$$

.

We can look at a slightly more complex example, that is detection of a content V as the candidate to be the principle predicate of a frame; for specific examples, we will standardize to Stanford NLP's tagger. . The baseline requirement for the vast majority of frames would require some fully formed clausal structure, which can easily be detected with a 2-stage mechanism. The stack required here is $\Sigma_1 = \{TOP, SBAR, SBARQ\}$, $\Gamma_1 :=$ S, adjuncts that can occur directly under SBAR, and SBAR complements; $\Sigma_2 = \{S\}$, and $\Gamma_2 :=$ any complement of S or VP (NP, VP, PP, ADJP, etc), to avoid locating any other clauses that might be embedded or adjunct. Since this would be a mechanism that is used on the extraction of a majority of frames, we specially name the stack elements as $\langle\Sigma_{S1}^{+\Gamma_{S1}},\Gamma_{S2}^{+\Gamma_{S2}}\rangle$, and the mechanism

$$\mathcal{F}^S(T) := \mathcal{F}'(A_{\langle q_0,Z\rangle}, \alpha^0, \Sigma_{S2}^{+\Psi_{S2}}, \Gamma_{S2}^{+\Phi_{S2}}, T_{S2} \in \mathcal{F}\surd(A_{\langle q_0,Z\rangle}, \alpha^0, \Sigma_{S1}^{\Psi_{S1}}, \Gamma_{S1}^{+\Phi_{S1}}, T))$$

. At times, we would need a partial function that only locates the first stage node (CP), for certain classes of extractions, especially when SPEC-CP/TOP position is important; this would be defined as the shorthand:

$$\mathcal{F}^C(T) := \mathcal{F}\surd(A_{\langle q_0,Z\rangle}, \alpha^0, \Sigma_{S1}^{\Psi_{S1}}, \Gamma_{S1}^{+\Phi_{S1}}, T))$$

.

Given the overall structure of the clause having been located as T^s, locating the content V requires finding either a finite or non-finite form of a content V, with additional auxiliaries, modals, negations, and other potential lexico-syntactic structures that may interpose in between; simply choosing the appropriate Γ_i for each stage allows us to accomplish the avoidance of those. This slightly simplifies the real situation, where some syntactically non-auxiliary elements performing semantically *deontic modal* expressions such as *desiderative* (e.g. *'want'*), *evidential* (e.g. *'seem'*), *commissive* (e.g. *'promise'*), and many others. Some of these, such as *'ought'*, or *'need'*, selectively exhibit auxiliary modal characteristics, such as in *'you ought not make a scene in the ball room'*. We will ignore these peculiarities for the purposes of this example.

Another important issues is that the immediate ancestor VP node of the VB that we seek has no fixed location relative to the IP structure, but it is always the last consecutive VP on some path from r^{T_s} to any of its subtree leaves. So the execution for this stage becomes $\mathcal{F}\surd(A_{\langle q_0,Z\rangle}, \alpha^0, \zeta_3, \eta_3, T_s)$, or $\mathcal{F}'(A_{\langle q_0,Z\rangle}, \alpha^0, \Sigma_3^{+\Psi_3}, \Gamma_3^{\Phi_3}\}, T_s)$, where the modification $\mathcal{F}\surd(\cdot)$ is specified in Ψ_3; and $\Sigma_3 = \{VP\}$, Γ_3 is the set VP-complements except VP itself. And the final stage in its detection would have $\Sigma_4 = \{VB, VBZ, VBP, VBD, ...\}$, the set of all possible finite and non-finite lexical V forms, and $\Gamma_4 = \varnothing$. The entire execution is carried out with the constraint stack with each $\langle \Sigma_i^{\Psi_i}, \Gamma_i^{\Phi_i} \rangle$ as described above.

5.4.3 Detecting relevant tree structure

The BPTA mechanism is mainly used in the sem-frame extraction module as the means to detect relevant constituents in syntax that require further processing by additional BPTAs in the stack; until all requirements are fulfilled.

5.4.3.1 Preliminary filtering

For any frame, the extraction starts with the main predicate, which is the most deeply located content V of the clause without crossing an embedded S-boundary. It then looks for a few additional logical properties associated with the main predicate, such as negation or the voice of the verb. The algorithm proceeds to enumerate through the NPs and PPs in appropriate positions in the tree, starting with the likely position of ARG0 (in the external or first internal position, depending on voice), and proceeds to find the constituents that are potentially obligatory arguments.

Many of the preliminary requirements can be verified through the examination of the set of terminal symbols alone, these are simple pre-filters for the existence of some $\sigma \in$ the surface string of symbols without its parsed form. One basic requirement is the presence of some candidate symbol that could be construed as the *predicate*, mostly in the form of a content V, but occasionally occurs in other forms such as predicate copula, and optionally may also include nominalized frames appearing in gerundic forms (normally with VBG tag).

5.4.3.2 Principal frame components

Relevant clauses are processed according to the aforementioned principles of sub-atomic semantics and frame decomposition. e.g., take the simplified parse “

$$[_{SBAR}\ [_{S}\ [_{S}\ [_{NP}\ Ian\ ][_{VP}\ [_{VP}\ sent\ [_{NP}\ Jane]$$

$$[_{NP}\ two\ pairs\ of\ bracelets[_{SBAR}\ which\ [_{S}\ were\ [_{VP}\ [_{ADVP}\ custom]$$

$$made\ [_{NP}\ \langle t\rangle]\ [_{PP}\ by\ a\ master\ goldsmith\ ]]\ ]]\ ]]\ ]]$$

$[_{ADVP}$ *last year* $][_{PP}$ *on Thanksgiving* $][_{PP}$ *through a courier* $]\ ]]$

". A number of constituents are presents as candidates for inclusion into the frame, a subset of these would be deemed to represent core components of the frame, such as the *predicate*, *external argument*, and essential *internal argument*s.

It first determines the location of the primary anchor of the frame, the location of the deepest embedded content verb without crossing a secondary SBAR boundary, which in this case would be '*send*' after being lemmatized. Next steps attempts to use runs of the machine B to detect voice, negation, and other relevant modifications to the VP that affect the structure of the frame, normally located in a region bounded by the immediately ancestral S and the first NP/PP structure within the VP, which come to $voice = 0$ (active) and $-negation$ for the matrix clause that we are concerned with. In the next step it seeks ARG0 in the region bounded by the modal/auxiliary branch and the immediately ancestral S (since this is an active construction), and locates the NP *'John'*. With a passive construction, this run would have a different set of constraints. Thereafter, more execution runs of the machine locates ARG1 and ARG2 within the VP structure headed by anchor.

5.4.3.3 Canonical arg-structure

Given that each frame at a semantic level has a structure similar to $\phi(A_0,\ A_1,\ ...\ A_k, e_i) \wedge X_1(e_i) \wedge X_2(e_i) \wedge\ \cdots$, where the essential arguments are named A_i, and obliques and non-essential components are X_j. For each lexical predicate, the set of possible adicity structures for the essential arguments can be read from the adicity data-base. The the potential adicity structures allows us to detect anomalous structures in some instances, which likely necessitate some transformation to re-create the semantic $\phi(A_0,\ A_1,\ ...\ A_k, e_i)$ part of the representation. A number of such scenarios exist, especially in cases where some essential argument has moved outside of the VP or the external argument position, or is otherwise covert. These will be discussed in sections of movement classes.

Some permissible adicity sequence in the adicity data-base, such as SVR, $SVOA$, $SVOS$, would need to be converted into a sequence of sets of symbols, such as R generally corresponds to $\{PP, PRT\}$, which then can be used to test against the set of candidate-arguments for a potential frame, to detect any argument type that might be missing. (The standard adicity structure represents a sequence of broadly defined predicate and argument cases, $\{S_{subj}, V, O, A, R, C, S_{vp_complement}\}$.) We will represent the list of argument structures as indicated in the adicity database as the shorthand $\mathcal{A}(\phi)$, and $\mathcal{A}'(\phi)$ being the canonical structures that have been translated into corresponding symbol sets, such as for *'gorge'*, $\mathcal{A}'(\text{`}gorge') = \left[\left\{\{NNP, NN, NNS, ...\}\right\},\ \left\{\{NNP, NN, NNS, ...\},\ \{PP, PRT\}\right\}\right]$ And for any putative set of arguments that can be seen from surface structure prior to any transformation, as T is parsed by a generic parser, we denote this *apparent* argument list as $\mathcal{S}(T) = \langle A_0, A_1,\ ...\ A_k\rangle$. Also let $\mathcal{A}'_{[i,j]}(\phi)$ be the i^{th} argument of the j^{th} canonical structure of ϕ, and $\mathcal{S}_{[i]}(T)$ be the i^{th} surface argument of the parsed form T, as shorthand representations. And let $\mathcal{P}(\langle A_0, A_1,\ ...\ A_l\rangle)$ be the set the permutations of such a list, and let $\mathcal{P}_{\langle i,j\rangle}(A) = \mathcal{P}(A_i, A_{i+1},\ ...\ A_{j-1}, A_j)$.

5.4.3.4 Need for transformation

The previous cases need no particular adjustment due to VP voice or other transformation, but in a clause such as:

$$[_{SBAR}[_{S}\ [_{NP}\ \textit{the bracelets}\ ][_{VP}\ \textit{were}\ [_{VP}\ \textit{made}$$

$$[_{NP}\ \langle t\rangle]\ [_{PP}\ \textit{by a master goldsmith}\ ]]\ ]]\ ]$$

. The structural modifications to the VP, since it is passive, is detected to have no argument in external position, but nonetheless, an external NP is detected in SPEC position. Given that we know the arity of $\phi_{make}(\cdot)$ to be required to be two in terms of obligatory arguments, the tree rooted at the trace position within the *'made by a master goldsmith'* VP would be inserted, and the tree in the trace NP, which is a terminal, would undergo a substitution operation $T_{VP} \circ_{\langle t\rangle} T_{NP_{extern}}$, with the tree of the NP in the external position, and thus the final predication structure is $\phi_{make}(ARG0_{...goldsmith}, ARG1_{...bracelets})$ in this case.

Another common example type of observed transformation would resemble the following:

$$[_{SBAR}\ [_{S}\ [_{NP}\ the\ master\ goldsmith\ ]$$

$$[_{VP0}\ [_{VP1}\ forged[_{NP}\ two\ pairs\ of\ bracelets]$$

$$[_{PP}\ forJane]]\ [_{CONJ}\ and\ ]\ [_{VP2}\ [_{PP}\ for\ Mary]]]\ ]]$$

. Note that there is inherent ambiguity in this surface form, as the NP *'tow pairs of bracelets'* and refer either to one pair for Jane, and one pair for Mary; or two pairs for each of Jane and Mary. The elucidation of this level of ambiguity requires the *cooperative* module analysis, and we here assume the *distributive* reading, so that semantically it represents two independent events, such that the goldsmith in one event forged a two pairs for Jane, and in another event two pairs for Mary. In that case, the coordinate structure can be rewritten as the two surface forms: *'the master goldsmith forged two pairs of bracelets for Jane'*, and *'the master goldsmith forged two pairs of bracelets for Mary'*.

In that case, we can see that this complex coordinate structure shares both the external argument *'the master goldsmith'*, the predicate bearing V *'forge'*, as well as part of the internal argument list *'two pairs of bracelets'*. In order to form two new syntacto-semantic structures that best represent two distinct events, we need to in some sense duplicate the aforementioned components in the context-free structure that correspond to essential components of these two semantic frames. Thus we need to detect such parallel structures, and then find the points in the tree, where one of the forms $[_{S}\ [_{NP}\ [t]_{extern}]$ $[_{VP2}\ [t]_{pred}[t]_{intern}[_{PP}\ for\ Mary]]\ ]]$ is only complete when additional structures, external argument, verb, and one internal argument are grafted onto it, and each of $[t]_{.}$ is such a point where these components are located, after we have re-transformed into a form that is relevant to completing the semantic frame. This can be done by duplicating the necessary parts, then grafting them; but in order to avoid unnecessary memory space usage, it is more efficacious to "virtually" graft the same structures onto those sites in the second event form. After this re-transformation, we no longer have a strictly context structure, but a DAG (discussed in detail in Section 5.4.4.8).

5.4.3.5 Tree reconfiguration

For each transformation class and subtype, once we have determined the constituent that has moved between the surface and deep-syntacto-semantic forms, the parsed tree structure needs to be reconfigured to serve as input for causal-relational extraction algorithms in Chapters 7 and 8. The principal function of the component is to create the location of the moving constituents in the context free representation of the structure at syntax-semantics interface, and graft the constituent there. As mentioned, this process of re-transformation into structures that better represent semantic frames potentially could have certain constituents become descendent to more than one subtree of the overall structure. This means that the process of re-transformation may turn the tree into a more general *directed acyclic graph.*

In some instances, it would be sufficient to increase the arity of an existing node in the tree. This can be seen in redirecting an external argument at the surface level to an internal argument position, such as in recreating the tree structure prior to a passivization transformation. The deep structure would be represented in the moving constituent, which at the surface level is in the SPEC position of an S or SBAR (depending on exact example), be redirected to be a daughter node of the VP; where the arity of the VP changes by +1. For instance, in previous our clause: $[_{SBAR}[_{S}$ $[_{NP}$ *the bracelets* $][_{VP}$ *were* $[_{VP}$ *made* $[_{NP}$ $\langle t \rangle]$ $[_{PP}$ *by a master goldsmith*]]]]] where $\langle t \rangle$ is not an overt position; the constituent $[_{NP}$*the bracelets* can be redirected from the SPEC-S position back into the VP: $[_{VP}$ *made* $[_{NP}$ *the bracelets*] $[_{PP}$ *by a master goldsmith*]]]]] (still need to move the instrumental case argument in a separate movement to a position above VP). In another example: *John finances Mary* $[_{S}$ $[_{VP}$ *to* $[_{VP}$ *conduct* $[_{NP}$ *the operation*]]]], the embedded structure has a *PRO* position that remains $[_{S_{emb}}$ *PRO* $[_{VP}$ to $[_{VP}$ conduct $[_{NP}$ the operation]]]]. The constituent $[_{NNP}Mary]$ is identified with PRO (not always, intricacies discussed in section 5.4.4.4).

In other types of transformations, since the automatic parsing doesn't generally take $[t]$

positions into account, the constituent would need to move into a level of hierarchy that is non-existent in the surface parse. An example such as $[_S$ *John* $[_{VP}$ *planned*]] *and* $[_S$ *Mary* $[_{VP}$ *carried out* $[_{NP}$ *the bank robbery*]]] is a type of coordination that at a deep semantic level, have two frames sharing the internal argument $[_{NP}$*the bank robbery*]. The redirection of this constituent to both VPs of the two conjoined clauses, such that the result becomes equivalent to $[_S$ *John* $[_{VP}$ *planned* $[_{NP}$ *the bank robbery* $]_i$]] *and* $[_S$ *Mary* $[_{VP}$ *carried out* $[_{NP}$ *the bank robbery* $]_i$]] , where the two internal arguments are coindexed to have identity with one another. This redirection of the same NP structure to two positions within the overall structure has the natural advantage of allowing the same event-argument (the robbery) to be identified with one anther through the structure itself; which may be important for certain types of further processing (for analysis of *cooperativity*, e.g.) . In another example such as *Jack* $[_{VP}$ *cooked* $[_{NP}$ *the bacon*] $[_{PP}$ *into a crisp*]] , the structure presented by the parser is not amenable for frame based representation, and misses the embedded frame entirely. The primary reason is that the parser does not utilize formal semantic knowledge of what possible adicity lists the matrix predicate *'cook'* should allow. The way to transform this structure into something more appropriate for the deep semantic representation is not to look for individual movements, and move to recreate the most appropriate representation of the embedded frame in a syntactic tree, where the entire embedded clause is an adjunct to the matrix event. This becomes *Jack* $[_{VP}$ *cooked* $[_{NP}$ *the bacon*] $[_S$ $[_{NP}$ *the bacon*] $[_{VP}$*BECOME* $[_{PP}$ *into a crisp*]]]] ,

5.4.4 Major movement classes

As we have seen, movements present between the S-structure and the eventual semantic form constitutes the main obstacle in accurately locating the boundaries of frames, and detecting and assessing the individual components of semantic frames; and this is the primary application of the automata based extractions procedures. The linguistic notion of movement dictates that the context-free structure be transformed from its most semantically relevant configuration (often corresponding to structure at LF), into its most well formed configuration in terms of permissible linear orders in the language. The elucidate of an individual movement involves locating the source and target locations of some moving constituent within the context-free structure, and then recreating the original prior to movement (often corresponding to LF).

The individual movement modeled for our purposes is a simplification of the formal syntactic notion of movement, including steps necessary to complete the transformation from the surface form back to a form amenable for frame analysis; but only the structural elements that are relevant to the representation by the machine parser, and need some adaptation for the peculiarities of each type of parser. For each class of movement, the general strategy is to perform two (or more in a few cases) detections, at least one on the current location of the moving constituent, and the other on the proposed origin; then perform some class of transformation (in the opposite direction of the linguistic transformation), through some tree operation or some combination of operations such as redirection or duplication.

There are a number of distinct class of movement operations that need to be treated separately, and most also have a complex taxonomy within each class with respect to the necessary detection mechanism. We will discuss the most frequent and essential transformation classes for sem-frame construction. We will do a series of linguistic analyses in the context of the forms of these movement classes presented by parsers, to provide the most appropriate application of the automata-based extraction mechanism for each class and each subtype within.

5.4.4.1 Clause with complex TAM structure

TAM structure is most common type of transformation that needs to take place, in order convert the parsed surface form into something more representative of the semantic frame structure.

5.4.4.1.1 General characteristics The components of interest here is the structure between the external argument position and location of the main content verb. The principal categories that reside there are auxiliaries and modals that express tense-aspect-mood of the frame. Other elements such as negation and adverbial also may be present. These structure contribute to the TAM components of the frame's meaning, for the most part representing their sequential relation with other event-frames (tense and aspect), as well as relation to reality and thought processes (modality). Thus these elements generally do not significantly contribute to the primary structure of the frame, and thus the primary function the extraction procedure performs is to ligate the external argument of the frame with the the predicate (usually content verb) and the rest of the argument list.

A mechanism that detects the typical case of TAM structure is relatively simple, and is formulated to first detect the presence of any candidate for external argument that can be an entity; then it is formulated to seek some form of VB that cannot be construed as auxiliary or modal. The baseline detection algorithm for these types of structures to transform is so formulated:

$$\begin{cases} \mathcal{F}^{\mathcal{X}}(A_{\langle q_0, Z\rangle}, \alpha^0, \Sigma_{AM}^{+\Psi_{AM}}, \Gamma_{AM}^{+\Phi_{AM}}, T_s \in \\ \qquad \mathcal{F}'(A_{\langle q_0, Z\rangle}, \alpha^0, \Sigma_{V}^{+\Psi_{V}}, \Gamma_{V}^{+\Phi_{V}}, T_s \in \\ \qquad \mathcal{F}^{\int}(A_{\langle q_0, Z\rangle}, \alpha^0, \Sigma_{EA}^{+\Psi_{EA}}, \Gamma_{EA}^{+\Phi_{EA}}, \mathcal{F}^{S}(T)\ )\ ) \end{cases} \tag{5.8}$$

where the parameters Σ_{EA} := the set of symbols of nominals appropriate for external argument of a full clause, Γ_{EA} := the set of symbols that are VP or VP-equivalent (e.g. a adjectival that is headed by a VBG). The parameters Σ_V := the set of symbols that correspond to the set of (terminal) tags that can be construed as verbs, Γ_4 := the set of symbols that correspond to VP complements (e.g. NP, PP, ADJP, etc). The parameters Σ_{AM} := the set of tokens that correspond to the closed list of auxiliaries and modals, and $\Gamma_5 = \varnothing$. The detected components are the external argument on one hand, and VP containing predicate and internal arguments. The context free structure then would be reconfigured, so that these constituents are redirected to a new immediate ancestor S, and that new S node would replace the existing S corresponding to the frame.

5.4.4.1.2 Complicating issues There are a number of additional complicating factors that need to be consider for this class. One is that, while many modals correspond to specific lexical entry, such as *'can'*, *'may'*, *'must'*, etc; the remainder utilize verbs that can serve dual roles as both main predicate in certain context, but also perform the semantic function of modal auxiliaries; examples include a small class, e.g. *'need'*, *'dare'*. As we can observe in the following:

1. *Stephen needs the certification for a future employment*
2. *Stephen need not obtain the certification during this year*
3. *Teresa dares Jane to take on the covert assignment*
4. *Teresa dare not volunteer for the most dangerous covert assignment*

We can see that in (1) and (3), *'need'* and *'dare'* perform the function of the predicate for a complete frame, Each of these

$$\phi_{need}(A_{Stephen}, A_{certification}, A_{SC}, e_i)$$

$$\phi_{dare}(A_{Teresa}, A_{Jane}, A_{SC}, e_j)$$

(which themselves require further analysis for their SCs) is a fully fledged frame, with appropriate adicity structures (in the data-base, SVOS for both counts). On the other hand, in (2) and (4), the same lexical items express modality; in (2) it expresses the *alethic* modality of necessity of the event described by the frame to correspond to the real-world; and in (4) it expresses a *dynamic* modality related to fear on the entity's part with respect to the prospective event.

Another issue is some periphrastic constructions also can have semantic content that express modality. Two classic examples would be *'have to'* and *'being able to'*, as in the following:

1. *Stephen must train for the job this month*
2. *Stephen has to train for the job this month*
3. *Teresa can take on the mission by herself*
4. *Teresa is able to take on the mission by herself*

We observe that (1) and (2) express the same event of $\phi_{train}(A_{Stephen}, A_{for\_job}, e_i)$, and the relation between the entity and the event remains one of *deductivity.* (3) and (4) also express the same event of $\phi_{undertake}(A_{Holie}, A_{mission}, e_j) \wedge X_{by\_herself}(e_j)$, both with the same relation of *permission.* In both, the periphrastic construction approximates the semantic function of the proper modals. To accurately represent the frame information in these types of examples, it is necessary to identify the periphrastic forms, and to treat these intervening structures the same way as modals.

5.4.4.2 WH-movement

Structures containing WH-movements are essential in both interrogative frames as well as relative constructions, and thus are a vital component of the analysis of linguistic data rich in entities and relations. Especially with respect to its role in in entity relative clauses, the correct analysis of this class plays a large role in understanding the topology of a social structure.

5.4.4.2.1 General characteristics The typical WH-movement involves the relocation of some constituent, which corresponds to some argument position, to a position within the CP but external to the INFL bearing constituent position; thus both external and internal arguments can be moved from positions in D structure to SPEC-CP. The WH-element (token) comes from a small closed set of surface tokens, and the WH-constituent similarly has a small set of { *WHNP, WHPP, WHADVP, ...*}, the WH-element is not always the head of the WH-constituent, such as in most if not all WHPPs. But all effective WH-constituent undergoing movement should contain some WH-element; which means that the location of the moving constituent is relatively easy to detect. For certain parsers, sometimes the overall clause is marked as SBARQ, which also further eases detection, but is not a reliable cue.

The mechanism for detecting WH-movement needs to first look for the last SBAR (or equivalent) in an uninterrupted chain of SBARs, which is expressed as $\mathcal{F}^C(T)$. The location of the WH-constituent then can be searched for from that point, followed by confirmation of the WH-element terminal token; these combined locates the moving constituent's S-structure location. The search for the trace location then proceeds from the sister(s) of the WH-constituent, constrained by the lexical properties of the WH-element (described in the following subsection). The selection of whether the external-position trace or internal position trace is based on whether there is a good external argument candidate:

$$\left\{\begin{array}{l}\imath(\mathcal{F}'(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{EN}^{+\Psi_{EN}},\Gamma_{EN}^{+\Phi_{EN}},T_t \in \\ \mathcal{F}\surd(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{S2}^{+\Psi_{S2}},\Gamma_{S2}^{+\Phi_{S2}},T_s \in \mathcal{W})\ )\ ) \\ \quad\Bigg| \qquad \mathcal{F}'(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{W}^{+\Psi_{W}},\Gamma_{W}^{+\Phi_{W}},T_s \in \mathcal{F}^C(T)\ )\end{array}\right. \tag{5.9}$$

as well as additional lexical properties from the WH-element itself. The baseline algorithm for this class is described for movement from external-arg position:

$$\left\{\begin{array}{l}\mathcal{F}\surd(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{S2}^{+\Psi_{S2}},\Gamma_{S2}^{+\Phi_{S2}},T_s \in \mathcal{W}) \\ \quad\Bigg| \qquad \mathcal{F}'(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{W}^{+\Psi_{W}},\Gamma_{W}^{+\Phi_{W}},T_s \in \mathcal{F}^C(T)\ )\end{array}\right. \tag{5.10}$$

And the following when the trace in internal-arg position:

$$\left\{\begin{array}{l}\mathcal{F}\surd(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{VP}^{+\Psi_{VP}},\Gamma_{VP}^{+\Phi_{VP}},T_t \in \\ \mathcal{F}\surd(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{S2}^{+\Psi_{S2}},\Gamma_{S2}^{+\Phi_{S2}},T_s \in \mathcal{W})\ ) \\ \quad\Bigg| \qquad \mathcal{F}'(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{W}^{+\Psi_{W}},\Gamma_{W}^{+\Phi_{W}},T_s \in \mathcal{F}^C(T)\ )\end{array}\right. \tag{5.11}$$

The parameters $\langle\Sigma_{VP},\Gamma_{VP}\rangle$ involve the set of VP or VP equivalent nodes, and VP complements; $\langle\Sigma_{EN},\Gamma_{EN}\rangle$ involve the set of tags likely able to serve as external argument constituents, and complement to NPs; $\langle\Sigma_{S2},\Gamma_{S2}\rangle$ are similar to the previously mentioned parameters of the same names. $\langle\Sigma_W,\Gamma_W\rangle$ are the set of WH terminal tokens, and $\varnothing$, which requires some specialization, since we need to recognize, within the BPTA, that WH-constituent and WH-element have a constraint relation on some path $T^r,\cdots,T_{WH-elem}$. So these has an internal structure of $\chi_W = \{\langle\upsilon_i,\rho_i\rangle\}|\Big((\upsilon_w = 1,\rho_w = 0)\wedge(\upsilon_z = 0,\rho_z = 1)\wedge(\forall w \in \mathcal{M}\ [\mathcal{M}_w \in WH-ELEM])\wedge(\forall z \in \mathcal{M}_w[\mathcal{M}_z \notin WH-ELEM])\Big)$

5.4.4.2.2 Lexical peculiarities The WH-movement detection involves detection of the WH-constituent, which can be made more accurate by locating its WH-element; as well as the detection of some argument potentially being missing. One of the advantages of WH-movement detection is that the token of the WH-element often informs us what type of argument we should be looking for. For *animate* entities, the external/internal-argument distinction is often made clear through the *'who / whom'* dichotomy (but not the *'what'* token for the *inanimate* entities, nor the WH-DET *'which'*). On a side note, entities whose NP takes a relative clause with *'who/whom'* element can be a good test for its *animacy*. If there is a sufficiently large data-set of social network data over the same set of individual people and object, we may be able to utilize this property to discover animacy for each entity.

Other elements such as *'how'*, usually indicating the $\theta-$role of *manner* (which is normally non-essential), all but guarantee that the argument position that the WH-constituent moved from is internal and non-essential. All WHPPs, which are not headed by the WH element, but the preposition, also all have strong inclinations in terms of what role they play in the argument structure; such as $[_{PP}$ *for whom/which/what* $]$ generally indicates a role of *purpose*, *benefactive*, or *recipient*, and consequently their likely position in the argument list; or the constituent $[_{PP}$ *by whom/which/what* $]$ generally conveys the role of *location* or *instrument*. These and other lexical properties of WH-elements allow us to better analyze the original role and position of these WH-constituents, and hence better perform re-transformation into more frame-relevant structures, and the baseline algorithm would be modified to take these into account.

5.4.4.3 Passivization

Passivized structures is also a frequent phenomenon, whose detection and analysis is necessary for the ordering of the argument list to be correct. It involves the altering the placements of normally internal and external arguments, and sometimes recognition of a missing essential argument.

5.4.4.3.1 General characteristics Passivized structures are generally used in cases where either the original external entity in a role such as *cause*, *agent*, *benefactor*, *force*, is unknown to the speaker, unspoken for some contextual reason in discourse, or that the semantic focus is placed on one of the original internal arguments. There are two major characteristics that can be reliably used to detect passivized constructions. One is that the predicate-V now occurs in a participial form, which sometimes can be distinguished from finite forms, but other times is identical to the aorist past form of the verb. The other is that the external argument position of the structure is occupied by one of the internal arguments, which generally permits all essential arguments in that position, provided that the complete argument-constituent is moved.

There are two basic types of forms of passivized structures, the non-agentive, and agentive. The baseline detection mechanism for passivized structures of non-agentive type can be formulated as:

$$\left\{\begin{array}{l} \mathcal{F}^{\mathcal{N}}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_P^{+\Psi_P},\Gamma_P^{+\Phi_P},T_t \in \\ \qquad \mathcal{F}^{\int}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{TN}^{+\Psi_{TN}}, \\ \qquad \Gamma_{TN}^{+\Phi_{TN}},T_s \in\in \mathcal{F}^S(T)\ )\ ) \\ \quad \Bigg| \quad \forall L_j \in \mathcal{A}'(V) \quad \Big[|\mathcal{S}(T)| < |L_j| \\ \qquad \bigvee \quad \nexists L_p \in \mathcal{P}(\mathcal{S}(T)) \\ \qquad \forall A_i \in L_j[\mathcal{A}'_{[i,j]}(V_{mtx}) = A_i] \Big] \end{array}\right. \tag{5.12}$$

The parameters $\langle \Sigma_P, \Gamma_P \rangle$ is responsible in detecting a structure of auxiliaries that indicates passivization. $\Gamma_P :=$ a set that contains terminals outside of auxiliaries, and all tags other than VP. $\Sigma_P :=$ consists of forms of past participial forms; it is pushed on when a *'to be'* form is detected, and is popped when a past participial is detected. The parameters $\langle \Sigma_{TN}, \Gamma_{TN} \rangle$ is similar to Σ_{EN}, except those tags that are the most appropriate for *theme*, *patient*, *phenomenon* and similar $\theta-$roles. And the agentive passive type can be formulated as the following:

$$\begin{cases} \mathcal{F}^{\surd}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{BY}^{+\Psi_{BY}},\Gamma_{BY}^{+\Phi_{BY}},T_u \in \\ \qquad \mathcal{F}^{\mathcal{N}}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{P}^{+\Psi_{P}},\Gamma_{P}^{+\Phi_{P}},T_t \in \\ \qquad \mathcal{F}^{\int}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{TN}^{+\Psi_{TN}}, \\ \qquad \Gamma_{TN}^{+\Phi_{TN}},T_s \in\in \mathcal{F}^{S}(T)\ )\ )\ ) \\ \quad \Bigg| \quad \Bigg[\Bigg[\exists L_j \in \mathcal{A}'(V)\ |\mathcal{S}(T)| = |L_j|\Bigg] \\ \qquad \bigwedge \quad \Bigg[\forall L_j \in \mathcal{A}'(V) \quad \nexists L_p \in \mathcal{P}(\mathcal{S}(T)) \\ \qquad\quad \forall A_i \in L_j[\mathcal{A}'_{[i,j]}(V_{mtx}) = A_i]\Bigg]\ \Bigg] \end{cases} \tag{5.13}$$

The parameters $\langle\Sigma_{BY},\Gamma_{BY}\rangle$ has to do with detection of argument that likely originated from external position in D-structure. Γ_{BY} is a set of tags of all clausal and VP nodes. Σ_{BY} contains nominals that can serve as complements to a small set of prepositions, especially *'by'*, which is pushed on when prepositions such as *'by'* is detected. Each subtype of passivized structures would require some additions and amendments to these basic mechanism.

5.4.4.3.2 Common complications In addition to their basic traits, there are some complicating issues for both *non-agentive* and *agentive* types. One important characteristic used to identify most passivizations is the lexico-syntactic structure associated with $\langle\Sigma_P,\Gamma_P\rangle$, such that the machine pushes on Σ_P. Whenever it observes $T_s|T_t \in N_T^+(r^{T_s}), \varsigma(T_s) \in$ the set of *'to be'* conjugated forms; and thereafter, whenever it observes $T_u|T_u \in dsc(T_s), T_v \in N_T^+(r^{T_u}), \varsigma(T_v) \in \Sigma_P$ in the same continuous path of VPs, it recognizes the characteristic. There are minor issues such as some RBs such as negations that can occur in the VP-chain, but most of the time the parser outputs a structure that could work with the mechanism, where the (RB not) is resides in a flat structure that also includes the *'to be'* auxiliary, such as in *'John (VP (VBD was) (RB not) (VP targeted))'*;

One complication is with the derivational process the participial forms of the verb and adjectivals often go through, so that is is sometimes very difficult for parsers or even analyses in theoretical syntax to tell between the two.

1. *John was promoted at his job this week*

2. *John was confused at his job this week*
3. *John was bored at his job this week*

In may parsers, all three of the previous example are parsed the same way, with a *(VP (VBD was) VP((VBN ··)))* structure that would be recognized by the baseline machine. But for most speakers, while (1) should be universally recognized as passivized, many will recognize (2) as a *predicate copula*, with the VBN having been zero-converted to a JJ, and (3) is almost always recognized as predicate copula. Since there is no difference in the structure except the identity of the content verb, the only possible solutions would lie in pure lexico-semantics or in discourse analysis.

Another issue pertains to the detection of this is that, infrequently, there will be occurrence of other VP-adjuncts in the middle of that chain, making it difficult to complete that part of the stack mechanism. A variety of adjunct structures may occur in these locations, all generally serve adverbial functions for the VP:

1. *Mary was* ***generously*** *praised by the professor*
2. *Mary was* ***in many ways*** *admired by her classmates*
3. *Mary was* ***as she checked her scores*** *encouraged by the her recent progress*
4. *Mary was* ***given that John and her participated in the robbery,*** *pursued by the authorities*
5. *Mary was,* ***being made aware her financial situation****, enticed by John to participate in the bank robbery*
6. *Mary was,* ***having been on the run for months from the authorities****, forced to flee to a neighboring state for the time being*

The issue here is that with adverbial adjuncts occurring in arbitrary places within the matrix clause, parsers often do not provide the correct output that clearly preserves the VP-chain needed to observe the lexico-syntactic substructure associated with $\langle \Sigma_P, \Gamma_P \rangle$ Some instances are relatively easy to detect, such as those with (1) or (2), the mechanism can easily be modified to accommodate constituents such as PP or ADVP that occur in that position. Those with longer embedded SC or clausal structures acting as adverbials are harder to resolve; But like in (3) and (4), they often contain a small, closed set of lexico-syntactic cues to indicate the presence of such constituents. Such can include *'as'*, *'given that'*, and others like *'since'*, *'as soon as'*, *'for the reason'*, that indicate adverbial clauses of *purpose*, *contrast*, *time*, *place*, *manner*, *reason*, *concession*, and a few other semantic categories. The high frequency types can be found by providing these lexico-syntactic cues, and the mechanism can be modified to incorporate them. Others such as (5) and (6) possess one of these frequently occurring cues, cannot be easily resolved with our mechanisms; these will require additional machine learning techniques or methods similar to the entire Chapter 8.

Another issue exists regarding the *agentive* type of construction, with regard to the detection of the original external argument now in the *'by'*-type PPs. We an observe in the following:

1. *The militants were observed to renew their offensive by the government's intelligence efforts*
2. *The militants were observed to renew their offensive by the oil-fields of Baiiji*
3. *The militants were observed to renew their offensive by giving more resources for reconnaissance aircraft in the region*
4. *The militants were observed to renew their offensive by the narrowest of margins*

First there is ambiguity within the lexical item *'by'* itself, as in (2), (we refer to the interpretation that the *'by'*-PP is attached to the matrix clause, not the SC, in all of these examples) where it has a locational sense. This can be checked (although not highly reliably) using a general NER procedure, and see if anything within the constituent under PP could be classified as LOCATION. Another possibility is that the constituent under the PP does not represent any kind of entity, but rather an event. This is likely evidenced by the head of that constituent, which in the case of (3) would be a VBG; the detection of which could be incorporated into our mechanism, and tells that it could not be the original *agent* or *perceiver*. Yet even more difficult cases are those such as (4), where the PP contains a periphrastic or semi-idiomatic expression that may appear to be an entity, and yet could not possibly serve as the external argument at LF. These cases are beyond the ability of our mechanism or traditional semantics to deal with.

5.4.4.4 Embedded small clause

Embedded SC constructions include those embedded clauses that have deleted full-CP or TP, while having a place for an external argument (not always overt), and usually contains no tense information (but does sometimes contain aspect part of the INFL).

5.4.4.4.1 General characteristics There are some variations in what linguistics consider SC; for our purposes, we will use a relative broad definition for SC, and consider small clause construction to encompass a few different types. They have a tendency to occur in tandem with *exceptional case marking* (ECM) class verbs in the matrix clause, such as *'want', 'believe', 'judge', 'consider'*, etc. These are highly relevant to SC constructions because they permit the raising of arguments of embedded clauses, where the matrix contains the ECM predicate.

Many SC-containing ECM matrix forms are represented as flat rather than hierarchical structures with respect to the adicity data-base. For instance, $\mathcal{A}(\text{`}see\text{'}) = \{SV, SVO, SVOA, SVS, SVOS\}$, of which SVOA, SVOS are representations of the ϕ_{mtx} directly over arguments in the embedded argument list (will be evident after next subsection). For this reason, for ECM ϕ_{mtx}, we need to treat all putative arg-lists with more than a single internal argument, to be a candidate for a more hierarchical analysis, with some embedded SC construction. The baseline detection algorithm for this class given the input T is formulated as:

$$\begin{cases} \mathcal{F}'(A_{\langle q_0,Z\rangle}, \alpha^0, \Sigma_{VP}^{+\Psi_{VP}}, \Gamma_{VP}^{+\Phi_{VP}}, T_t \in \\ \quad \mathcal{F}^{\int}(A_{\langle q_0,Z\rangle}, \alpha^0, \Sigma_{EA}^{+\Psi_{EA}}, \Gamma_{EA}^{+\Phi_{EA}}, T_s \in \mathcal{F}^S(T)\ )\ ) \\ \quad \Big| \quad \forall L_j \in \mathcal{A}'(V_{mtx}) \quad \nexists L_p \in \mathcal{P}(\mathcal{S}(T)) \\ \qquad \Big[\forall A_i \in L_j[\mathcal{A}'_{[i,j]}(V_{mtx}) = A_i]\Big] \end{cases} \tag{5.14}$$

where the parameters of Σ and Γ in the instances have some variability according to the subtype within the class. Σ_{EA} := *appropriate symbols for some entity that resides in external position of the embedded clause, or that can be moved from PRO*, Γ_{EA} := *any complement to nominal categories that can occur there.* Σ_{VP} := *some appropriate complement to the embedded VP*, Γ_{VP} := *any other VP complement or S complement.* Most subtypes would have additional characteristics to be discussed in the following subsections, each requiring additional modifications.

5.4.4.4.2 SC surface types These include the type of overt-verbless predication structure, the type of subordinate structure with only a participial or gerundic verb form serving as the predicate, as well as the type of IP-clause that contain a non-finite verb to be included. Consider the following:

1. *John considers* $[_{SC}$ *Mary a good partner for robbing banks*]
2. *John wiped* $[_{SC}$ *the table clean*]
3. *John and Mary blasted* $[_{SC}$ *the vault door open*] *with some explosives*
4. *John finds* $[_{SC}$ *Mary picking the lock at the bank vault*] *at night*
5. *John wants* $[_{SC}$ *Mary recruited for the next job in the financial district*
6. *John sees* $[_{SC}$ *Mary arrive at the bank yesterday*]
7. *John trains Mary* $[_{SC}$ *to break into a bank vault with decoding equipment*]
8. *John finances Mary* $[_{SC}$ *to carry out the operation of a series of heists around the country*]

For examples (1), (2), and (3) , the constituent marked *SC* has approximately the same semantics if the non-finite copula *'to be'* is introduced between the two arguments, as in *'Mary to be a good partner'* or *'the vault to be open'* . These clauses contain no inflectional information, and their would-be predicates, contribute no significant semantic information toward the frame, and can thus be treated as predicate copula embedded clause. The difficulty in deciphering the type in (1-3) above resides in the variation that most syntactic parser treats these structures, which often produce anomalous structures which are difficult to recognize as a complete frame. Often the parser is unable to analyze the example similar to (3) beyond the chunking level, and the entire structure between the matrix S and $[_{NP}$*the vault door*] and $[_{ADJP}$*open*] would remain unanalyzed. There are similar cases where the the constituents [$_{VBZ}$*blasted*], [$_{NP}$*the vault door*], and [$_{ADJP}$*open*], are partially analyzed, such that all three constituents are direct descendents of the matrix VP. For this class of cases, since the syntactic parse does not offer a hierarchy within the matrix VP, and the flat structure within contains at least two constituents that appear to be internal arguments of the matrix-S, one of the which is a nominal, while the other would be headed by and ADJ, PREP, VBG, etc. After reanalysis, the samples with matrix verbs outside of the ECM generally has the NP representing the argument between the matrix and embedded predicates to be structured with the matrix clause, but leave PRO in the embedded structure, as we can observe in (2) and (3).

We need to utilize our knowledge-base of the possible set of valence structures, to see whether the matrix-V in question permits the aforementioned two-internal argument sequence. Another variation of inadequate parsing appears frequently as in (2), where the parse may be *John* [$_{VP}$ *wiped* [$_{NP}$ [$_{NP}$ *the table*] [$_{ADJP}$ *clean*]] (possibly due to the non-productive structures that could be analyzed this way such as *'courts martial'*). These generally need to be deconstructed from their parser produced structure, and reanalyzed as the previous type, if they are to be correctly analyzed, since in their original parsed forms they do not convey an embedded clause; thus cannot be used to extract the embedded frame directly.

Examples (4 - 6) contain an non-INFL form of the predicate verb of the embedded frame, and are usually parsed so that the embedded clausal structure is evident. There is some *aspect* information but not tense, that is conveyed through the morphology of the verb; although for a limited number of examples, such as those SCs with a specific subset of embedded predicates such as *'arrive'*, *'leave'*, etc has a main verb with less clear interpretation, and can occur in what appears to be non-finite forms, but is most frequently tagged as VBP. These have the external argument position filled in the embedded clause by an entity, such as *"John sees* [$_{S_{emb}}$ *Mary* [$_{VP}$ *picking* [$_{NP}$ *the lock*] [$_{PP}$ *at the bank vault*]]]*"*. In terms of the semantic relation between the matrix and embedded clauses, in most instances, these type can easily be interpreted as the entire embedded frame being an argument of the matrix frame, where the ECM verb of the matrix frame also takes an

SBAR as well as SC argument with approximately the same semantic content, such as in the following pair meaning roughly the same set of events:

i *John observes Mary leave the bank at night*

ii *John observes that Mary left the bank at night*

So for these subtypes, it is usually appropriate to equate these forms with embedded SC with a form that contains an embedded SBAR with the same embedded predicate and argument list.

Examples (7) and (8) also contain some non-finite form of the predicate verb within the embedded structure, and are usually parsed so that the embedded structure is presented as S but not SBAR. Such samples contain an infinitival, and are generally postulated to have a cover *PRO* residing in its external position; and thus any argument manifested in the linear sequence between the matrix and embedded predicates cannot be part of the embedded frame, but must be part of the matrix structure. The matrix V permitted in this subtype also reaches beyond ECM verbs, including a variety of semantic classes. Each sample has a set of essential components and clausal structure, except the external argument which remains covert; thus the frame structure can be elucidated once the entity represented by the external argument is identified.

5.4.4.4.3 Semantic interpretations We can see that at a deeper semantic level for the embedded frame, this class include both of so-called *complement small clause*, and *adjunct small clause.* The former interpretation considers the entire event described in the embedded structure to be an argument of the matrix predicate, while the latter interprets the internal argument of the matrix predicate to be identical to an argument in the embedded frame.

A *Jack considers the meat raw*

B *Jack eat the meat raw*

C *Jack sees Jane committing the robbery*

D *Jack ignores Jane to commit the robbery*

E *Jack deceives Jane to commit the robbery*

F *Jack convinces Jane to commit the robbery*

For (A) and (B) above, we observe that whether or not the matrix verb belongs to the ECM class determines whether the semantics is interpreted as *complement SC* (A) or *adjunct SC* (B), for the type of surface form where the overt embedded verb is absent. We can analyze the structures as *Jack considers* $[_{SC}$ *the meat raw*] and *Jack eat the meat* $[_{SC}$ *PRO raw*]. The first is semantically similar to *'Jack considers it a fact that the meat is raw'*; while the second is similar to *'Jack eats the meat which is raw'*. In both of these scenarios, the structure of the embedded frame remains the same, while the matrix frame's internal argument would be an event in the *complement* case, and an entity in the *adjunct* case.

For those surface structural types that contain some form of embedded predicate verb, the situation is somewhat analogous. If the matrix verb is ECM in these cases, then the SC is almost certainly interpreted as *complement* by the speaker. When the matrix verb is outside the ECM class, the SC is always interpreted as *adjunct* by most; but unlike the previous *adjunct SC* examples where the overt embedded verb is absent, such as (B), the determination of the meaning is far more complex, principally with the identification of PRO in the embedded clause. For examples (A-D), each is intrinsically ambiguous in their syntacto-semantic structure, such that the PRO in each may be coindexed with either the external or internal argument of the matrix S. The meaning of each interpretation could be paraphrased as the following:

- a *Jack ignores Jane (as to not heed her warning) so that Jack carries out the robbery*

  b *Jack ignores Jane (as to passively turn a blind eye to her plans) so that Jane carries out the robbery*

- a *Jack deceives Jane (as to conceal his plan) so that Jack can carry out the robbery*

  b *Jack deceives Jane (as to trick her into unknowingly act) so that Jane carries out the robbery*

- a *Jack convinces Jane (as to refrain from interfering) so that Jack can carry out the robbery*

  b *Jack convinces Jane (as to participate) so that Jane carries out the robbery*

For (D), the interpretation [a], which has the matrix external argument coindexed with the embedded PRO is far more dominant, (it is questionable whether [b] is allowable interpretation at all); while for (F), the interpretation [b], which has the matrix internal argument coindexed with the embedded PRO is dominant instead; for (E), both interpretations are very plausible, and neither is overwhelmingly the favored choice. So this subtype of SC constructions have the external argument of the embedded frame that is generally ambiguous, and the combination of syntactic and $\theta-$structure related information alone cannot provide a definite answer. More information is needed from discourse and possibly finer semantic classifications of dyadic verbs would be necessary.

5.4.4.5 R-extraposition

Extraposition of constituents generally involve the repositioning of "heavy" elements to the right of the canonically ordered constituents of a clause. This is a class within which is much variation, and some cases will be mostly irrelevant to our goal of frame extraction, some will be very difficult to extract correctly because of ambiguity at a lexico-syntactic level, while others can be extracted with relatively good confidence and would be useful in correct understanding of frames.

5.4.4.5.1 Basic types It is generally assumed that right extraposition of maximal projections in language is associated with the scarcity of short-term memory capacity, and the need for a more $L \longleftrightarrow R$ balance in the S-structure. The distinguishing characteristic here is that the governor of the moved category disagrees with its head in S-structure. There are a number of different types of extrapositions in linguistic theory, the following describe the most prevalent types:

1. *Jack gave to Jane, due to his affection for her,* ***some honey that he personally collected from a beehive***

2. *Jack allows Jane to bring her laptop to the mansion* ***which her parents purchased for her as a graduation present***
3. *Jack brought a bicycle to Jane* ***as her transportation on campus***
4. *Jack introduced a intern sales-woman to Jane* ***as her liaison at the firm***
5. *Jack draws a portrait on the blackboard* ***depicting Jane's facial expression in great detail***
6. *Jack suggested it to Jane* ***that she move back to a west coast location for the summer for surfing lessons***
7. *Jack brought it about in time* ***that Jane took part in the heist at downtown***

From above, (1) is an example where a single maximal projection, corresponding to exactly one essential frame component, is extraposed. We can unambiguously see that this is extraposition, separated from its governor *'gave'*, and not simply shifting within the S, because of the observed clausal adjunct that intervenes between the main clausal sequence and the constituent in question. (2) is the extraposition of a relative-clause that describes the underlined argument in the canonical order, so the core component of the argument did not move, but only its adjunct; the relative clause can be analyzed independently, and then the identity of *which* can be coindexed with the element within the matrix frame. (3/4) also has the same structure, except with the extraposed adjunct to the argument being a PP. (5) is a case where there is a VP headed by a gerund, and in fact should semantically behave as an SC '[$_S$ [$_{NP}$ *a portrait*] [$_{VP}$ *depicting* [$_{NP}$ *Jane's facial expression*] [$_{PP}$ *in great detail*]]]' . (6) or (7) is the an *'it'*-extraposition, which behaves similarly to (2) in that the core element (in this case a PRP) of the argument remains in situ, while its adjunct moves to the right.

Examples where the extraposed constituent is adjunct to one of the essential arguments of the frame, such as (2/3), are frequently ambiguous if the original frame has more than one internal argument; and these are usually not essential for arriving at a correct interpretation of the frame structure. We can see in (4) that sometimes, even for a human speaker, it is necessary to utilize information in the discourse context, in order to associate the adjunct such as *'as her liaison at the firm* with the appropriate argument. Even an example such as (2), which most individuals would be able to discern the association of *'graduation present'* to *'laptop'* not to *'mansion'*, but requires extra-linguistic information that are contextual and cultural. For examples such as (5), there are essentially two frames, sharing the argument *'a portrait'*. That fact, and the lack of information for the frame structure, from the adjuncts in (2-4), allows us to avoid these types. The only benefit in analyzing this is to elucidate the indexing of the WH-element of an extraposed relative clause, such as in (2), which then aids in the analysis of that embedded frame (of the adjust relative clause, not the matrix clause). An *'it'*-extraposition like (6/7) is something that cannot be dismissed, because that the matrix clause's frame does not have a complete set of components; the *'it'*-argument conveys no information about the event, state, or concept that the extraposed constituent represents, it is simply a device for indexation. So we must analyze and transform it back to its pre-extraposed form, in order to have full set of frame components.

5.4.4.5.2 General procedure We will limit ourselves to those relevant for frame structure, those that involve a single constituent that corresponds to some essential component of the frame. The baseline detection mechanism for R-extraposition can be formulated as:

$$\left\{\begin{array}{l}
\mathcal{F}^{\mathcal{X}}(A_{\langle q_0, Z\rangle}, \alpha^0, \Sigma_A^{+\Psi_A}, \Gamma_A^{+\Phi_A}, T_s \in \\
\qquad \mathcal{F}'_{\blacktriangleleft}(A_{\langle q_0, Z\rangle}, \alpha^0, \Sigma_{VP}^{+\Psi_{VP}}, \Gamma_{VP}^{+\Phi_{VP}}, \\
\qquad\quad T_t \in N^-(T_s) | T_s \in \mathcal{U}, \varsigma(T_t) \in \Sigma_S)\ ) \\
\qquad \mathcal{F}'_{\blacktriangleright}(A_{\langle q_0, Z\rangle}, \alpha^0, \Sigma_{DJ}^{+\Psi_{DJ}}, \Gamma_{DJ}^{+\Phi_{DJ}}, \\
\qquad\quad T_t \in N^-(T_s) | T_s \in \mathcal{U}, \varsigma(T_t) \in \Sigma_s) \\
\qquad \mathcal{U} = \mathcal{F}^{\int}(A_{\langle q_0, Z\rangle}, \alpha^0, \Sigma_S^{+\Psi_S}, \Gamma_S^{+\Phi_S}, T_r \in \mathcal{F}^C(T)\ ) \\
\quad \Big| \quad \forall L_j \in \mathcal{A}'(V_{mtx}) \qquad \left[\mathcal{A}'_{[i,j]}(V_{mtx}) \neq A_i\right]
\end{array}\right. \tag{5.15}$$

where the parameters $\langle \Sigma_{S2}, \Gamma_{S2}\rangle$ are previously mentioned, and pertains to IP structures. The parameters $\langle \Sigma_{DJ}, \Gamma_{DJ}\rangle$ the set of symbols that can be adjunctive to NP, these would be pushed on once maximal projection of such a phrase is seen, and pop when the head of such a phrase is read. The parameters Σ_{VP} := the set of symbols that correspond to the set of VP-heads and equivalents, which is pushed on when the maximal projection of such is seen, and popped off when a head of content V terminal is seen. The parameters Σ_A := the set of symbols that consists of heads of argument constituents, when maximal projection of of argument constituents, whose detection pushes it on, and is popped off once the head of such constituents is seen.

5.4.4.6 Internal argument shifting

Shifting is in many ways similar to extraposition, in terms of some maximal projection of an argument moves to a non-canonical position; and generally in shifting, this involves the entire argument's surface form. The major difference here is that the governor of the moving constituent agrees with its head in the S-structure syntactic tree. (Hence it is also not subject to considerations on subjacency) Due to the fact that it is bounded by the maximal projection of its governor, the general result is that the moving constituent is not moved away from the rest of the frame components, but still forms a continuous surface form with them, only the ordering of the components may be changed from the canonical order. As the name of the class indicates, it generally shifts some argument constituent to a different position at surface:

1. *Jack allows Jane to bring to the mansion* ***her laptop***
2. *Jack brought to Jane* ***a bicycle for transportation on campus***
3. *Jack draws on the blackboard* ***a portrait depicting Jane's facial expression in great detail***

And we can see that, multiple movements of this type is generally not permitted, and thus such case can be ignored.

i *Jack exchanged with Jane a case of smuggled diamonds for a bag of money from the bank heist in Canary Wharf*

ii *Jack exchanged a case of smuggled diamonds for a bag of money from the bank heist in Canary Wharf* ***with Jane***

iii *Jack exchanged a case of smuggled diamonds* ***with Jane*** *a for a bag of money from the bank heist in Canary Wharf*

iv *Jack exchanged a case of smuggled diamonds for a bag of money from the bank heist in Canary Wharf* ***with Jane*** v ?* *Jack exchanged for a bag of money from the bank heist in Canary Wharf* ***a case of smuggled diamonds*** *with Jane*

vi * *Jack exchanged for a bag of money from the bank heist in Canary Wharf* ***with Jane*** *a case of smuggled diamonds*

The baseline detection mechanism for shifting is similar in form compared to R-extraposition, and provided that the parser output indicates that these constituents are in VP-adjunct position at S-structure. with the same basic movement pattern as well as the same mismatch in adicity conditions. The shifted component can always be formulated as a movement of some internal argument toward the right (whether or not the the direction is right at a theoretical level). The difference in mechanism accounts for the lack of need to search in positions that are IP-adjunctive. This baseline algorithm can be formulated as the following:

$$
\left\{
\begin{array}{l}
\mathcal{F}^{\mathcal{X}}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_A^{+\Psi_A},\Gamma_A^{+\Phi_A},T_s \in \\
\qquad \mathcal{F}'_{\blacktriangleleft}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{VP}^{+\Psi_{VP}},\Gamma_{VP}^{+\Phi_{VP}}, \\
\qquad\quad T_t \in N^-(T_s)|T_s \in \mathcal{U},\varsigma(T_t) \in \Sigma_S)\ ) \\
\qquad \mathcal{F}'_{\blacktriangleright}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{DJ}^{+\Psi_{DJ}},\Gamma_{DJ}^{+\Phi_{DJ}}, \\
\qquad\quad T_t \in N^-(T_s)|T_s \in \mathcal{U},\varsigma(T_t) \in \Sigma_s) \\
\qquad \mathcal{U} = \mathcal{F}^C(T)\ ) \\
\quad \Big| \quad \forall L_j \in \mathcal{A}'(V_{mtx}) \qquad \left[\mathcal{A}'_{[i,j]}(V_{mtx}) \neq A_i\right]
\end{array}
\right. \tag{5.16}
$$

All parameters are similar to previous discussion's. For various subtypes, these will need to be adjusted accordingly.

5.4.4.7 Elliptical constructions

Ellipsis is a phenomenon that generally involves two or more adjacent clauses sharing some constituent within the IP at a deep level, which is necessary to complete the frame of each. This class of transformations require either a duplication of identical constituents at two locations at LF, or this sharing phenomenon can be expressed through expanding the context free structure into a DAG. Depending on the parts of the syntactic trees that are identical and thus shared, and the constituents that remain distinct and thus present in the

S-structure form of each class, there is a variety of different types within the class.

5.4.4.7.1 Basic types

There are a wide variety of possible elliptical expressions. Some of which are not directly applicable to our frame-extraction task, such as ellipses in Q&A, which is limited to a very narrow format of discourse. Other types are very difficult to analyze correctly without additional resources from supervised learning, such as nominal ellipsis which requires highly accurate NER for covert nominal constituents; the need to determine the exact antecedent, sometimes over very long linear distance, makes accuracy here impossible. So we will focus on a few types that are the most doable as well as useful, as in the following:

1. *this semester, Fred will apply to grad school in philosophy, and Hollie _ in music*
2. *Fred plans to search for the lost treasures in Bolivia, and Hollie _ the ancient scrolls in Crete, on the other hand*
3. ?* *Fred made the preparations so that he will find a way to rob the bank at midnight, and Hollie _ in the morning*
4. *Fred is able to succeed in his major course-work, and _ in his extracurricular activity as well*
5. *Hollie has been working on her graphic design portfolio, and Julia _ also*
6. *in the fall, Fred will perform at the local music festival, Hollie will _ too*
7. *The magician and his assistant planned to _, and they did perform the grand illusion that night*

From examples above, each example contains _ as locations of constituents that are not overtly expressed. (1/2) are examples where some parts of the syntactic structure commanded by the matrix SPEC-element, including the heads of all VPs in the intervening sequence, but having some argument or adjunct to the innermost VP being overt; this is typically termed *gapping* in coordinate structures; sentential adverbial adjuncts of each clause may be present, but does not affect the analysis of the clausal pair; these may include *'too'*, *'in addition'*, *'on the other hand'*, etc. It can be seen in (2) that this intervening covert sequence can cross an IP boundary, so that only the matrix external argument and an argument in the embedded SC remain overt in the surface form. As the intervening sequence begin to cross CP boundary and have more clauses included, as in (3), the grammaticality of the gapping construction decreases dramatically.

The examples (4/5) above are examples where the second clause has only one argument, and some sentential adverbials such as *'also'*, *'similarly'*, *'as well'*, etc that express close comparison between the semantics of the two frames, are usually present. The parts of the surface form related to TAM, predicate verb, and the remaining arguments are not overtly present. This is usually termed *stripping*, and the semantics of the second clause basically entails everything in the first clause with that cover argument replaced.

For (6/7) above, the external argument and the TAM elements are present in the surface sequence, but the inner-most VP of the clause with the main predicate and essential internal arguments are missing. This type is the *full VP-ellipsis*. Its structure is somewhat easier to analyze, since the elided part of the clause is a single constituent; sequentially it is also cleaner than the other variant, with only adverbial adjuncts that can appear in positions after the TAM elements. The complication in this case comes from the fact that either the first or the second clause can be the one elided; and the only way to give a full treatment for this type is to test for ellipses in both directions.

5.4.4.7.2 General procedure One of the issues of execution is that the second clause in the pair is often not parsed correctly, due to it missing some over elements; this especially occurs frequently when the predicate-V is missing from the surface form. The parsed example *“*[$_S$ [$_{NP}$ *John*] [$_{VP}$ *bought some peaches*], [$_{CC}$ *and*] [$_{NP}$ [$_{NP}$ *Mary*] [$_{NP}$ *some pears*]]*”* , where the second S is incorrectly parsed into a compound noun consisting of *'Mary some pears'*. These types of mistaken parsing needed to be corrected before any further analysis and processing can occur. The general strategy is to locate the constituents that are overt in the second clause, and replace the corresponding constituents in clause 1 to form a semantically relevant form of clause 2. Since there is general some sequence of structures missing in the middle of the second clause's parse, we look for the overt structure sequentially before (usually external arg and/or TAM) and after (usually some subset of internal arguments); with the precondition that either the length of the argument list in clause 2 is insufficient, or that the predicate itself is not overt.

For here, we will limit the formulation to two clauses in such relations, for the two trees T^1 and T^2. In general, there may be permitted a constant maximum number of clauses in a single elliptical construction, up to the maximal arity of the trees permitted. The generalized baseline detection mechanism for elliptical expressions can be formulated as:

$$\left\{\begin{array}{l}
\mathcal{F}'(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{IN}^{+\Psi_{IN}},\Gamma_{IN}^{+\Phi_{IN}},\\
\qquad T_t \in N^-(T_s)|T_s \in \mathcal{U},\varsigma(T_t)\in\Sigma_S)\ )\\
\quad \mathcal{F}'(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{EN}^{+\Psi_{EN}},\Gamma_{EN}^{+\Phi_{EN}},\\
\qquad T_t \in N^-(T_s)|T_s \in \mathcal{U},\varsigma(T_t)\in\Sigma_s)\\
\quad \mathcal{U} = \mathcal{F}\checkmark_{\forall}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_S^{+\Psi_S},\\
\qquad \Gamma_S^{+\Phi_S},T_r \in \mathcal{F}_{\forall}^{S}(a(T^1,T^2))\ )\\
\\
\quad \Big| \quad \Big[\Big[\exists L_j \in \mathcal{A}'(V_{mtx}) \quad [\forall L_p^1 \in \mathcal{S}(T^1)|L_p^1| \geq |L_j|\Big] \bigwedge\\
\qquad \Big[\forall L_j \in \mathcal{A}'(V_{mtx}) \quad [\forall L_p^2 \in \mathcal{S}(T^2)|L_p^2| < |L_j|\Big]\Big] \bigvee\\
\qquad \mathfrak{o}(\mathcal{F}^{\int}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_V^{+\Psi_V},\Gamma_{AM}^{+\Phi_{EA}},T_s \in \mathcal{F}^S(T^2)\ )\ )
\end{array}\right. \tag{5.17}$$

The parameters $\langle \Sigma_{EN}, \Gamma_{EN}\rangle$ and $\langle \Sigma_{IN}, \Gamma_{IN}\rangle$ generally correspond to the set of possible external and internal arguments; each pushes onto the stack once the machine detects the appropriate maximal projection for the argument, and pops off once the appropriate nominal is detected. These need to be tuned carefully for each type and subtype among elliptical constructions.

The parameters $\langle \Sigma_V, \Gamma_{AM}\rangle$ is used to detect whether there is any content-V that can serve as the predicate (all verbs except auxiliaries and modals); either the absence of the content-V, or having less arguments in the clause than permissible according to the predicate (when a content-V is present). For the stripping type constructions, since there is normally one of a small set of aforementioned sentential adverbials expressing similitude between the frames present, the mechanism is made more accurate by adding the detection of these at a layer above that of the external nominal argument ($\langle \Sigma_{EN}, \Gamma_{EN}\rangle$). These sets of parameters will be Σ^S_{ADV} and Γ^S_{ADV}, each pushes on a set of possible symbols when it sees ADVP, and pops off the stack with the detection of one of the limited set of RBs that can work as aforementioned sentential adverbials for stripping, such as *'too'*, *'also'*, *'well'*, etc. For the VP-ellipsis variant, we must also allow T^1, T^2 to be interchangeable, and try both possibilities, since the elided components can occur in either clause 1 or 2.

5.4.4.8 Coordination

Coordination occurs in parallel structures that are conjoined by explicit or implicit means. Coordinations potentially have some portion of the structure independent to each clause (tied with conjunction if explicit coordination), and portions that are independent; the amount of shared structure can vary from zero to almost all (generally all except one argument or predicate). The types of coordinations we are interested in are those that have explicit conjunctions between the parallel components and sharing of some S-structure in syntax between the two.

5.4.4.8.1 Basic types For coordinate structures that are completely independent clauses, and do not share any constituents within the clauses, there is no need to treat them any differently than two recognizably separate clauses. In the future, when there is a need to analyze social network structure and *cooperative* relations, the information about their logical connection may need to be preserved, but not at the level of individual frames. For those coordination where two clauses are separated, but some constituents are shared within the clauses, these are normally already processed as elliptical expressions in 5.4.4.7.

The type of clause need to be analyzed in this section are single clause, where some components of the frames are shared, while other components are parallel and distinct in the surface form. Some of the typical cases are demonstrated below, where a variety of different configurations are possible here:

1. *the wild foxes and the domestic canines hunted the herds on a cattle range in Montana*
2. *Jason and Melanie performed the grand illusions on the stage of the concert hall*
3. *the wild foxes tracked, hunted, and killed the herd of cattle on the ranch*

4. *Jason studied, mastered, then performed the illusions in the "grand illusions" magic company*
5. *the wild foxes hunted the herd of cattle as well as the free-range horses in Montana*
6. *Melanie masterfully performed at a jazz concert, in a Broadway play, and as a magician's assistant*
7. *the foxes of continental Europe and of North America hunted the livestock of the ranchers.*
8. *Melanie took part in the magic performances in the Bellagio, the Venetian-Palazo, and Caesars Palace of the Vegas strip*

For examples (1/2) above, the external argument is the locus of the coordinate structure; for (3/4) the coordinate structure is that of the predicate verbs, and both the structure above and below the VP are shared; for (5/6) the coordinate structure is located at one of the internal arguments. The examples (7/8) illustrate a situation where the coordinate structure is actually located within one the subtree corresponding to one of the arguments, but not the head of the argument-constituent itself.

5.4.4.8.2 General procedure The generalized structure of such coordinations is that there are portions of the structure at the highest levels that are shared among frames (always have the CP and IP structures shared); and often have some of the deepest structures shared between them as well (always have some internal argument shared, unless it is a monadic frame, or internal argument coordination); but there exists some portion of the frame in the intervening portion that exists independently for each frame, and is part of the actual syntactic coordination.

The generalized baseline detection mechanism for elliptical expressions can be formulated as the following; where the individual parameters are defined according to each subtype:

$$\begin{cases} \mathcal{F}'_{\forall}(A_{\langle q_0,Z\rangle}, \alpha^0, \Sigma_{CCT}^{+\Psi_{CCT}}, \Gamma_{CCT}^{+\Phi_{CCT}}, \\ \qquad T_t \in N^-(T_s) | T_s \in \mathcal{U}, \varsigma(T_t) \in \Sigma_S) \\ \quad \mathcal{F}^{\int}(A_{\langle q_0,Z\rangle}, \alpha^0, \Sigma_{CNJ}^{+\Psi_{CNJ}}, \Gamma_{IN}^{+\Phi_{CNJ}}, \\ \qquad T_t \in N^-(T_s) | T_s \in \mathcal{U}, \varsigma(T_t) \in \Sigma_S) \\ \quad \mathcal{F}'(A_{\langle q_0,Z\rangle}, \alpha^0, \Sigma_{PRE}^{+\Psi_{PRE}}, \Gamma_{PRE}^{+\Phi_{PRE}}, \\ \qquad T_t \in N^-(T_s) | T_s \in \mathcal{U}, \varsigma(T_t) \in \Sigma_s) \\ \quad \mathcal{U} = \mathcal{F}\surd(A_{\langle q_0,Z\rangle}, \alpha^0, \Sigma_S^{+\Psi_S}, \\ \qquad \Gamma_S^{+\Phi_S}, T_r \in \mathcal{F}_{\forall}^{S}(T)\ ) \end{cases} \tag{5.18}$$

The parameters $\langle \Sigma_{PRE}, \Gamma_{PRE} \rangle$ is designed to detect those frame components' possible tags that are possible above the parent of the conjunction, it varies according to the subtype of single-clause coordination. The parameters $\langle \Sigma_{CNJ}, \Gamma_{CNJ} \rangle$ is used to detect the limited set of conjunctions themselves. $\langle \Sigma_{CCT}, \Gamma_{CCT} \rangle$ is designed to detect those arguments (or other useful constituents) within each conjunct subtree of the coordination. For certain subtypes, there will be more machinery required to resolve further complications, as described next.

5.4.4.8.3 Structural complexity The actual coordination structure is normally headed by a conjunctive adverbial, such as *'and'*, *'then'*, etc. A second point of departure exists for the shared structure within the coordination, within which is a second shared portion of the syntactic structure between the frames. Again using the example *'Melanie took part in the magic performance* [$_{PP}$ *in* [$_{CRD}$ [$_{NP_1}$ *the Bellagio,* [$_{NP_2}$ *the Venetian-Palazo*], [$_{CONJ}$ *and*] [$_{NP_3}$ *Caesars Palace* [$_{PPX}$ *of the Vegas strip*]]]] *'*. We will term the coordinated structure CRD, which is an NP inside a PP-argument. Within this coordinate NP there are three conjuncts, NP_1, NP_2, NP_3; one of these the PP *'Caesars Palace of the Vegas strip'* actually contains a second PP (tagged PPX) that should be shared among the three coordinated NP entities, where all three of these casinos are in fact described as being *'on the Vegas strip'*.

Here we have a three-fold division of the structure-space within the IP: the structure beyond the parent of the conjunctive which is shared among the frames; the coordinate structure directly commanded by the conjunctive (minus the inner shared portion); and some innermost portion that is again shared among the coordinated subtrees. The outer and the coordinated portions should always exist among all subtypes, the inner shared portion exists in only some subtypes. The boundary between the outer shared and the coordinated structures is detected through the presence of the conjunctive; while the boundary between the coordinated structure and the inner shared is detected through examining whether one of these parallel conjunct subtrees has extra components than the rest. In the example case, the conjunct *'Caesars Palace of the Vegas strip'* contains an extra subtree of a locational PP. It is actually not certain that this inner PP is indeed shared; and could be that only *Caesars Palace* is on the strip; but generally for this type of NP, especially when the extra component is on the final conjunct, the extra structure is shared. So we adopt this heuristic that if such extra structure occurs within the sequentially final conjunct of the coordination, we assume that the structure is shared. Otherwise, when this extra set of components occurs within one of the conjunct subtrees that is not sequentially final, then it is assigned to the frame corresponding to that conjunct alone.

Particularly with the coordination that occurs within the external argument, there is an additional concern for how the semantics of the individual frames fit together, in terms of the *cooperativity* among the multiple entities. This regards the way a set of frames reflect single or multiple events in the real-world (where individuality of events is not black and white). These can correspond to *distributive*, *collective*, and *cumulative* type of cooperative relations, and the classification of these will be a future module of the study, beyond mere frame semantics. The examples involving external argument coordination is purposely designed to be ambiguous in whether each should be interpreted *distributively* or *collectively.*

5.4.5 Entity information

Individual nominal constructions within these surface forms often contain complex internal structures. One type of adjunct, the relative clauses where the relative pronoun is coindexed with the nominal head of the NP is taken care of given the recursive design of any clausal embedding in the corpus. The remaining structural complexity is analyzed through similar automata mechanisms as in movement. These extracted NP internal structures are expected to be a significant part of the eventual set of entity-relations that are present in any social network corpus, and contribute to its graph theoretic representation.

5.4.5.1 Common relational types

Excluding what occurs inside any embedded clauses that are adjuncts to NPs within arguments, the vast majority of NP-internal structures that could contain information about other distinct entities, basically fall into three categories, with respect to their lexico-syntactic structures. One type is that of *genitive*-case construction, which appears to be a PRP$ tag for a pro-form of that function, or a pair that has a nominal followed by a special POS token. The structure inside such a generative nominal is usually limited in length, but sometimes can be structurally complex and could even itself contain frames, such as *"the man whose dog destroyed the garden and ravaged the flower-bed's house"*. The relation it actually represents is difficult to determine:

1. *the rich lawyer* $[_{WHCP}$ *that works for the biggest firm in the city*$]$*'s case*
2. *the businessman* $[_{WHCP}$ *who just made a series of major deals*$]$*'s real-estate property*
3. *the skilled acrobat* $[_{WHSC}$ *performing the most stunning aerial maneuver*$]$*'s performance company*
4. *the girl* $[_{WHCP}$ *who got into the top universities of the country*$]$*'s recent scores*

Each of the above examples contains a syntactically similar form of genitive construction, but has a semantically distinct type of relation between the base entity/object and the modifying genitive. (1) here connotes a working relationship of the lawyer on a task, similar to *subject matter*; (2) connotes a possession relation of the real-estate property, which is possibly the most common type; (3) connotes a notion of membership of the individual acrobat in the performance group; and (4) connotes some external description, in the form of academic record, on the girl entering university. Each of these subtypes, and many others, require detailed knowledge of the ontology of the real-world and current culture to fully and correctly elucidate, such as the *'score'* here likely refers to a performance metric in the girl's course-work, rather than some musical piece that she has composed, which has an impact on the type of relation. So these variation in semantics of this morphosyntactic structure is largely dependent on real-world referents and pragmatics.

The second type is that of *endocentric* compositional nominal, a subclass of compound noun where the head of the compound contains the baseline category for the entire structure, such as *'bus station'*, *'poll station'*, *'titanium bracelet'*, etc. This is the most prevalent type of compound nominal, and is the only type where its semantics can be extrapolated with only lexical semantics semantics from its components (*exocentric* compound nouns has meaning beyond some combination of its lexical semantic parts). Although compound nouns have a great deal of linguistic theory associated with it, on a practical level, one may generally view a two-part endocentric nominal $\langle W, V\rangle$ set-theoretically, such that $V \in 2^U$ is a subset within the relevant universal set U that contains the base-semantics of the compound. Then the set $\mathbb{R}_{U\cap T}(W) \in 2^{U\times(U\cup T)}$, where $T \supseteq \varnothing$ (W may not be $\subseteq U$) are entities that may be above and beyond U, where $\mathbb{R}_{U\cup T}(X) := X \times Z \subseteq U \cap T$, is the set of relations involving $x \in X$ whose nature is unspecified. The overall set theoretic structure of $V_W(x) := x \in V \wedge x \in X \mid \mathbb{R}_{U\cup T}(W) = W \times X$. There is a wide variety of possible relations denoted by this type of construction, such as:

- *railway station* Relation: W in service of V
- *shooting range* Relation: W is the location of where V takes place
- *course outline* Relation: W is a description of V
- *news cycle* Relation: W is a conceptual organization of (the information in) V
- *paleontology curriculum* Relation: W is the subject matter of V
- *justice league* Relation: W is the motivation for the existence of / the goal of V

among many types. To attain the semantic granularity of these types of relations, which is similar to the granularity of the types of cognitive linguistic definition of semantic frames, it is not possible to individually classify these accurately without a significant series of study of its own merit. Such a study would be very beneficial for completing the relational component of a graph representation of the entity network, and may be undertaken at a future point.

A third common type is a nominal modifier that is consisted of a PP, whose complement is itself a nominal entity. The relation is easier to define than the previous two, because some semantic content about the nature of the relation is carried in the preposition itself (although within each preposition type, some degree of relational ambiguity still exists). Extracting each complement nominal of a NP modifying PP, such as *'Edinburgh'* within '$[_{NP}$ *the magistrate* $[_{PP}$ *from Edinburgh* $]$ $]$', and its associated PP-head *'from'*, would give one a good idea that this is a locational entity likely having the relation of *source* with respect to the entity *'magistrate'*.

5.4.5.2 Generalized mechanism

The baseline detection algorithm for this class given the input T_n, that is an NP-subtree corresponding to an argument found in the overall structure is formulated as the following, for the genitive type nominal modifier:

$$\begin{cases} \mathcal{F}'_{\blacktriangleleft}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{BB}^{+\Psi_{NN}},\Gamma_{NN}^{+\Phi_{NN}},T_u \in \\ \quad \mathcal{F}'(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{POSS}^{+\Psi_{POSS}},\Gamma_{POSS}^{+\Phi_{POSS}},T_t \in \\ \quad \mathcal{F}^{\int}_{\blacktriangleright}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{NN}^{+\Psi_{NN}},\Gamma_{NN}^{+\Phi_{NN}},T_n)\ )\ ) \end{cases} \tag{5.19}$$

where the pair $\langle \Sigma_{NN}, \Gamma_{NN} \rangle$ is designed to detect nominal terminals or nominal maximal projection; and the pair $\langle \Sigma_{POSS}, \Gamma_{POSS} \rangle$ is designed to detect PRP\$ or the special POS token. Next, the formulation of the mechanism for the endocentric compound nominal:

$$\begin{cases} \mathcal{F}'_{\blacktriangleleft}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{BB}^{+\Psi_{NN}},\Gamma_{NN}^{+\Phi_{NN}},T_u \in \\ \quad \mathcal{F}^{\int}_{\blacktriangleright}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{NN}^{+\Psi_{NN}},\Gamma_{NN}^{+\Phi_{NN}},T_n)\ ) \end{cases} \tag{5.20}$$

Then, the formulation of the mechanism for the nominal complement to the PP that is complement or adjunct to a argument nominal:

$$\begin{cases} \mathcal{F}'_{\blacktriangleright}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{BB}^{+\Psi_{NN}},\Gamma_{NN}^{+\Phi_{NN}},T_u \in \\ \quad \mathcal{F}^{\int}_{\blacktriangleleft}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{POSS}^{+\Psi_{P}},\Gamma_{P}^{+\Phi_{POSS}},T_t \in \\ \quad \mathcal{F}^{\int}(A_{\langle q_0,Z\rangle},\alpha^0,\Sigma_{NN}^{+\Psi_{NN}},\Gamma_{NN}^{+\Phi_{NN}},T_n)\ )\ ) \end{cases} \tag{5.21}$$

where the pair $\langle \Sigma_P, \Gamma_P \rangle$ is designed to detect one of a set of allowed prepositions for inter-nominal relations.

5.5 Evaluation and application in overall system

5.5.1 Brief evaluation

To briefly test the preprocessing module, we look at the precision of the output from the frame extraction process. We annotated a randomly selected set of samples from purposefully noisy data that has a variety of surface orderings, from the same dataset of **marriage and relationship forums** in Section 4. The input to this preprocessing module first was lemmatized, tagged, and parsed, then was annotated until there was at least 100 samples that were correctly parsed prior to input into the preprocessor. (where the input to this preprocessor was itself correct). We annotated 105 randomly selected samples from the output of the preprocessor of that corpus.

The fraction of incorrect output of the parser to this preprocessor is 18.4%, where the parser output does not match expected linguistic analysis of the clause. These are mostly a result of incorrect POS tagging, such as in (1) below where *"knows"* is tagged as NNS, resulting in the parsing failure of the entire sentence; or some error in the original writer of that utterance, such as the run-on in (2) below where the run-on sentence by the user between *"USER_S2H16Y get married"* and *"USER_RAJ is not married"*, results the two sentences being parsed as a single frame and results in incomprehensible structure (*USER_S2H16Y* and *USER_RAJ* ends in the same frame with *marry* as the predicate).

1. *... and USER_ANONYMOUS12 knows of many men who are all scamming their women ...*
2. *... and they would get to know each other before USER_S2H16Y gets married USER_RAJ is not married ...*

After we discount those where the input parsed data is incorrect, there are 102 samples annotated that are correctly transformed into tree structures expected from transformational grammar. These are annotated to see whether each of the frames is correct according to a possible outcome of linguistic transformations based on formal grammar.

In order to be correct, the top level frame should have the same *predicate*, *polarity* (presence of negation), all of the *essential arguments* of the frame in the same order, as the a possible grammatical transformation; for any clauses that are embedded, either which are themselves arguments of a matrix frame or is a complement of an argument (e.g. *"the girl [who carried flowers]"* with the relative clause as the NP complement), down to the 3rd level from the matrix clause. For the 102 correctly parsed samples, 92 are correctly transformed (90.2%). Most of the transformation errors are due to issues with argument identification or ordering, such as below:

a *he will let USER_JUST_ME10 down after he gets the citizenship trust*

b *they may want curly hair because they may want to try something other than the dead straight*

In (a) above, the particle is mistaken as the second essential argument of the matrix frame. The matrix frame should have $\phi(ARG_0, ARG_1, ARG_M)$, where ARG_0 would be correctly recognized as *"he"*, ARG_1 as *"USER_JUST_ME10"*, and *ARG_M* as the embedded frame; But the particle *"down"* was tagged as JJ (which is an possible syntactic tag), but the separable verb *let-down* was not recognized as a single predicate; as a result the preprocessor marked *"USER_JUST_ME10"* as an adjunct and placed it after all of the arguments in the transformed structure. Another example of (b) above, the embedded clause *"they may want to try something other than the dead straight"* was not analyzed correctly with respect to the complex argument *"something other than the dead straight"*. It was analyzed that the internal argument in this case was *"the dead straight"* rather than the entire structure, thus moved *"something other"* to the incorrect adjunct position; and the post-transformation structure had those two pieces as separate constituents in different parts of the tree.

5.5.2 Performance as a preprocessor

For this module to behave as a preprocessor, its performance in the system differs from its performance *per se* with regard to correctly producing a structure predicted by theoretic models in generative-transformational grammar. The down-stream modules are primarily tasked with complex pattern recognition, with which we may predict the likelihood that a linear (Chapter 7) or tree (Chapter 8) structure contains characteristics that are likely to be recognized as causal by human subjects. Since such pattern recognition algorithms fundamentally detect whether structures are similar, not whether they are theoretically "correct", we need this preprocessing module to produce similar outcomes when two different S-structures should have the same underlying structure at the syntacto-semantic interface, but the produced structure need not necessarily have the same form as the theoretically predicted structure.

This also means that for the transformations performed by the frame preprocessor which are theoretically correct, it always results in a form that is usable by later modules (since S-structure that should have the same syntacto-semantic structure would map to the same "correct" form). Even for those transformations that are incorrect with respect to the theoretical form at syntacto-semantic interface, there is still a chance that they are usable by the later pattern recognition procedures, as long as the separate S-structures that should map to the same theoretical structure at syntax-semantic interface do map to some output structure that is "incorrect" but are the same. We can illustrate this with a simplistic example drawn from the beginning of this section:

1. *With Tuesday came the agreeable prospect of seeing him again, and for a longer time than hitherto;*
2. *The agreeable prospect of seeing him again came, and for a longer time than hitherto, on Tuesday*

As we have seen, these are largely equivalent structures at the syntacto-semantic level. If they both of their post-transformation structure is identical:

- *[$_{VP}$came [$_{NP}$the agreeable prospect of [$_{SC}$ seeing him again [$_{PP}$ on Tuesday]]]], and for a longer time than hitherto*

Even though this transformed structure may not be "correct" according to rules of transformational grammar, if the module output them as identical or very similar forms, there is still a chance these can contribute to a common pattern if both are in the training set; or if one is a positive sample in training, while the other is in the testing set, then there is a chance that model that the sample in training contributed to would help identify the sample in the testing set.

5.5.3 Utility in downstream modules

The extraction process for the semantic frames has three main types of utilities for the remaining modules of the system. First, it provides the identities of the critical components, mostly the predicate and the essential entity arguments, thus allowing this information to be used in further modules of the Chapters 7 and 8. These specific components of a frame sometimes need specific treatment within the process of determining causality. Second it provides a way to identify those elements that are outside of the normal frame components, such as clausal adverbials, adjuncts, and the general topology of the tree outside the IP structure. The goal of the modules in Chapters 7 and 8 is to find lexico-syntactically complex causality, and sometimes causal structures with more than one link in the chain; there the overall structure and elements outside of the basic clausal form (which encodes the frame components) takes primacy. It would sometimes be advantageous to examine determination of causality with purely those types of information, to the exclusion of frame-internal information, as we will see later. The extraction of frames and its components (predicates and arguments) also plays an essential role in the construction of a graph theoretic representation of a social network (where the corresponding events and entities are key).

Chapter 6

Causal Relations in Language

Causation is a relation that is well known and studied both inside and outside of the domain of linguistics. Originally formalized by Aristotle in the Metaphysics (before 322 BC), the concept *τι ὅ ου κινούμϵνον κινϵί*, or the prime mover, necessitates chains of causation that effects all subsequent events in the system. We know that from the foundational study of metalanguage, where causal relations as expressed in some object language, this class of relations has certain defining properties (Burks, 1951; Simon, 1952). There also have been series of attempts to probabilistically provide a definition of causality using quantifiable concepts (Robins & Greenland, 1989; Pearl, 1999/2000; Tian & Pearl, 2000). These properties include: ❶ the causal expression in language involving two events $e_1 \xrightarrow{caus} e_2$ denotes the logical relation between the extension of e_1 and the extension of e_2, not their symbolic counterparts. ❷ Causal expression of $e_1 \xrightarrow{caus} e_2$ logically entails the falsehood of the propositions that $\neg(e_2 \xrightarrow{caus} e_1)$ as well as $(\bar{e_2}) \xrightarrow{cause} \bar{e_1}$ (by $\bar{e_i}$ we mean a world scenario in which e_i does not take place); since in ordinary usage of language if e_1 causes e_2, then e_2 cannot cause e_1 (establishing the direction in a link in the chain of causation), and also the absence of the occurrence of e_2 cannot be taken to imply the absence of e_1; such as in *"the force from the push causes the cart to move"* cannot be taken to mean *"if the cart does not move then there is no force pushing on it" (force could be insufficient to overcome friction).* ❸ Although sometimes there is correlation between causality and temporal precedence, the caused event is not always required to succeed the causing event in temporal order; this is especially true when each event in the causal chain occurs over a significant duration of time, such as *"the trip to Mars causes the bone-density of the astronauts to decrease"*.

It is distinct from co-occurrence relations, such that in one of the logical types that can be described as causal, the occurrence of the *cause* is entailed by that of the *effect*, given a particular state of the system. It is not a superset of reciprocals, as reciprocity does not necessarily imply temporal ordering of the relevant events, and that it may (and often does) involve more than two entities. We can easily find examples of reciprocities that are non-causative, and causatives that are not reciprocal:

1. *John and Mary bumped into each other on the running track*
2. *John made Mary kick Patrick in the shin.*

So for either of these concepts, the linguistic expressions of one is not a subset of the expression-set of the other; and thus they are independently defined linguistic concepts. Causation is a concept that requires both some type of atemporal logical relation such as entailment and temporal relation between the contained events.

One of the ways of analyzing surface form that correspond to the notion of causation at a deep semantic level is to utilize syntactically relevant sub categorization frames (Pinker, 1989), which is critical for ditransitive causal constructions, generated within *construction grammar* (Fillmore, 1989; Goldberg, 1995 / 2002; Kay & Fillmore, 1999; Boas, 2007). The types of causative constructions are a set of related and paradigmatic set of scenes involving real-world actors, (Kodama, 2004) whose encoding could aid in the construction of social network structures. This applies to a whole range of causatives that cover most linguistically expressed causatives, as well as additional linguistic features described in other sections.

6.1 Taxonomy of linguistic causal structures

Linguistically, it is one of the complex features that involves more than one eventuality in the semantics; it contains a number of different subtypes that must be treated distinctly in the text as well as in their logical form. The area of language meaning that normally falls under "causation" often also include the associated logical relational types of *permission* (Wierzbicka, 1998 / 2002) and *enabling* (Wolff et al, 2002; Wolff & Zettergren, 2003; Wolff 2007;), which connotes different environment surrounding the real-world event, but present a similar linguistic structure to standard causative constructions. The construction should contain a minimum of two events by definition, one where a cause acts on an effector, and another where the result is effected; it also arguably contains a minimum of three events (Mandelblit, 1997), within certain typological frameworks.

An essential role of a linguistic causative is to mark the structure for an increase in *valency* (Mandelblit, 2000). In most instances, this would mean an additional entity becomes part of the frame structure, with the entity(s) of original eventuality before the valancy increase becoming related in some manner. There are certain types of linguistic constructions which are considered causative, which contains types such as relations between entities, objects, or qualities, such as below:

1. *the salinity of the water made me thirsty*
2. *the obtrusive threshold made the children stumble at the door of the house*

Here a *'causer'* is not a recognized *'agent'* of any accepted definition. One does not become thirsty simply due to the existence of salty water, but rather the event of consuming the saline water is necessary. One also does not tumble due to only the presence of a threshold, but walking over the location of the threshold is required. These appear to be elliptical expressions of causality that does not contradict the nature of causality, but omits a part of the formulation, (Scheffler, 1992[a]) causality itself fundamentally entails some relation between a pair of events e_1, e_2.

6.1.1 Causative manifestations

Causatives in language encode a complex event with multiple constituents, and a difference in the surface structure of different types have real world implications on their semantics and pragmatics. In cognitive studies, these linguistic constructions corresponds to a sequence of events with spatial, energetic, or logical contiguity. (Langacker, 1988 / 1991) Each single cognitively recognized event can be encoded through separate verb, or the entire causative chain can be encoded with a single verb stem, where the adicity of the verb mediates its structure. In its single clause expressed form, the conventional thematic archetype, and along with it the case system, may allow fewer participants than the total number with a causation chain. Cognitively, this over-utilization of the normal capacity of argument structure in language is the reason why there is a large range of variation in the expression of underlying chains of causation. (Langacker, 1991). Langacker illustrates with the common way of French in expressing an underlying causative chain with three entities:

1. *Paul a fait jeter le caillou par la femme*
2. English: *Paul made the woman throw the pebble*

In this case, the French expression incorporated the intermediate participant as a periphrastic PP *par la femme*, while the corresponding English allowed for the analytic construction to directly incorporate both the *causer* and the immediate *agent* to convey the same underlying relational structure. This demonstrates the basic need for additional processing between morphosyntax and semantics, in order to correctly analyze deep semantic relations such as causation.

Causative being constructed through morphological means are well attested cross-linguistically, such as the extensive systems in Matses (Fleck, 2002), Olutec (Zavala, 2002), Hebrew (Saad & Bolozky, 1980, 1984), Korean (Yeo, 2005), and Navajo (Gessner, 2001), and many others. There is some evidence within modern English of some vestige of a morphological mechanism that is no longer productive, such as *short**en***, *deaf**en***, *optim**ize***; although this strategy is no longer generally productive in the language. Old English had a prevalent causative-inchoative-stative triad (Dowty, 1979; Parsons, 1985 / 1990), with some cases of suppletion.

Since causitivization is a valence-increasing operation that may apply to a variety of eventuality types, causatives are not a monolithic set of argument structures at the morphosyntax-semantics interface. Olutec (Maldonado, 2011; Dryer, 1986), for example, contains morphologically distinct constructions for causitivizing underlying intransitive and underlying transitives, it also contains *applicative* morphemes that sometimes used to causitivize certain classes of verbs. It is important to realize that causitivization is a phenomenon that is diverse in both morphosyntax, in terms of the strategies of forming them at the surface, and in semantics, in terms of the logical forms that they represent. We will examine the most important and prevalent variants here.

6.1.2 Causative-inchoative alternation

Causative and inchoative constructions normally occur together in semantically related pairs, and these corresponding pairs are by and large morphologically related, as cross-linguistically attested in languages such as German, Hungarian, Polish, (Piñón, 2001b), Japanese (Yamaguchi, 1998), Russian, German, (Kjell, 2001) etc., where the forms are frequently morphologically related. According to Lexicon Uniformity Principle (Reinhart, 2000 / 2002), there must be a unity in the underlying concept behind both surface constructions, which then can be related using arity operations (Rappaport-Hovav & Levin, 2011). The pair of arity operations are valency-increasing and valency-decreasing, which in some area linguistics are called *causative diathesis* and *recessive diathesis*, producing a pair of thematic representations of the same real-world event. (Sasaki, 1987) This class of verbs is sometimes also known as *unaccusative*, distinguished from *unergative*, in that these contain sometimes covert layers in syntactic structure to allow them to alternate (Hale & Keyser, 2002), It is also at times deems as “lexicalization” of more complex causative structures (Fodor, 1970), although we will see that this is not always an appropriate description.

The ability to form this alternation is licensed by the type of real-world event with is necessary participants, that corresponds to the lexical entry. The typology of the real-world events referred to ultimately determines the types of compatible argument structure at the $\theta-$level, which in turn permits specific lexical entries to undergo this alternation; this concept is referred to as *thematic cores* in the fields of language acquisition and childhood development. (Pinker, 1989; Coppock, 2009) Whenever the inchoative variant is present in the language for a given alternation, the *inchoative* form can be termed *anticausative.* A cross-linguistic study (Haspelmath, 1993) shows that derivation in both the *inchoative* $\longrightarrow$

causative and vice versa are widespread in certain languages.

Where the *inchoative* $\longrightarrow$ *causative* direction is predominant in UG, there is generally accepted framework a triad, where an inchoative verb of the meaning i_p is derived from a stative predicate with a meaning of a theme of the clause being in state s_p through derivational morphology; then a transitive verb of the meaning t_p is in turn derived from i_p through a valence raising operation. (Dowty, 1979) One can think of the *dead* $\longrightarrow_i$ *die* $\longrightarrow_t$ *kill* as an often cited example, where the logical form of *kill* can be expressed as *CAUSE(BECOME(dead))*. Inchoative clauses in many languages also can resort to reflexive construction in syntax (e.g. German). This allows for the construction of the corresponding transitive form without the need to change verb valency. (Holmes, 1999) Some maintains the opposite direction of derivation, that the inchoative variant is the actual derived from the causative-inchoative variant. There is some evidence for this in the higher level of morphological complexity in languages such as French, Italian, and Russian (Koontz-Garboden, 2009; Rappaport-Hovav & Levin, 2011). Other models of derivation also exist, such as both the inchoative and causative form deriving from a common verbal stem.

An causative-inchoative construction contains a minimum of two eventualities according to the basic analysis (Parsons, 1990). In other analyses, it may require three distinct eventualities (Levin & Rappaport-Hovav, 1995; Mandelblit, 1997), with two of the eventualities being the causing event and the inchoative event, while a third eventualities is the representation of the logical connection between the two. In the case of the three distinct eventualities, the theory usually specifies that the causing event would consist of an *agent* actively engage in some activity with some form of mental state assigned to the agent entity, otherwise, there would be little distinction between the causing eventuality and the connecting eventuality, and little typological motivation in separately treating the two.

Cases can be made that the underlying inchoative is monadic or dyadic in terms of existentially bound arguments, whether an underlying *cause* role is presumed to exist independent of the *theme*. Levin & Rappaport-Hovav (1994) sees inchoative as dyadic in structure, with a surface unspecified *causer* argument. They argue that in a construction such as *the tree fell*, it can easily be augmented by a PP anaphore to form *the tree fell* ***by***

itself, and hence that underlying *cause* argument is present at some semantic level, despite not being expressed in the morphosyntax of the clause. Piñón (2001a, 2001b) argued that this does not definitively support a dyadic hypothesis, given a corresponding examples such as *the baby girl stood by herself for the first time*, where the PP is also compatible with non-inchoative intransitive verbs. But his objection maybe due to the thematic ambiguity of the surface form *by itself*, and seems to be restricted to animate subjects, as *the tree stood by itself* does not seem to contain any meaning of causation. He also argued that the felling of the tree in this case is not caused by an act done by the tree itself, but by some external *force*, whether by wind, erosion, disease, human-action, etc. However, this no longer is a problem, if we posit that the inchoative construction contains a deliberate under-specification of the *cause* argument, rather than attributing the *cause* role to the subject by default. This will have implications in the graph theoretic construction of inchoatives and their causative-inchoative derivations, which we will explore in Section 6.2.3.2.

The $\theta-$role *cause* in a causative-inchoative construction does not appear to exhibit selection restriction by the verb with regard to sentience, animacy, or a variety of other nominal classifications (Fillmore, 1970; Hall, 1965; Rappaport-Hovav & Levin, 2011). This can be demonstrated in the following:

1. ***Jack / the lumber company / the circular saw / the force of the motor*** *felled the tree by the mountain-road*
2. ***Jane / the stove / the heat from the gas-range / the act of cooking*** *melted the butter in the pan*

Rappaport-Hovav & Levin (2011) also argue for the lack of selection restriction in the case of the *theme* of the inchoative class:

1. *Antonia broke* ***the vase/the window/the bowl/the radio/the toaster***
2. ***The vase/the window/the bowl/the radio/the toaster*** *broke*

Although this line of argument is less convincing than that for the *cause* argument. Consider one of the clearest cases of causative-inchoative alternation of *fall* and *fell* (historically related in morphology, which since has become non-productive). We see in the following that there is some agreement in the latitude of selection restriction between the two forms:

1. ***The tree / the cedar / the telephone pole*** *will fall*
2. *Jack will fell* ***the tree / the cedar / the telephone pole***

However, this is not true of possible themes of *fell* with other types of real-world referents, as we can see in the following alternations:

1. ***The column / the statue / the building / the termite mound*** *will fall*
2. *Jack will fell* ***? the column / ? the statue / ! the building / ! the termite mound***

We are not encountering a phenomenon of polysemy for *fall*, since most will recognize *a tree falling* and *a building falling* as using the same basic sense of the words and the same motion in the real-world. This is also not a phenomenon specific to this lexical pair, but others exhibit the same pattern of acceptability, such as in the inchoative-causative alternation of *clear*, which contains no consistent pattern of where the selection restriction on the *theme* might be manifest, except appealing to the lexical meaning and its senses:

1. *Jack clear* ***the table / the counter / the room / ! the sky / ! the weather / ! the shuttle launch date***
2. *The climate pattern cleared* ***! the table / ! the counter / ! the room / the sky / the weather / the shuttle launch date***
3. ***! the table / ! the counter / the room / the sky / the weather / ! the shuttle launch date*** *cleared*

So we cannot generalize causative-inchoative alternations as having the same broad selection restriction on themes. This, along with the fact, which Rappaport-Hovav & Levin recognized, that it is possible for the inchoative form to have narrower selection restriction than the causative-inchoative form, such as in the following:

1. *The hostess cleared the dining room of patrons*
2. *The dining room cleared of patrons*

3. *The hostess cleared the table of dishes*
4. *! The table cleared of dishes*

Here, the selection restriction on *clear* in its inchoative form is narrower, as evidenced by example 4. So, as evidenced, each of the inchoative and the causative-inchoative forms of any alternation is capable of forming selection restriction independent of others. And this Rappaport-Hovav & Levin posited as the existence of two different lexical verbs of the surface form *clear*, one is specified as $[+c+m]$, while the other having no agentive requirement as $[+c]$, in Reinhart's (1996, 2000, 2002; Reinhart & Siloni, 2003) framework. This finding of lacking in uniformity in selection restriction of the forms will have significant repercussion on the graph-theoretic construction with respect to causative-inchoative alternations, in how to optimally compose different graph regions that correspond to these. The presence of selection restrictions also makes the automatic extraction of causative-inchoative frames substantially more complicated; it requires some knowledge-base hashed by frame's *predicate*, as well as its *theme/patient/subject matter/benefactive/recipient/etc* argument. It must either be hand built by linguists or learned from very large corpora.

6.1.3 Ditransitive lexical causatives

Lexically ditransitive causatives are in many ways analogous to causative-inchoative constructions, as it is simply a "causative-transitive" construction with the causitivization of an argument structure already replete with an *agent* and a *patient/theme*; in which then the original agent becomes the intermediary, or the *pivot* in the causative construction. .

6.1.3.1 Basic behavior of ditransitive causatives

This class basically includes any ditransitive verb that can be logically analyzed using the structure $Entity_x \xrightarrow{cause} (Entity_y \xrightarrow{perform} Entity_z)$ or $Event_1 \xrightarrow{cause} Event_2$. There are certainly other valid analysis of this class of surface forms, but the eventual graph construction step necessitates that the typology of the graph structure be kept simple. This means that in so far as possible, we need to analyze argument structure types with valence ≥ 3 into simpler logical representations with valence ≤ 2.

The members set of this class is a numerically smaller number of lexical verbs that express the causation of a set of events which themselves contain two obligatory $\theta-$roles, and is generally less researched than the causative-inchoative alternation. For the causative-inchoative alternation, there is generally a large number of inchoatives of the pairs that are missing (Parsons, 1985 / 1990), the situation is generally the opposite for lexically ditransitive causatives, such that many of the corresponding causative-transitives of the transitive forms of verbs are not in the language. (This generally fits in with the predominance of of dyadic frames in many languages) As with the causative-inchoative alternation, and even more so, the vast majority of the causative-transitive alternation are morphologically unrelated, or have undergone suppletion.

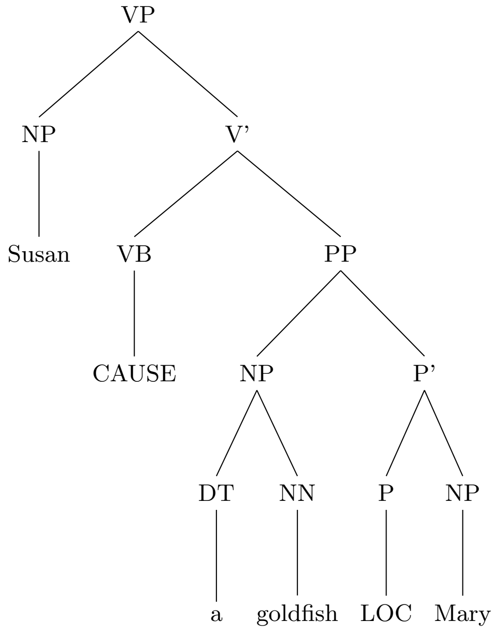
A prototypical lexical item of this type is the verb *show*, in the form: $Entity_x$ *shows* $Entity_y$ $Entity_z$, which contains the meaning of the event corresponding to $Entity_y$ *sees* $Entity_z$. This contained event posits an *experiencer* and a *theme* role, with the containing statement positing the meaning $CAUS(Entity_x,\ SEE(Entity_y,\ Entity_z))$. For certain languages, the causative pivot (direct effector of the caused event, $Entity_y$ above) exhibit different case markings depending on its semantic role in the causative chain. These are attested in language such as Quechua, Kannada, Japanese, Hebrew, and Hungarian, and the distinctions among the subclasses of case selection seem to be based on agentivity and experientiality of the pivot (Cole, 1983; Langacker, 1991), which are parameters in determining $\theta-$roles cross-linguistically. If these subclasses can be readily distinguished, then it would be useful to give them distinct treatments in the construction of the social network, since they each represent a different relational configuration among three entities.

6.1.3.2 Morphosyntactic-semantic interface

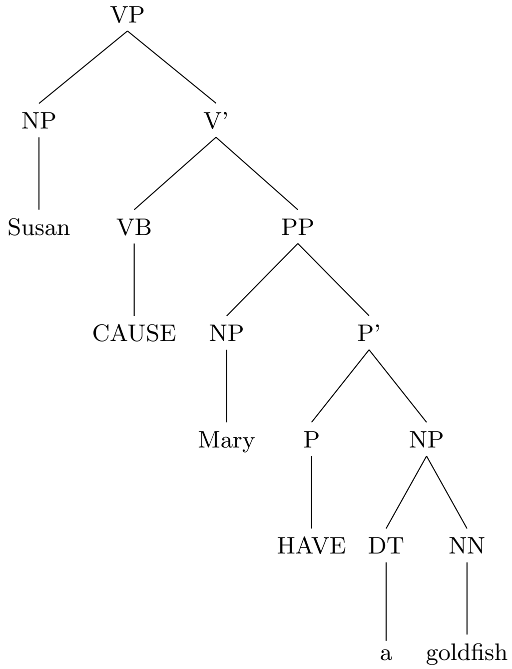
For some time, generative semantics has analyzed ditransitive constructions in English (Harley, 1995, 2004; Kratzer, 1996; Richards, 2001; McIntyre, 2005, etc) to possess a complex internal structure that is composed of multiple content VPs/PPs at a deep level. In this schema, many ditransitive constructions contain additional covert morphemes that occupy head positions within this multi-level structure, and entails many ditransitives to be analyzed causatively. The central question is whether the representation at the syntax-semantics interface of these ditransitive structures contains a single or multiple layers of structure in terms of a unit of representation (e.g. in terms of semantic frames). The variations of the morphosyntax with respect to the semantics of each frame potentially can be used to express pragmatic variations in emphases or implicatures; but it could also indicate individual, ideosynchratic variations based on the speaker's linguistic experience. These structures are necessary to explain the syntactic behavior of many idiomatic expressions in English (Richards, 2001). Consider the following:

1. *Susan gave a goldfish to Mary*
2. *Susan gave Mary a goldfish*
3. *Mary got a goldfish*
4. *Mary received a goldfish*
5. *Mary inherited a goldfish*

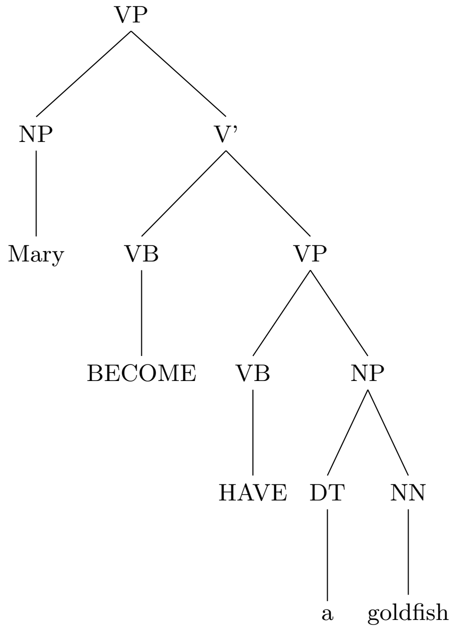
(*'giving'* often does not involve a single physical action, but involves some abstract transfer of possession) The examples (1) and (2) can be analyzed as below, (from Harley, 1995 and Richards, 2001). These analyses posit a covert morpheme *CAUSE* in the syntactic structure of these constructions, and would necessarily force a causative semantic analysis for many similar ditransitive verbs. This causative analysis of such polyadic frames applies in the same way as the analyses of causative-inchoatives, and allows individual polyadic frames to be further analyzed into simpler constituent dyadic frames, where the same *CAUSE* morpheme could be used to explain the alternation without the change of valency. So this essentially posits that each ditransitive verb behaves as an idiomatic expression of an underlying complex structure, as in the following that posits an embedded frame of a copula with locative information, within an overall causative frame:



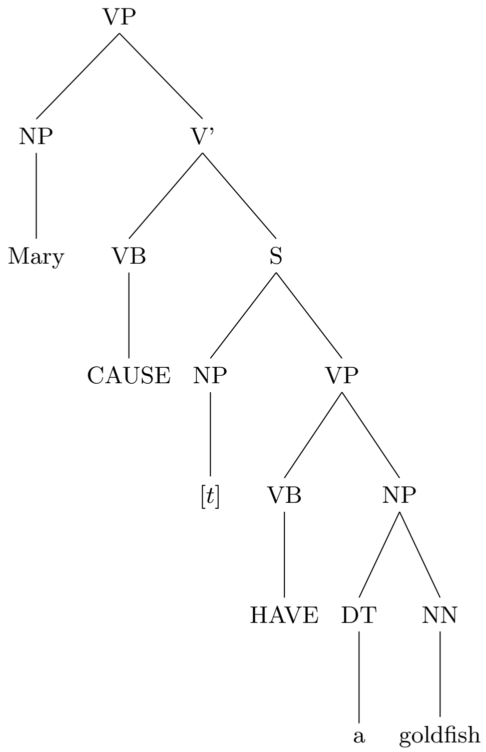
or potentially another interpretation where the embedded frame itself is dyadic, and conveys possession, within a matrix causative frame:



Also, the monotransitive example (3) corresponding to (2) has an analogous structure, which posits that the intermediate *actor* in the causative is identified with the *causer*. In this theoretical framework, ditransitive verbs, such as *give*, *send*, *show*, all have a morphologically complex internal structure that could be represented as separately in syntax, each of which can occupy a different head within a distinct syntactic constituent. (Richards, 2001) If this structure corresponds closely to LF, then the semantics of causative-transitives is analogous to that of causative-inchoatives. The underlying structural complexity is identical to that of (2), but utilizes the inchoative-specific compound verbal constituent of $[BECOME \quad (HAVE(\cdot))]$ in the place of the previous $[CAUSE \quad (HAVE(\cdot))]$ for (2). There would need be some special sub-categorization and transformation rules that are associated with this class of lexical verbs, a construction with a different lexical verb *'receive'* in (4) or *'inherit'* in (5) that achieves roughly the same meaning does not seem to have this peculiarity. This can be represented as in the following:



An possible alternative interpretation of (3) contains the reflexive idea, of an argument representing a single entity occupying both the original *causer* and *causee* roles of an underlying ditransitive construction, which would be a self-loop in a graph representation of the network. In an economic reciprocity framework (Fehr & Gachter, 1998 / 2000), this would correspond to the structure of a *unary reciprocity* relation. Such interpretation can be represented lexico-syntactically as:



We see that even morpho-syntactically simple structures can obscure covert transformations into something semantically more complex.

6.1.3.3 Semantic complexity of causative-transitive alternations

At a logical level of representation, many within this class of causative constructions are closer to that of 'let' constructions. These are a class of causative construction that contain the 'counterfactual logic' form (Wolff, 2003 / 2007), where e_1 is the causing event, and e_2 is the caused event, and $\neg e_1 \longrightarrow \neg e_2$. For instance, observe the following ditransitive constructions with additional qualifiers:

1. *John showed Mary the landscape, but she did not see it*
2. *John sent Mary the letter, but she did not receive it*
3. *John fed Mary some soup, but she did not eat it*

Some of these may sound slightly odd to some listeners, but generally are acceptable. Such examples show that many ditransitive lexical causatives have the property of e_1 providing the necessary, but not sufficient pre-condition for e_2 to take place; where e_1, e_2 are events with associated semantic frames in the linguistic form. This is saying that $(Entity_y \xrightarrow{e_2} Entity_z) \vDash (Entity_x \xrightarrow{e_1} Entity_y)$, with the semantic entail $\vDash$ having meaning similar to *let* or *allow ... to*, in affording pre-conditions for e_2.

While some ditransitives such as *feed* have a simple correspondence to a transitive (in this case *eat*), other lexical verbs falling in this category exhibit much more complex semantics that includes a causative force in the construction along with other predications about the causal relation. This behavior is possible in ditransitive causative forms in part due to the lack of shared surface forms (here the different surface forms often come into being through suppletion as a diachronic process) for the transitive and the causative-transitive forms, so it is possible to observe suppletion of one base transitive verb by multiple causative transitive lexical verbs, each occupies a slightly different semantic space in terms of complex relations in the real-world. Consider the following examples from an alternation that is rich in meaning variation:

1. *Mary has a fiberglass surf-board*
2. *John gives Mary a fiberglass surf-board*
3. *John hands Mary a fiberglass surf-board*
4. *John buys Mary a fiberglass surf-board*
5. *John awards Mary a fiberglass surf-board*
6. *John allocates Mary a fiberglass surf-board*

Example (1) is the base transitive form, having Mary as the *possessor* or *actor* depending on the granularity of the $\theta-$role system employed, and fiberglass surf-board as the *theme*. (2) here gives a default causative-transitive meaning, approximately denoting $(John \xrightarrow{someaction} surfboard) \xrightarrow{\vDash} (Mary \xrightarrow{possess} surfboard)$. (3) implies that John physically handled the surf-board in some manner, to cause Mary to have it, denoting $(John \xrightarrow{manipulate} surfboard) \xrightarrow{\vDash} (Mary \xrightarrow{possess} surfboard)$. (4) entails that John makes some type of payment to a third party, so that Mary may possess the surf-board, denoting $(John \xrightarrow{pay} Entity_z) \xrightarrow{\vDash} (Mary \xrightarrow{possess} surfboard)$.

The meaning of (5) implies a more complex set of events, which with the normal meaning of *award*, involves the fact that Mary somehow deserves the reward in the form of a surfboard, by gaining some accomplishment, hence

$$(Mary \xrightarrow{perform} Entity_w) \xrightarrow{\models} \Big((John \xrightarrow{manipulate} surfboard)$$

$$\xrightarrow{\models} (Mary \xrightarrow{possess} surfboard)\Big)$$

. The meaning of the last example has a different form of complexity, in that the choice of the predicate *allocate* implies that John assigns a set of objects to a group of individuals, of which Mary is a member, so $\Big((Mary \in E) \bigwedge \forall e_i \in E[John \xrightarrow{assign} e_i]\Big) \xrightarrow{\models} (Mary \xrightarrow{possess} surfboard)$.

As we see, these causative-transitives should normally be conceptualized as constructions of at least three predications, (except possibly with the default causative meaning, i.e. *give*), to obtain the complete meaning of each ditransitive verb. The baseline structure is $e_1 \xrightarrow{\models/E_0} e_2$, while frequently each of e_1 and e_2 are simple events with a single predication each, with e_2's predication remaining constant, and the predication in e_1 having semantic variation to give the specific ditransitive meaning. Certain distransitve verbs add on additional structure of semantic complexity to the $e_1 \xrightarrow{E_0} e_2$, such as example (5), where an additional layer of causality brings the structure placing E_0 as the result of a matrix causal relation, to $E_4 \xrightarrow{E_0'} (e_1 \xrightarrow{E_0} e_2)$, and such as (6), where the precondition is a set of parallel events predicated on a group of entities, relations in causality, sometimes forming multiple relations among eventualities by a single surface predicate.

6.1.3.4 Instrumental strategy

There is an alternative to causitivizing transitive constructions without raising valence at a syntactic level for some languages, as mentioned earlier in French (Langacker, 1991). There is a similar construction in English as well that accomplishes the causitivization with similar means, but generally involves replacement of the transitive of the form V_{tr} in $Entity_x \xrightarrow{V_i^{tr}} Entity_y$ with a causative-inchoative verb V_j^{caus}. The causative-inchoative V_j

would have a more generalized meaning than V_i, and is usually a hypernym of V_i. And the intermediate agent $Entity_x$, the causative pivot, would become an instrumental argument, or be in an adjunct that conveyed instrumentality, such that the final construction resembles $Entity_z \xrightarrow{V_j^{cause} \wedge INSTR(Entity_x)} Entity_y$, such as example (2) below:

1. *Mary kicks the ball into the corner*
2. *John moves the ball into the corner* ***with Mary / using Mary / by Mary's help***

This accomplishes the causitivization of the original clause, but the transformation process causes the semantic content of the original verb to be lost, such as losing the manner of moving the ball in *kicking*, when forming the above example. This loss of specificity of the verb meaning make the identification with the original event surface form much less confident. Instrumental strategy forces the intermediate argument to $[-m]$ cluster, to be a non-volitional entity, thus placing a further restriction on the argument structure.

6.1.4 Analytic constructions

A number of different causative constructions in English exist that are capable of describing the same caused event, these are also often termed *periphrastic causative*s by many semanticists (although this term is ambiguous in this context, so we will avoid this usage). These are called analytic causative, characterized with distinctive lexico-syntactic structure, range from lexically ditransitive causatives, to what we can analytic causative constructions in this subsection, to clearly multi-clausal constructions in the next subsection. Some speculate that this complexity-range of different constructions have the same underlying structure near the syntacto-semantic interface (e.g. Lakoff, 1965; Dowty, 1979), consider the following set of causatives that denote some causal relation between some action of John, and the cooling of the soup, with subtle semantic distinctions among them. These are usually organized on a continuum of the level of integration between the expressions of e_1 and e_2 in the linear structure of the sentence, as seen in the range of expressions capable of expressing the same causal relation of $John \xrightarrow{caus} (Soup \xrightarrow{cool})$ below:

a *Jane cooled the soup*

b *Jane made the soup cool*

c *Jane had the soup cooled*

d *Jane got the soup cooled*

e *Jane caused the soup to cool*

f *Jane caused the soup to become cool*

g *Jane brought it about that the soup was cool*

h *Jane caused it to come about that the soup was cool*

The analytic constructions are those that express causation through quasi-modals such as *make*, *have*, or *get*, which do not unambiguously command an embedded IP/TP, which conveys causality through a special Small Clause construction, as in examples (b) through (d) above. However, as we will see in a later section, there are subtle semantic distinctions among different parts of this range that will be important to their extraction from text and their implication in a social network structure.

6.1.4.1 Morphosyntactic issues with analytic causatives

The middle part of this range of constructions each with varying degree of surface form complexity constitutes what we call analytic construction (Maldonado, 2007; Kemmer & Verhagen, 1997), which utilizes a separate auxiliary / (non-productive class of) modalized content verbs, along with the content verb that indicates the caused event, to form the surface construction. These do not need to occur with inherently causative verbs, but sometimes occur with inchoative forms of some causative-inchoative alternation. The following are some examples:

1. *John made the ice on the counter melt*
2. *John had the ice on the counter melted*
3. *John had Mary melt the ice on the counter*
4. *John made Mary wash the car on the driveway*
5. *John had the car on the driveway washed*

6. *John had Mary wash the car on the driveway*
7. *John let Mary wash the car on the driveway*

In particular, ‘make’-causatives are the most frequent occurring causative construction type (Stefanowitsch, 2001; Lauer, 2010), at least in spoken English, such as in the Switchboard Corpus. Thus any reasonable sample of linguistic causatives should include a robust method for extracting analytic causatives. There is some evidence that the corresponding D-structure contains more than one fully-formed clause, much of it lies in the fact that ‘make‘ and ‘have‘ used in these constructions are incompatible with normal notions modals or other auxiliaries.

1. *John should wash the car on the driveway*
2. *John did wash the car on the driveway*
3. *! John made wash the car on the driveway*
4. *! John had wash the car on the driveway*

(Here, there exists a forced external argument for the embedded clauses.) It seems that the content verb, in this case ‘wash‘, needs to select for an external argument that is distinct from the external argument of the overall matrix clause ‘John‘, such as forms like *John made Mary wash the car* and *John had Mary wash the car* are grammatical. This shows that the content verb resides in a clause distinct from the matrix clause, and that ‘make‘ and ‘have‘ here do not behave as normal modals and auxiliaries. In this case the *intermediate agent/causee* would need to occupy both an internal argument of the matrix clause, and the external argument position of the embedded clause at some time during transformation; this is sometimes classified as *perception-cognition-utterance (PCU) verbs* in cognitive linguistics (Givón, 2008). Although, these also behave unlike other voice or modality expressing verbs in the English lexicon, which generally require a *small clause* (SC) as a complement:

1. *! John made Mary [*$_{SC}$ *to sing at the recital]*
2. *John made Mary sing at the recital*
3. *John asked Mary [*$_{SC}$ *to sing at the recital]*

4. *! John asked Mary sing at the recital*
5. *John wants Mary* $[_{SC}$ *to sing at the recital]*
6. *! John wants Mary sing at the recital*
7. *John needs Mary* $[_{SC}$ *to sing at the recital]*
8. *! John needs Mary sing at the recital*

Another significant issue in separating an analytic causative into two clauses is the would-be matrix clause's inability to undergo syntactic transformations such as passivization without the over presence of the SC, such as the attempt at passivizing (1) below:

1. *John made Mary sing at the recital*
2. **! Mary was made sing at the recital*
3. *Mary was made* $[_{SC}$ *to sing at the recital]*

where it is shown a strong possibility that the embedded clause has a SPEC position that must be overtly occupied, such that there must be some indication of its presence at the S-structure. The passivization requires not just a simple movement of a constituent of an analytic causative, but also augmentation of the embedded clause to become a well formed SC. Another analytic causative 'get'-construction, normally normally takes a passivized complement, and is in alternation with the unaccusative 'get'-construction (example (1) below). (Haegeman, 1985; Fleischer, 2006 / 2008) This construction, analogous to 'make'-causative's inability to undergo normal passivization, cannot undergo de-passivization without the presence of the SC structure at surface, illustrated in the following examples (adapted from Fleischer 2008):

1. *John and Mary got arrested*
2. *John got Mary arrested by the police*
3. **! John got the police arrest Mary*
4. *John got the police* $[_{SC}$ *to arrest Mary]*

6.1.4.2 Semantic space of analytic causatives

The earlier examples (a) - (h)'s types range from those of the tightest morphosyntactic integration of the causal and caused events (causative-inchoative, being contained in one word), to those of the loosest, and yet unambiguously causal relation (multiple levels of embedding between the events). The readings of these demonstrates the generally believed trend that tighter morphosyntactic integration of two events leads to stronger semantic relation between the two events; the precise mechanism and semantic differences among these forms need further elaboration. The underlying principle likely lies in tendencies within human cognition, and has not been precisely determined with respect to correlation between structural complexity in morphosyntax and semantic distance between two events. In certain languages with morphologically expressed causatives, such as Navajo (Gessner, 2001), the meanings in English of the 'make'-causative constructions as well as lexical causative constructions (such as causative-inchoative) correspond better (than other forms of English causatives) to Navajo morphological causatives in their semantic space. The 'make' and 'have' causative constructions are frequently classified as members of *bare-stem complement verbs* (Givon, 1993) This class exhibits a tighter integration of the two events that comprise the causative. (Givon, 1993; Stefanowitsch, 2001; Hollmann, 2003). Some analytic constructions also exhibit additional semantic content to pure causation, such as *'make'-causative* often (but no universally) being analyzed as *directive* or could be by default interpreted as*coercive* causation (Lauer, 2010). (Other interpretations of *'make'*-construction are possible, but would be semantically *marked.*) Consider the following three scenarios involving John, Mary, and Patrick:

i *John is the commanding officer of the platoon, of which Mary is a member; John gives a command for Mary to fire on Patrick, an enemy combatant*

ii *John is being hunted by Mary on the special forces; but he tricks Mary into shooting Patrick instead, by exchanging his distinctive uniform with Patrick*

iii *John is training Mary as a new recruit, using Patrick as a mock target of the exercise; but he inadvertently mistook a live round for a blank, and handed Mary the gun with the live round, and Mary then shot Patrick*

All three scenarios are instances of causation, with John being the *cause* of the shooting event, but all three instances differ with respect to the relations among the three entities in subtle ways. For scenario (i), it is clearly appropriate to use *John made Mary shoot Patrick*, as John had both the authority and volition to cause Mary to undertake the shooting. For scenario (ii), most would also judge it to be semantically compatible with *John made Mary shoot Patrick*, but most would also feel some awkwardness comes from an infelicitous statement that by itself seems to indicate John had a more direct relationship with the shooting; this is likely due to the lack of direct control of John over the actions of Mary, except by deception. For scenario (iii), it is unlikely to be judged in any interpretation to be appropriate for the situation, this is due to the lack of both causer control and authority over the intermediate *agent*, but also the absence of intention or volition on his part. In many ways, this class of constructions is closer to the meaning expressed in *John commanded Mary to shoot Patrick* (Shibatani, 1976), than any other construction, although without explicitly stating the authority of the causer. So, some level of volition or control of the *agent/causee* is presupposed but such a construction, unless otherwise stated in the surface form.

Lauer (2010) found evidence that casts doubt on ‘make’ being a hyponym of ‘cause’. He found examples where ‘make’ is exactly appropriate for the intended meaning of the speaker, while an analogous construction with ‘cause’ sounds marginal, such as below:

1. *I made myself work out three times a week*
2. *?? I caused myself to work out three times a week*

If the ‘make’ construction conveys the idea of volition or control that the ‘cause’ construction does not, in addition to the meaning contained in the plain ‘cause’ construction, then it is likely a violation of Gricean maxim of quantity to use the latter when the former is more appropriate. So this does not constitute clear evidence that the ‘make’ construction does not entail the ‘cause’ construction, but it does show a clear semantic distinction between the two, conveyed through the difference in morphosyntax; we will see more implications in later sections.

6.1.4.3 Permission subclass analytic constructions

The lexical item *let* in English has a highly variable semantic space, and the constructions expressed using 'let' in English to have highly variable corresponding structure across other languages. It is suggested (Bally, 1920; Wierzbicka, 1988) that this is caused by variations in the basic concept of *causation* among the languages; and this subclass of analytic causative constructions is a good example of variations at the margins of this core concept. The 'let' causative is to be treated slightly differently in its real-world semantics, expressing permission (Wierzbicka, 2002), by removing obstacles to the completion of the caused event. The more generalized concept here is $P_u(e_1) \leq P_t(e_1) \Longrightarrow P_u(e_2) \leq P_t(e_2)$, where t, u are indices along a linear sequence (e.g. time) and $u > t$. It is closer to counterfactual logical formula (Comrie, 1989; Shibatani, 1973 / 2002). If the causing event is e_1 and the caused event is e_2, then the 'let' causative indicates that $[\neg e_1 \longrightarrow \neg e_2] \bigwedge \neg[e_1 \longrightarrow e_2]$, with e_1 being a necessary, but not normally sufficient condition for this construction. Consider the following examples:

1. *!? John put the dishes on the table away, so John let the table become cleared*
2. *!? John forced Mary to walk into the room, so John let Mary walk into the room*

Such constructions are clearly anomalous with respect to the semantics. Whenever the causer is the main force in bringing the caused event about, then the 'let' construction becomes problematic. So this subclass of causative construction merits particular treatment in probability theory and formal representation, when converting into its logical representation, and eventually into social network structure.

6.1.5 Full embedding multi-clausal strategy

Multi-clausal strategy as defined for our purposes is the set of causative constructions that involves more than one CP at the surface form, excluding the discourse level causations (which offer very different characteristics for extraction and analysis). We place the strategies that unambiguously involve multiple clauses in the D-structure of the causative construction in this category. By unambiguous embedded clause, we mean that it must at a minimum contain the evidence of a *small clause* in its surface form, as the following constructions contain:

1. *John caused the ice in the driveway* $[_{SC}$ *to melt]*
2. *John got Mary* $[_{SC}$ *to rent the car for the trip]*
3. *John forced Mary* $[_{SC}$ *to rob the corner store]*

This comes with the caveat that multi-clausal constructions other than those 'causes' or 'bring - about' could have additional denotations on the causing event (e_1) or on one of its arguments other than the minimal semantics of a causative. Counterfactual causals are a special class which are only semantically causal if one interpret it in a *possible-world* scenario (more detail in 7.2.2). For instance in example (3), the predicate *force* assigns the property of $\{-volition\}$ on the *causee* of e_1 Mary, the intermediary in the causal construction, denoting that the caused action with Mary as the *agent* is not voluntary in nature. A number of lexical verbs have been identified by the psycholinguistics community as being able serve to indicate causation within a multi-clausal causative construction (Wolff et al., 2002; Wolff & Song, 2003), these include the following:

cause, bribe, compel, convince, drive, have,
impel, incite, induce, influence, inspire, lead,
move, persuade, prompt, push, force, get, make,
rouse, send, set, spur, start, stimulate

In our classification, we have separated out those that do not take a full SC as a complement into analytic causatives, which include 'make', 'have', 'let', and some constructions involving 'get', because they exhibit different semantic behavior. This is in some ways an extension of the analytic type, with the distinction being that the structure can be arbitrarily deeply embedded, and the there is a large variety of different lexico-semantic cues for its construction. So we will term the above set of lexical verbs, less the set of modality-expressing verbs used in analytic causatives 'make', 'have', 'let', the multi-clausal causative indicator candidates, or *manner-of-causation.* When one finds one of these lexical verbs as the content verb in the matrix clause, with an embbeded small clause, then the entire construction has high likelihood of being in this class of causatives.

The multi-clausal construction with ‘allow’, ’permit’, or ‘enable’ heading the matrix clause follow a similar logical representation as the ‘let’ analytic constructions, which have an underlying counterfactual logic. These should be included within the analysis of causative, whenever we embrace a more expansive definition of causation in language, as these share many structural similarities to causatives at surface level and in their deep logical representation (Wierzbicka, 1998 / 2002).

6.1.6 Discourse level causation

Discourse level causative constructions do not have a monolithic definition, as by its conception, they occur as pairs events represented within a discourse with some distance in between. The minimal distance between the two representations of events is a pair of adjacent clauses, and the pair does not occur in a matrix-embedded clause structure. The clearest examples of discourse level causation are cued by explicit *discourse connectives* (Do et. al., 2011), which are extra-CP conjunctions such as *because*, *therefore*, *so*, *thus.* The following are some examples of this category:

1. *John gave Mary a ring, therefore Mary called her parents*
2. *Since John gave Mary a ring, Mary called her parents*
3. *Mary called her parents, because John gave her a ring*

This type of connectives marks one of the pair of linked eventualities either as the *cause* or the *effect.* The surface indication of the pair is usually adjacency, or being two clauses in the same sentence. This type of causality normally has a three event structure in the semantic representation, as: $(Entity_x \xrightarrow{perform} Entity_y) \xrightarrow{cause} (Entity_u \xrightarrow{perform} Entity_v)$. While sometimes $Entity_y$ identified with $Entity_u$, there is no intrinsic structural reason that it must be the case, such as in *“because the President ordered the strike, the Russians launched their missiles in retaliation”*.

A number of recent works have focused on semi-supervised or unsupervised methods in extracting causal pairs from linguistic corpora (Beamer & Girju, 2009; Riaz & Girju, 2010; Do et. al., 2011), also utilizing discourse connectives, in order to find latent links across wide context. These are focused on data type of traditional linguistic corpora, and face similar issues as domain-specific NER such as for NER tool-kit from Stanford NLP group (Manning et al. 2014), with regard to the data type in social media. Moreover, most of the methods of extracting discourse causative pairs focus on causal relation between types of events as represented by their predicates, without the focus on the entities participating in the causal relation, this makes these methods alone unsuited for contributing to the topology of a specific real world social network. There is no broader context beyond short text snippets in typical social media data, and no overall linearized ordering of text that serves as distance measure; so given the lack of data of the appropriate type, there is not a significant number of causal relations across distance to identify. Also there is no resource that is appropriate for the data type homologous to the Penn Discourse Tree Bank (Prasad et. al., 2007).

6.1.7 Relevant structures that are not explicitly causal

Many non-embedded but subordinate multi-clausal structures have some conditional or counterfactual relation between the events that are represented in each clause. Some of these may occur such that one of the pair is logically subordinate to the other using the explicit semantics of the connectives, which for instance could have *“whenever* $[_{Clause1}\ Entity_x\ ...]$ *,* $[_{Clause2}\ Entity_y\ ...\ ]$*”* or *“*$[_{Clause1}$ *should* $Entity_x\ ...\ ]$ *,* $[_{Clause2}\ Entity_y$ *would ...* $]$*”*, the first represents a conditional relation of $P(e_2|e_1)$, while the second is counter-factual of $P(e_{2_{e_1}})$. while others pairs may not have logically explicit connective, and merely occur with general connectives such as ‘and’.

These by themselves cannot be generally regarded as causative, since they generally only indicate a relation that may have a statistical association between the occurrences of e_1 and e_2. Association between two variables in a system potentially encode many subtypes relations in a Markov model, of which causality is one. Consider the following examples with a very common type of such structure in the form of *if ... , then ....*:

1. *if John is not reachable this afternoon, then Mary has called for a meeting during that time*
2. *if the caribou migrated across the brooks range during the fall, then other native spieces likely have done the same*
3. *if we observe large radio lobes approximately along the magnetic poles of an AGN (active galactic nucleus), then there should be a strong EMR source visible in the x-ray spectrum at the center*

Given some real world knowledge, we know that (1) likely indicates that Mary likely called for the meeting prior to the change of John's schedule; so if the relation is causal, it is likely $e_2 \xrightarrow{caus} e_1$. For example (2), it is likely that some third event, such as change in weather pattern during the fall, precipitated both e_1 and e_2, forcing both groups to migrate. For example (3), we know that the occurrence of both the radio lobes and the x-ray source are the result of some material in the galaxy that falls into the central rotating black hole of the AGN, forming an accretion disk. So for both (2) and (3), both e_1 and e_2 are the result of some unmentioned e_3.

In such a case where specialist knowledge is required for this discrimination task, it is not certain whether any individual's perception on the relation would be causal, but the underlying real-world relation is causally structured; thus, these examples may demonstrate some conflicts between the cognitive and world-logical definitions of causation. The above are explicit conditional structures that show some association between e_1, e_2, but there is no expressed or assumed sequential connection among the events; this is typical of explicit conditionals, which lack the temporal element of causality; and cannot be generally regarded as causal without additional information immediately surrounding the frames or in the larger context. . Thus, we cannot regard this class of multi-clausal constructions as generally causal, even when there is strong dependence in conditional probability between e_1 and e_2. But measuring this may be useful in arriving at a working definition of causality in SN, especially in terms of the counter-factual pairs of clauses that are present in the corpus; this we will examine in a later section.

6.2 Common issues in linguistic causality

There are a number of issues in associating causative construction with real-world events and entities. The discussions of several of these will elucidate important findings with regard to some system of taxonomy for causality in terms of their extensions in real-world events and relations. It also has a role in how types of linguistic causative construction would find correspondence in a real social network. Both the perceptions of speakers (individually determined) and logical structures (determined by the system of logic subscribed to) play some role in the formulation of the multi-variate taxonomy/ontology. We will discuss several of these below.

6.2.1 External argument selection

A clausal causative construction requires the use of an external argument, the presence of which can be used as a test for whether a lexical entry participates in the causative-inchoative alternation. (Rappaport-Hovav, 1988; Zubizarreta, 1992, Cortez, 1995). The role of *cause* at the morphosyntax-semantics boundary may represent one of several semantic roles at a logical level of representation. Consider the situation where *John compelled Mary to throw the brick through the window pane.* Consider the following partial representations of the event above:

1. *John made the window pane break*
2. *Mary made the window pane break*
3. *the brick made the window break*

All of these construction in the analytic causative form are well formed, although the meanings on the level of logical representation with respect to the real-world event references are quite different from each other. The three different real-world roles in this complex event might be analyzed as being the $\vec{cause}^i$, $\vec{cause}^d$ with respect to directness of the construction (or possibly $\vec{agent}$, depending on the view whether causative-inchoatives are true agents, or alternatively $\vec{effector}$ role assigned to the external argument (Williams, 1981 / 1994), in some literatures), and $\vec{instrument}$, By $\vec{cause}^i$ here, we mean a mediated actor (of some distinct entity or concept at the same granularity) in bringing about the change of state, and by $\vec{cause}^{d}$, we mean an unmediated actor in bringing that about. We differentiate from the previous *cause* role at the shallow morphosyntax-semantic interface, by an extra $\vec{\cdot}$ symbol. At the same time, the causative-inchoative constructions does not seem to be semantically compatible with the original event for the first of these three representations, with *John* being only an indirect cause of the state transition that *the window pane* underwent, as in the following:

1. *! John broke the window pane*
2. *Mary broke the window pane*
3. *the brick broke the window pane*

So the above shows that some causative constructions have selection restrictions that eliminate any indirect $\vec{cause}^{i}$, but allows for the selection of and $\vec{cause}^{d}$ that is directly involved in the action of changing the state. There are some indications as to what could serve the role of $\vec{cause}^{i}$ as well, potentially with regard to *volition*, in the following pair:

1. *John brought about the circumstances such that the window broke*
2. !? *the brick brought about the circumstances such that the window broke*

As many have previously observed (McKoon & MacFarland, 2000; Levin & Rappaport-Hovav, 1995; Wright, 2001/2002), there is also a distinct class of causative-inchoatives that selects only arguments with $\vec{instrument}$ at a deep semantic level. For example, in the event such that *John compelled Mary to submerge the circuit board in hydrochloric acid during the etching process*:

1. *! John corroded the exposed copper on the circuit board*
2. *! Mary corroded the exposed copper on the circuit board*
3. *the hydrochloric acid corroded the exposed copper on the circuit board*

We observe the selection restriction of the verb *corrode* precluded both the $\vec{cause}$ and $\vec{agent}$ from appearing in the *cause* position of the causative-inchoative, while only an $\vec{instrument}$ in the deep sense can be selected.

So we see that there is at a minimum a three-way distinction in terms of the breadth of the selection restriction on the *cause* position in the construction. The analytic constructions almost always will present the broadest selection restriction, being able to select some *ca*$\vec{u}$*se* that are a number of steps away from the actual state altering event; whereas the lexical construction would have at most the same breadth in selection restriction, and often narrower depending on the specific class of causative-inchoative or lexically ditransitive causative verb.

6.2.2 Presence of a logical ca$\vec{u}$se

For most causations that we see in life, there is some identifiable *ca*$\vec{u}$*se* that observers can point to, once all of the relevant information about a causative construction becomes known. But there are certain semantic classes of inchoatives where no such *ca*$\vec{u}$*se* is necessarily present, even in the real world referents of the event, consider the following:

1. *the bomb detonated*
2. *the bomb was detonated*
3. *someone or something detonated the bomb*
4. *the child becomes blind*
5. *the child was blinded*
6. *the wrong medical treatment blinded the child*

The meanings of the first sentence above is not necessarily be identical to the other two in their real-world forms. The *bomb* in this case could have been caused by some external entity (with or without volition) to detonate, but it also may be possible that no such entity exists, and that the detonation is an event brought on by some event internal to the *bomb* itself, such as the trigger ignition spontaneously lighting, with no observable cause in the real-world. Where the other examples contained either a cover (2) or an overt (3) *ca*$\vec{u}$*se* for the event of detonation. So we have to recognize that there are possibilities of purely random occurrence in these classes of causative-inchoative pairs that does not depend on external cause. Such processes often manifest themselves, like above, as sub-categorizations of some lexical verbs in syntax. A similar sequence of reasonings exists for (4) - (6).

Some lexical verbs involved in causative-inchoative alternations are much more likely than others to be involved in such “cause-less” events represented in the inchoative. There are some, in fact, by default in the absence of additional context, presumes that there is no external cause to the event that followed:

1. *the building crumbled*
2. *the main beam fractured*
3. *the pitches from the violin strings harmonized*
4. *the sleeping girl awakened*
5. *the building crumbled from the demolish charge*
6. *the main beam fractured from the falling asteroid*
7. *the pitches from the violin strings harmonized by the tuning of a musician*
8. *the sleeping girl awakened by the raindrop on her cheeks*

Here, examples (1-4) seem to have different default meaning than (5-8), which have explicit causes built into the surface form. The *buildings* and *beam* could fall apart from disrepair, the *violin strings* could harmonize if they so happen to be in tune at that time, and the *girl* awakes from a natural sleep. It seems that all of the examples where an uncaused inchoative may exist, is in the context of a situation where the natural / default outcome expected from the current state of events, which will take place over time, with the absence of an external force that can alter the course of events.

So we can see that even though that the pairs in causative-inchoative alternations should have the same lexical meaning, it is dangerous to assume that causation is mandated by these lexical verbs that **are able to** serve as inchoative predicates. For certain lexical verbs that can appear in causative-inchoative alternations, there could also be some cause-less description, such as *‘the bomb spontaneously detonated’*, or *‘the building by chance collapsed’* that are legitimate as well. The *causative-inchoative* forms of these alternations certainly do always contain some causation as a part of the meaning, as the $ca\vec{u}se$ is represented in the surface form; but the *inchoative* forms of these alternation should only be reliably deemed as caused event, if some adjunct (usually in the form of a PP) to denote something in the chain of causation. The other classes of causatives do not present this problem, since all of their structures require some *causer* argument to be a part of the semantic frame, even if the $ca\vec{u}se$ is unknown, such as: *"**Someone** made the girl wake up"*.

6.2.3 Chain of causation

In the real-world, events of causation often operate in a transitive fashion, with the resultant state of the system providing the necessary and sufficient condition for some further causal event. This leads to the concept of "chain of causation" that lies at the heart of how causation orchestrates the interaction among the set of events in the natural order, including the aforementioned *prime mover* in the Aristotelian world view. So language's treatment of such conception is also an essential component of how causation is expressed.

6.2.3.1 Variations among classes of causatives

There is a common observation that lexically constructed causative generally only describe direction causation while the embedded multi-clausal strategy permit the description of distant causation chains, which emphasize the sequential nature of series of interrelated events in their real-world extensions, where direction causation is in the same sense as $\vec{cause}^d$ mentioned earlier. This has been widely recognized in the field of semantics for a considerable period of time (Brennenstuhl & Wachowicz, 1976; Comrie, 1985; Croft, 1991; Dowty, 1979; Levin & Rappaport Hovav, 1994; McCawley, 1978, Parsons, 1990). This difference in denoting the chain of causation is not only expounded in theoretical semantics, but has also be experimentally verified in psychology experiments with human participants (Wolff, 2003), and thus is robust with respect to real-world events. Consider a complex chain of causation in the statement: *Adam Smith brought about the recognition of the force of the 'invisible hand', which brought about the free market system, which in turn provided the precondition for the industrial revolution, which then created the need for abundant source of fuel, which led to the felling of most trees of the forests of Western Europe.* With this in mind, examine the following causative construction for compatibility with the above statement:

1. *! Adam Smith **felled** most trees of the forests of Western Europe*
2. *? Adam Smith **made** most trees of the forests of Western Europe **fall***
3. *Adam Smith **caused** most trees of the forests of Western Europe **to fall***
4. *Adam Smith **brought it about** that most trees of the forests of Western Europe would **fall***
5. *Adam Smith **acted** in such a way as to ultimately result in the eventuality that most trees in Western Europe would fall*

Statement (1) here is clearly incompatible with the original statement (provided Adam Smith didn't fell trees in an unrelated way), and shows that a causative-inchoative has very limited selection restriction with regard to the length of the chain of causation. Statement (2) is also judged by some as incompatible with the original. And while, as we showed earlier, analytic causatives contains the ability to represent a chain of causation containing more than one event, its length is not unlimited. One of the issues here that causes difficulty for some speakers seems to be intentionality, volition, and causer authority, which for many is implied with the usage of these analytic constructions, which is not the case with the multi-clausal strategies. The temporal integrity of the entire construction mentioned in Section 6.1.4 likely also plays a role in making the this example very awkward in conveying the entire sequence of causation. Statement (3) - (5) are judged by most as compatible with the original, and shows that with a multi-clausal strategy, a chain of causation of significant length can be represented. (3/4) still present some awkwardness of interpretation for some speakers, but (5) should be completely acceptable for almost everyone. This also applies to other forms of analytic causative constructions. Consider the following scenario: *John needed to create a diversion for his get-away, so he sent his partner in crime, Mary to fire a gun shot into the crowd in the dining room, and the guests all rushed out of the hotel.*

1. *! John **cleared** the hotel dining room, in order to create a diversion*
2. *John **had** the hotel hotel room **cleared**, in order to create a diversion*
3. *John **caused** the hotel room **to be cleared**, in order to create a diversion*

4. *John* ***brought it about*** *that the hotel room is* ***cleared****, in order to create a diversion*

The lexical construction does not seem to be compatible with the aforementioned scenario, while the analytic and multi-clausal constructions seem to be able to express the appropriate sequence of events.

The explanation from the psychology community has been to posit a phenomenon of *metonymic clipping* (van Valin & Wilkins, 1996; Wolff, 2003) for the *causer* of the linguistic constructions that are capable of denoting long chains of causation among real-world events. In effect, a sentient original *causer* may stand in for all of the subsequent intermediaries in the chain, up to the ultimate effector of the final simple event. This is similar to the psychological effect of *windowing of attention* (Talmy, 1996), where a volitional original *causer* can treat all the intermediate means in accomplishing an ultimate goal as a single invariant. This has significant implications for which types of causative constructions we can reasonably rely on for contribution to a social network structure. As we have seen, multi-clausal strategy is virtually unbounded in terms of the separation in time, space, and the length of logical sequence the causation can follow in the real-world. Many instances of such multi-clausal causations cannot be relied on to contribute to a reasonable topology of a social network structure that represent real individuals and events.

6.2.3.2 No-intervening-cause criterion

In some expressions of seemingly transitive chain of causal events in natural language, when an explicit intermediary entity exists between the original cause and the ultimate resultant event, the underlying real-world sequence of events may be better described as direct causation. This type of direct causation with intermediary entity occurs because of the existence of *no-intervening-cause criterion* (Wolff, 2003 / 2007) in establishing direction causation. For a set of examples were this criterion is determinative, consider the following:

1. *Daniel used the key to close the lock (by turning the key in the lock)*
2. *Daniel used Emily to close the lock (by commanding her to lock it)*

3. *Daniel caused the falling rock to close the lock (by accidentally pushing it off the ledge)*
4. *?? the key caused the lock to close*
5. *Emily caused the lock to close*
6. *the rock caused the lock to close*

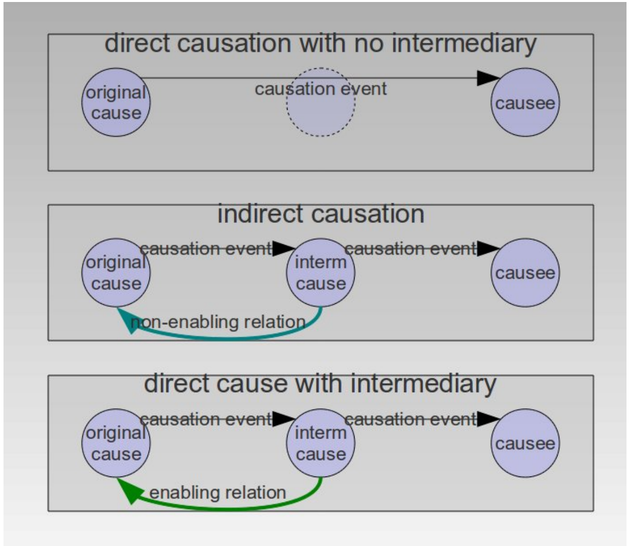
We can see the difference between examples (1) and (2), which Talmy (1988) and Wolff (2003) ascribed to the no-intervening-cause criterion. This is evidenced in examples (4) and (5), where the *key* is not normally assigned the role of the *causer*, whereas a human entity such as *Emily* could readily take the intermediate causer role in the chain of causation: $Daniel \xrightarrow{cause} (Emily \xrightarrow{cause} (Lock \xrightarrow{BECOME(open)}))$. And this is not simply an issue of volition or animacy on the part of the intermediary, as demonstrated in (3) and (6), where a *rock* can equally well fit this criterion to serve as the intermediate cause.

The underlying explanation of whether an entity can serve as intermediate cause is explained with the notion of *enabling*. If the intermediate entity has a semantic enabling relation with the original causer, then the criterion is fulfilled, and if the intermediary has some other semantic relation with the original causer, then the entire construction must be indirect, as described in Figure 6.1:

In our examples, it would be perfectly legitimate to describe the situation as *"the key enables Daniel to open the lock"*. But the corresponding descriptions *"the rock enable Daniel to open the lock"* and *"Emily enables Daniel to open the lock"* are highly suspect at best. It may seem like a some loop in that the the *rock* or *Emily* enables Daniel to open the lock, and at the same time *Daniel* causes either the *rock* or *Emily* to behave in a certain way so that the lock is eventually undone. But the key distinction is that the former, e.g. the enabling relation involves ***the presence*** *of the rock or Emily* that enables *Daniel* to act; but the latter involves *Daniel* bringing about some action in the *rock* or in *Emily* herself, so that the eventual effect takes place through such an action. It is merely the ambiguity in the linguistic representation that effects this illusion of the violation of logical consequence.

Also notice that there could be an *enabling* relation predicate from the intermediary to the original only if the original causer is sentient and is capable of exercising volition. This effect can be seen in the following:

Figure 6.1: no-intermediate-cause criterion and enabling: here the 3rd is case is sometimes conceived as *"enabling"*, as in the presence of a precondition, but the the intermediary itself may be involved as a member of the causal chain, such that it is a link between the original causer and the eventual effect; this can be seen in an example such *"the credit card enables Daniel to open the lock"*, where the presence of the *"credit card"* is enabling, and is itself an instrument in bringing the effect about



1. *the freezing rain caused the key to close the lock (by expanding while freezing and forcing the key to turn)*
2. *the key caused the lock to close*

In this case, the description *"the key caused the lock to close"* seems to be a much more apt for the situation, given that an inanimate object as *freezing rain* cannot be imbued with volition.

The indication of the *enabling* condition of this criterion has been experimentally verified as having broader reach than just the usage of the predicate *'enable'* (Wolff 2003). A number of verbs capable of making such predications have been identified (Wolff et al. 2002; Wolff & Song, 2003) in experimental psychology, which are *aid, allow, enable, help, leave, let, permit.* When any of these is present between the intermediary and the original causer, the no-intermediate-causer criterion can be utilized to determine the causal chain as a single conceptual event by English speakers. Some adjunctive and embedded causatives also express enablement relations.

6.2.3.3 Causal chain and change of state

There is also a requirement on the sequential state of the system that the causal eventualities operate in. Consider the case of a world-system U such in which a sequence of states of the system $Q^U =< q_0, q_1, ..., q_n >$ succeeds one another, and also a sequence of eventualities $E^U =< e_0, e_1, ..., e_m >$ occur, such that $\forall q_i, q_{i+1} \in Q^U$ $[\exists\ \tau \wedge (\tau(q_i) = q_{i+1}) \wedge \exists e_j \in E^U[e_j \longrightarrow \tau]]$, where τ is some effect that stems from the eventualities e_j, thus $q_i \xrightarrow{e_j} q_{i+1}$. There are also eventualities $e_k \in E^U$ such that $q_i \xrightarrow{e_k} q_i$. So we will call the subset of eventualities that cause a transition to a different state in U as E^U_τ. Those that do not fall into two separate groups, one such that describes an integral eventuality where there are permitted changes in state during the eventuality's occurrence, but at the conclusion the entity returns to the original state prior to this occurrence; an occurrence such that (where e^x_j is a temporally coherent portion of the occurrence of the eventuality) $e_i \xrightarrow{e^1_j} e_h \cdots \xrightarrow{e^l_j} e_i$ is permitted; this class we term $E^U_{\neg\tau}$. The remainder of the predications all do not allow the involved entity to undergo any state transition during the occurrence, whether or not the end state identifies with the onset state, this class we will term E^U_ς $\forall e_j \in E^U_\varsigma$ $[q_i \rightarrow q_{i+1}] \Longrightarrow q_i = q_{i+1}$.

Certain classes of verbs distinctly encode the change of state in the system, indicating that one of the entities that are involved in the predication, has a different state immediately following the eventuality compared to immediately prior to the event. This set of eventualities $\in E^U_\tau$ include those denoted by *fall*, *rise*, *kill*, *scatter*, *build*, etc. Another set of eventualities distinctly returns the system to the exact same state as immediately before the eventuality's occurrence, these $\in E^U_{\neg\tau}$ include those denoted by *blink*, *breathe*, *flap*, *bounce*, etc, eventualities that are cyclical in nature, with respect to the entities' state. And predicates (may not be eventualities in certain conceptual frameworks) $\in E^U_\varsigma$ are stative in nature, such as *BEING(heavy)*, *BEING(bright)*, *BEING(cruel)*a *hail from*, *descend from*, etc. (There are also many verbs that are epistemically ambiguous, such as *turn.*)

Whenever in an expressed causal relation, the types $\tau, \neg\tau$ of the causing eventuality and the caused eventuality are the same, the causation is always logically sound, and the type of each eventuality remains constant. While a causation between two instances $\in E^U_\varsigma$ seems to depend on the relative degree of 'permanence' of these two predications. Consider the following examples:

1. *Daniel killing Ferdinand caused Emily to fall to the floor*
2. *Daniel blinking Ferdinand caused Emily to breathe heavily*
3. *?! Daniel being cruel to Ferdinand caused Emily to be descended from Ganymede*
4. *Daniel being descended from Ganymede caused Emily to be cruel*

For examples (1) and (2) where $e_i \xrightarrow{caus} e_j \mid e_i, e_j \in E^U_\xi, \xi \in \{\tau, \neg\tau\}$, these are semantically and logically well formed. The examples (3) and (4) show that $e_i \xrightarrow{caus} e_j \mid e_i, e_j \in E^U_\varsigma$ sometimes are well formed, while not at other times.

There are various interactions among these classes of eventualities, whenever two of different classes are involved in a single expressed causal relation. Consider the following examples:

1. *Daniel killing Ferdinand caused Emily to blink*
2. ▲ *Daniel blinking (as a signal) caused Emily to kill Ferdinand*
3. *! Daniel killing Ferdinand caused Emily to descend from Ganymede*
4. *Daniel being descended from Ganymede caused Emily to kill Ferdinand*
5. *! Daniel blinking caused Emily to descend from Ganymede*
6. *Daniel being descended from Ganymede caused Daniel to blink*

In three of the examples above, (1), (4), (6), the forms are perfectly grammatical and semantically well formed. In two of these, $e_i \xrightarrow{caus} e_j | e_i \in E^U_\varsigma, e_j \in E^U \setminus E^U_\varsigma$, thus any eventuality $\in E^U_\varsigma$ may cause any type of eventuality, with nothing anomalous in the semantics. While for (1), $e_i \xrightarrow{caus} e_j | e_i \in E^U_\tau, e_j \in E^U_{\neg\tau}$, which also seems perfectly reasonable, and acts according to the defined behavior of the classes.

In examples (3) and (5), the expressed causation seem to be semantically anomalous and logically faulty, in the form of $e_i \xrightarrow{caus} e_j | e_i \in E^U \setminus E^U_\varsigma, e_j \in E^U_\varsigma$; this seems to indicate that any eventuality $\in E^U_\varsigma$ cannot be the result eventuality with a different type being the causing eventuality. Finally, example (2) seems to be well defined in semantics and logic, being in the form of $e_i \xrightarrow{caus} e_j | e_i \in E^U_{\neg\tau}, e_j \in E^U_\tau$, but the type of eventuality that e_i belongs to seems to have been altered by the virtue of being involved in this causal relation; now e_i behaves more like those $\in E^U_\tau$, given that it through the causation changed the state of the system. So when e_i is in the causing eventuality position, and the caused event e_j being $\in E^U_\tau$, e_i undergoes a type change of $E^U_{\neg\tau} \rightsquigarrow E^U_\tau$, and takes on the type of the caused eventuality. The *'blinking'* action of *Daniel* in (2) now has a real-world consequence that is acyclic, unlike its unmarked version.

6.2.4 Volition and Expectation

The concept of *volitional entity*, when applied to causality, largely overlaps with the classical concept of *teleology*, where some entity is aware of the expressed purpose for initiating the chain of causation. While there are some other modes in $\theta-$structures that can influence volition and force among different entities, such as in the use of reflexives in Spanish (Maldonado, 1988), no $\theta-$structure type has the flexibility to elucidate the nuance and complexity of the concept of volition in an multi-entity event. According to the concept of volition, among simple transitive and inchoative constructions, verbs can be classified into three groups: those describing strictly volitional events ❶ (drink, throw, ask), those describing strictly non-volitional events ❷ (forget, lose, trip), and those describing events that could describe both volitional and non-volitional events ❸ (drop, break, offend). (Vendler, 1967; Brennestuhl, 1975). According to these studies, the external argument (Jackendoff, 1990) selection among these classes principally differ in that class ❶ requires the appearance of human subjects, whereas classes ❷ and ❸ do not. Instead of 'human' being the criterion, the key criterion here is likely better set as 'animate'. (Butt & Ahmed, 2007)

Both psychological concepts of *volition* and *expectation* can in fact be coalesced in the *mental state* concept within Reinhart's systematization of the $\theta-$roles (Reinhart, 2002; Everaert et. al., 2012). Accurate assessment of volition in the real-world requires knowledge of the internal cognitive processes of an entity *causer* or *agent*, and thus cannot be expressed in a universally agreed on set of possible worlds. This is encoded as the $\theta-$system primitive $+/-\ m$, which is distinct from the causal primitive $+/-\ c$. This mental state of awareness or acknowledgment is central to human's categorization (at least linguistically) of the events surrounding them (Dowty, 1991). And in this system, the semantic differentiation between the roles of *cause*, *agent*, and *instrument*, all of which have $[+\ m]$ in their feature cluster (Jackendoff, 1987; Reinhart, 2000), is whether the mental state is present (agent), absent (instrument), or under-specified/undecidable(cause), for the computable semantics of the construction. The presence and absence of features in each cluster largely determines the number and type of clusters permissible in each frame; one of such is the prohibition of identical clusters in the same frame for most underlying event types (with exceptions from reciprocals and such) (Kremers, 1999; Reinhart, 2002). This is corroborated by Langacker's (2002) findings that languages like Japanese and Hungarian demonstrate semantically distinct types of causatives constructed with the same lexical verb, but with different morphosyntactic cases for arguments. In particular, the morphosyntactic ACC and ABS for the intermediary entity in a ditransitive causative construction normally indicates the intermediary to be non-volitional or non-aware, where as if the intermediary is expressed in DAT or INSTR case it is normally volitional or agentive. This indicates that there are two semantically distinct types of causative constructions cross-linguistically (if we consider the whole range of meanings all to be some manner of causation).

In a number of languages with morphological causatives, such as (Turkish, Finnish, Hungarian), given a *causer* that is non-volitional (stone, water, heat), and an caused event that is volitional (drink, throw, ask), morphological causatives cannot be used, rather only periphrastic construction is possible. (Brennenstuhl & Wachowicz, 1976) We can make cross-linguistic observations that this type of real-world situations requires linguistic constructions that are less tightly integrated between the causing and the caused event. For English, the situation is less complex than languages with morphological causatives, but the capacity for the *causer* and *causee* to possess for volition does make distinctions among different constructions, consider the following examples:

1. *John made Mary evacuate the room, when the cloud of tear gas poured in*

2. *John had Mary evacuate the room, when the cloud of tear gas poured in*
3. *John let Mary evacuate the room, when the cloud of tear gas poured in*
4. *John evacuated Mary from the room, when the cloud of tear gas poured in*
5. *The cloud of tear gas caused Mary [*$_{SC}$ *to evacuate the room]*
6. *? The cloud of tear gas made Mary evacuate the room*
7. *! The cloud of tear gas had Mary evacuate the room*
8. *! The cloud of tear gas let Mary evacuate the room*
9. *! The cloud of tear gas evacuated Mary from the room*

Here, all types of constructions are semantically acceptable, when the *causer* is a volitional entity. But when the *causer* is a non-volitional entity, then the constructions with tighter integration between the events would not be acceptable, only the ‘make’ construction is marginally acceptable with this inanimate *causer*; the ‘have’ construction is not acceptable, since ’have’-causatives entails at least some volition from the *causer*; while the construction with the embedded caused event clause as at least an SC is fine. But when the caused event is non-volitional by nature, such restriction does not apply with the same rigidity. The following are a range of examples with a non-volitional caused event:

1. *The broken railing made Mary stumble onto the lawn, as she tried to evacuate*
2. *The broken railing had Mary stumble onto the lawn, as she tried to evacuate*
3. *The broken railing sent Mary onto the lawn, as she tried to evacuate*
4. *The state of the economy had Mary worried about her prospect for employment after graduation*
5. *The state of the economy worries Mary, in terms of her prospect for employment after graduation*

Both (1) and (2) are analytic constructions that have inanimate *causer* and volitional intermediary of the caused event. Where the *‘have’*-construction has a strong tendency volitionally interpreted, the *‘make’*-causative has no such filter, and both are considered acceptable. Example (3) is not a direct suppletion of the intransitive/inchoative *stumble*, but is the closest to its meaning in this context. In cases where there is a causative-inchoative alternation, such as in the case of ‘worry’ as in Example (5), both a analytic construction and the causative-inchoative verbal form are acceptable, when the caused event is non-volitional.

There are a few types of analytic causative constructions that requires a nuanced treatment with regard to volition. The best of such example are the ‘make’-causatives, consider the following:

1. *John made Mary move away from the lawn*
2. *the dog made Mary move away from the lawn*
3. *the seismic activity made Mary move away from the lawn*

In the case where *John* is the external argument, who is an *animate* and *intelligent* being, the interpretation is strongly biased toward a volitional meaning. In the case where *the dog* is the external argument, which is not *intelligent*, but still *animate* and capable of making decisions, there still seems to be some predisposition toward interpreting this as volitional. Finally, in those cases where the argument is not *animate* or *sentient*, the interpretation must be non-volitional, as in the case of *seismic activity.*

The ‘let’ constructions, on the other hand, may be an exception, since the logical form that is represents has certain peculiarities due to its counterfactual nature, and the following form with a volitional caused event would also be acceptable:

1. *! The cloud of tear gas let Mary evacuate the room*
2. *The gap in the wall let Mary evacuate the room, as the thick tear gas filled the room*
3. *The broken railing let Mary stumble onto the lawn, as she tried to evacuate*

While example (1) is problematic mainly based on its lexico-semantics that *‘tear-gas’* typically impedes movement; where $P_{t+1}(e_1) > P_t(e_1) \implies P_{t+1}(e_2) \leq P_t(e_2)$. It is possible, however, to interpret it the gas giving cover for Mary to escape detection while leaving the room. (2) shows that it is possible for ‘let’ construction to have an inanimate *causer* entity, and contain a volitional caused event. This might be due to ‘let’ construction having the provision $\neg[e_1 \longrightarrow e_2]$ in its logic form, which preempts any presupposed connection of the *causer* to the volition in the caused event, since the *causer* cannot force e_2 to come about by its definition.

6.2.5 Temporality

Temporal distance is also likely an issue in the usage of ‘make’ and ‘have’ causative constructions (Givón, 1993). It does not affect the length of the chain of causation, or otherwise affect the logical structure of the meaning, but only governs whether the e_1 and e_2 are loosely ‘simultaneous’ or ‘contemporary’, as in occurring within the same period of some pragmatically defined temporal granularity. Co-temporality is also related to the previously mentioned issues of control and authority on the part of the *causer*. Consider the following scenarios where intentionality or volition is not an issue:

i *John stretched a thread across Mary’s door to measure its dimension, just as Mary walks into the front door, causing her to stumble and fall as she walks into her room*

ii *John stretched a thread across Mary’s door to measure its dimension, around the time when Mary left for school; as Mary returned in the afternoon, and the thread caused her to stumble and fall as she walks into her room*

iii *John stretched a thread across door of a guest room that would later become Mary’s bedroom to measure its dimension; years later, after John sold the house, the new owner’s daughter Mary moves into the house, and the thread caused her to stumble and fall as she walks into her new room*

iv *John stretched a thread across door of a guest room bedroom to measure its dimension, at the height of the Roman republic; two millennia later, on a trip to visit ancient ruins during her summer vacation, Mary steps into the now abandoned house, and the thread caused her to stumble and fall as she walks into that room*

In all of these situations, it was never John’s intention to cause anyone to fall. The logical sequence in causation is also identical for all three scenarios, where $[_{E1}\quad John \xrightarrow{setup} wire] \xrightarrow{E3} [_{E2}\quad Mary \xrightarrow{trip}]$, with no clear intervening event in the chain of causation. The key difference among these scenarios is the difference in the amount of time that passes between e_1 and e_2. Here are some different causative constructions with varying levels of appropriateness in expressing these scenarios:

1. *John tripped Mary at the door*

2. *John made Mary trip at the door*
3. *John let Mary trip at the door*
4. *John caused Mary to trip at the door*
5. *John brought it about that Mary tripped at the door*
6. *John caused it to come about that Mary tripped at the door*
7. *John provided the circumstances arranged in such a way as to lead Mary to trip at the door*

Example (1) here is clearly unacceptable semantically, without additional modification, for scenarios (ii), (iii) and (iv); and can be acceptable for scenario (i) with some speakers noticing the infelicity with respect to John's intentions. Examples (5) - (7) are acceptable semantically for all scenarios, and causes no pragmatic difficulties for most speakers. Example (4) is generally acceptable for all scenarios, where some speakers may notice infelicities for scenarios (iii) and (iv). These are consistent with the findings in the previous sections, and with the requirements on volition and intention.

The most distinguishing cases are the analytic constructions in (2) and (3). Both are acceptable for scenario (i), when the listener take the sentences to mean that John in neglect did not remove the thread in a timely manner as to preempt any chance of an accident. However, both are unacceptable for scenarios (iii) and (iv), when taken to have the same meaning of John's negligence, with the only difference from (i) being the quantity of time that has passed between e_1 and e_2. For scenario (ii), the semantic acceptability for both seem to be varied, with the 'make' construction may or may not be acceptable depending or various assumption that the listener makes, while the 'let' construction is likely not to be. The use of deeply embedded causatives expresses $e_1 \xrightarrow{caus} e_2$, while providing the connotation that ❶ no volition on the part of the *causer* in e_1 is necessary, ❷ the temporal separation between e_1, e_2 can be considerable, and ❸ the logical sequence between the two can also be rather convoluted.

There may be a number of different semantic and pragmatic factors at work, and it is difficult to tease them apart. But it seems to us that the key distinction here is whether the *causer* and the *causee* are contemporaries at the same location. If the *causer* John and the *causee* Mary likely to not have awareness of each other's existence across vast swathes of time, then these analytic constructions are not likely to be acceptable. Without additional context, we don't actually know whether they knew each other's existence in most of these scenarios, but that knowledge can be reasonably inferred through the amount of time that has passed. So temporal inference will likely be a factor in analyzing some of the analytic causatives that we observe, and analytic causatives can help elucidate the likely amount of time passed in certain situations.

6.3 Relevance and feasibility of each class

As we have seen, a key difficulty in discovery and processing of causality is the large of variation in possible linguistic expression of the same causation event. If it is true of most causation events that the information pertaining to a specific one is presented once in the SN stream, and that the users of the network may use any of the range of variations of expressing causation, then it would be necessary to accurately locate and extract all instances of causative constructions. In this case, we must devise a reasonably precise method for each of the causative construction types in order to obtain an accurate picture of causation events within the network. Given that there is a wide variety of strategies discussed in Section 6.1 (as well as other unmentioned), this would require a highly complex task with many sub-modules each with a different strategy specifically devised for it.

However, the organization of linguistic data in many social networks mitigates the need to extract all types of causative construction. The predominant form of linguistic data are those present in content streams, and when a shared experience such as a social gathering among users, or a time sensitive new item are discussed there within, it is normally discussed multiple times by different users each with his/her own way of expressing and phrasing predications that have the same real-world referent event. In these instances of streams, it is not necessary to capture all different classes of causative constructions, but only several of the frequently occurring classes of causatives in SN data would suffice to achieve good coverage. Thus, we need to identify the classes that have the mostly available methods that are also reasonably frequent.

6.3.1 Feasibility of causative-inchoative alternation

The most obvious approach to finding causative-inchoatives is to rely on a knowledge base to inform us of the lexical entries that are relevant. However, There is currently no complete database for all lexical entries involved in causative-inchoative alternation for any language; although wiktionary on-line contains a semi-comprehensive list of causative-inchoative pairs that are homophones. However, as discussed earlier, in many languages, including English, there is a high rate of morphological identity and correlation between forms of verb entries describing the same object/material transition with (causative) and without (inchoative) the description of an external *causer* role. In English, most such pairs are identical in surface forms, such as *melt*, *clear*, *fill*, *burn*.

Also, the vast majority of causative-inchoative pairs are in fact observed, in their surface forms, to be a subset of those content verbs from the same free-morpheme that polymorphically appear in monadic and dyadic frames. The inchoative form usually exhibits a single *theme* argument at surface, and the causative form exhibits the *theme* argument and some actor as *causer*, as expected by the semantic theory from the previous section.

1. $\langle_{THEME}$ *the bomb* $\rangle$ *detonated*
2. $\langle_{THEME}$ *the bomb* $\rangle$ *detonated due to* $\langle_{CAUSE}$ *John's action* $\rangle$
3. $\langle_{CAUSER}$ *John* $\rangle$ *detonated* $\langle_{THEME}$ *the bomb* $\rangle$

As we saw in Section 6.2.2, the first example of an inchoative is only potentially causative, given the possibility that the event occurs as part of a natural progression of the current state of events. The inchoative (2) has the same structure as (1), but now the cause is explicitly represented, as in the causative-inchoative construction of (3); these are the minimum level of structures needed if a causative were to be ascribed to the surface form with confidence.

In English surface structure, there is another set of semantic frames that frequently exhibit this type of polymorphism, consider the following examples:

1. *John piloted yesterday*

2. *John piloted a commercial jet yesterday*
3. **! a commercial jet piloted yesterday*
4. *Mary surfs on the weekends*
5. *Mary surfs the waves on the weekends*
6. **! the waves surfed on the weekends*

This is usually an under-specification within the semantic frame that allows some generalization through not specifying the internal argument of the frame. But here, the semantic classes of arguments residing in direct object position in the transitive forms, such as *vehicles*, in *"John piloted ⟨vehicle⟩ yesterday"*, cannot appear in subject position of the corresponding intransitive examples This behavior is fundamentally different from the ergative behavior within the causative-inchoative pair.

6.3.2 Feasibility of lexically ditransitive causative

Causative-active alternation is a much smaller class than that of the causative-inchoative alternation. Analogous to the behavior of the causative-inchoative, the pair should exhibit dyadic and triadic forms in an exhibition of polymorphism in the semantic frame. But unlike the majority of causative-inchoative alternations, the causative-transitive forms are usually not *zero-derivations* of their transitive counterpart. So the vast majority forms within these pairs are morphologically unrelated, and recognizing the alternation relation at the surface forms cannot be relied on. As we have seen, the modification of the semantics from the transitive form to the causative-transitive form is also more complex, and there usually are multiple targets for the transformation of a single transitive.

Determining the the set of ditransitives is possible with any sufficient set of data in the language (not necessarily the target data in social media), and with sufficient diversity of adicity among its verbs. One of the issues in finding the adicity is the separation of arguments from adjuncts, whereas bare NPs are highly likely to be arguments barring certain special classes such as time expressions, but PPs in the surface form may be either arguments or adjuncts, depending on the lexical verb and the remaining arguments:

1. *John sent Mary some flowers*
2. *John promised Mary a vacation*
3. *the accident earned Mary a ticket*
4. *John sent some flowers to Mary*
5. *? John promised a vacation to Mary*
6. *?! the accident earned a ticket to/for/at/by/... Mary*

As in these example, the ability to use PP in place of an direct NP argument in a triadic predicate varies widely among different lexical verbs, with many having marginally acceptable forms. There are some among the triadic predicates that, when their frame takes three full NPs directly, it makes them at best marginally acceptable, such as in the following:

1. *John will win the first place prize for Mary*
2. *? John will win Mary the first place prize*
3. *John will hunt some wild pheasant for Mary*
4. *?! John will hunt Mary some wild pheasant*
5. *John will design a new room for Mary*
6. *?! John will design Mary a new room*

But we know that all three of these components in each case are arguments, and not adjuncts, because when replaced with a pro-form, these seem to be able to freely appear in bare NPs:

1. *John will win her the first place prize*
2. *John will hunt us some wild pheasant*
3. *John will design me a new room*

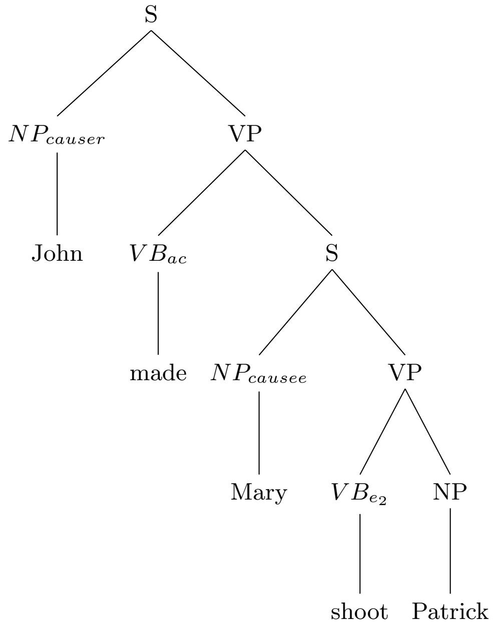
So in terms of the adicity of lexical verbs in English, we can determine whether a verb is triadic, with three arguments, by looking at whether there is a significant fraction of 3 full-NP forms in the distribution of its frames in a sufficiently large corpus, or if there is a significant fraction with 2 full-NP plus a pro-form.

For any lexical verb with triadic frames that has been identified, in English, the determination of the corresponding dyadic frame through some valency lowering operation is a far more difficult task. The primary reason here is because English is a highly analytic language, and contains little morphological means to express the same underlying event with differing valencies and structures of the semantic frame. Even with a traditional corpus of sufficient size and training data, there is no a priori reason to assume that the linear context of the triadic frame would be generally similar in any way to the corresponding dyadic, since they express fundamentally different real-world event types.

This, along with the fact that *causative-transitive*s are a small class relative to other causative construction types, makes hand-building a semantic resource in the recognition of this type of causative construction very attractive. The only ways of guaranteeing reasonably accurate detection is to rely on a existing knowledge base of causative correspondence to select one corresponding dyadic predicate and its lexical verb for each lexical verb found to have a triadic frame, if the triadic predicate is determined to be causative, by employing a semanticist to do the same, or to train a model on some data-base that is labeled according to these correspondences.

6.3.3 Feasibility of analytic causatives

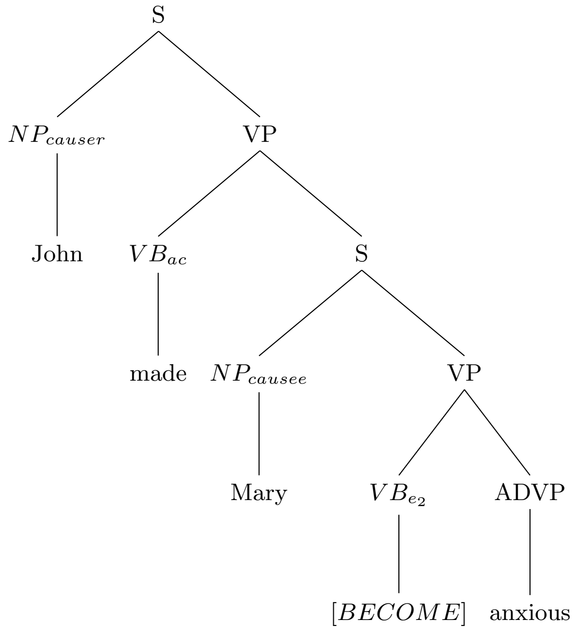
While the constructions in causative-alternations are made less transparent by the lexical expression of these classes of causatives, especially in cases where a pattern of suppletion is observed (most causative-transitives) where individual verb lemmas need to be recognized through a knowledge-base as causative in nature, the analytic constructions do not suffer from such lack of systematicity. Analytic constructions uniformly contain a small set of modality expressing verbs, such as *'make'*, *'have'*, *'let'*, as well as specific and highly recognizable features in its syntactic tree (given a specific parser). This class is among the most amenable causatives for automatic detection and extraction due to its morphosyntactic transparency and uniformity. All instances of this class exhibits this syntactic structure when parsed correctly, where the term VB_{ac} refers to the analytic causative verb, and VB_{e_2} is the predicate of the caused event, and where $NP_{theme/patient/goal}$ is relevant only if the embedded predicate is dyadic:



Other structures with the same modality-expressing verbs ‘make’, ‘have’, ‘let’, but without content verb in the predication of e_2 in the archetypal analytic causative. This related type contains an adjectival argument, which has a head of ADJ(JJ), PP, or VBD at the surface, that functions as a copula prediction. These can also be analyzed as containing a covert SC, with *BECOME* as part of the predicate, include the following:

1. *John had Mary worried about the driving test*
2. *John got Mary on the dean’s list with his connections at the school*
3. *John made Mary anxious with his stern manner*

These can be analyzed to include a embedded phrase structure analogous to the *CAUSE(BECOME(A))*-construction of those within the causative-inchoative alternations, where A is the aforementioned internal argument, and a parse that is structurally similar to that of the typical analytic constructions. This may take a variety of different forms when processed through an automatic parser, but a valid parse consistent with that of the archetypal analytic construction is as follows:



But other syntactically (at least at S-structure) identical, and also constructed with the same modality-expressing verb, do not actually have any semantic causality associated with them. This is especially true with many 'have'-constructions, as seen in these examples:

1. *John got [$_{PRED}$ Mary mixed up with someone else's daughter at the PTA meeting]*
2. *John had [$_{PRED}$ Mary pegged as one of those over-achievers]*
3. *John has [$_{PRED}$ Mary going off to college in a few years]*
4. *John has [$_{PRED}$ Mary winning the first place in a national contest]*

For all of these examples, whether the predicate is inherently monadic or dyadic, the contained predication about the *causee* Mary does not refer to any event or process that Mary is involved in in the real world. These types generally represent some irrealis predication in the mind of the *causer* about the state or action of the *causee*. These predications express some type of divergence between the real world extensions of these states or events, and the perceptions of the entity denoted by the external argument of the matrix clause, in other words, his/her misconception of reality. For analytic constructions that have an inner embedded content verb that in aorist form, this is almost never an issue; as in the examples below, where the aorist form of the verb always indicates causation, with no ambiguity in this regard, across all open classes of verbs:

1. *John had [$_{PRED}$ Mary reach the peak of Mount McKinley]*
2. *John had [$_{PRED}$ Mary hit the policeman with a rock]*
3. *John had [$_{PRED}$ Mary climb Mount McKinley]*
4. *John had [$_{PRED}$ Mary run from the policeman]*
5. *John had [$_{PRED}$ Mary sing in front of the auditorium]*

while other manifestations of the embedded content verb, such as the participial form, exhibited very different behavior. There is a clear distinction among different lexical aspects of content verbs, particularly along the *durational* (semelfactive and achievement) / *nondurational* (activity and accomplishment) distinction. In participial forms, *nondurational* verbs in the embedded clause position, such as 'reach' and 'hit', do not exhibit a large amount of ambiguity in interpretation, these is a strong preference to be interpreted by most speakers as irrealis as in examples (1) and (2) below. On the other hand, the construction with *durational* verbs are highly ambiguous, and equally likely to be interpreted in a causative or irrealis way, such as (3), (4), and (5) below:

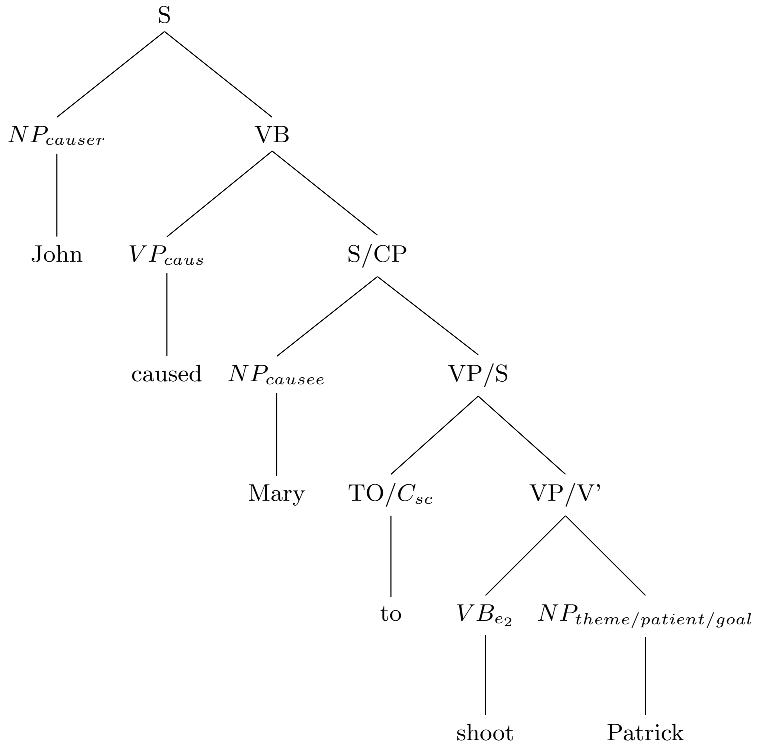
1. *John had [$_{PRED}$ Mary reaching the peak of Mount McKinley]*
2. *John had [$_{PRED}$ Mary hitting the policeman with a rock]*
3. *John had [$_{PRED}$ Mary climbing Mount McKinley]*
4. *John had [$_{PRED}$ Mary running from the policeman]*
5. *John had [$_{PRED}$ Mary singing in front of the auditorium]*

For those that correspond to monotransitive causatives (the embedded clause has only one obligatory argument, this test is inadequate, as examples (3), (4), and (5) above can be interpreted either as causative or irrealis, especially when the embedded verb is a participle. This seems to be related to the ability of durational verbs to encode the embedded event that is concurrent with the presumption made by the entity of the matrix clause; this allows the entity to make assumptions about yet incomplete event.

Since there is no morphosyntactic cue to distinguish these from the true monotransitive analytic causatives, the only consistent way of deciding whether some particular ‘have’ or ‘get’ construction is causative or irrealis is to decide on the basis of either the lexical entry of the contained predicate verb, or use the potential inferences from the social network structure surrounding the involved entities. If we require that this decision be informed on the basis of only the text that directly expresses the causal predication, without the luxury of examining the surrounding linear context, which does not usually exist in SN, then devising a discrimination task between lexical verbs that can be involved in analytic causatives and those that do not becomes necessary (lexical verbs now carries the bulk of the discriminating information for this task).

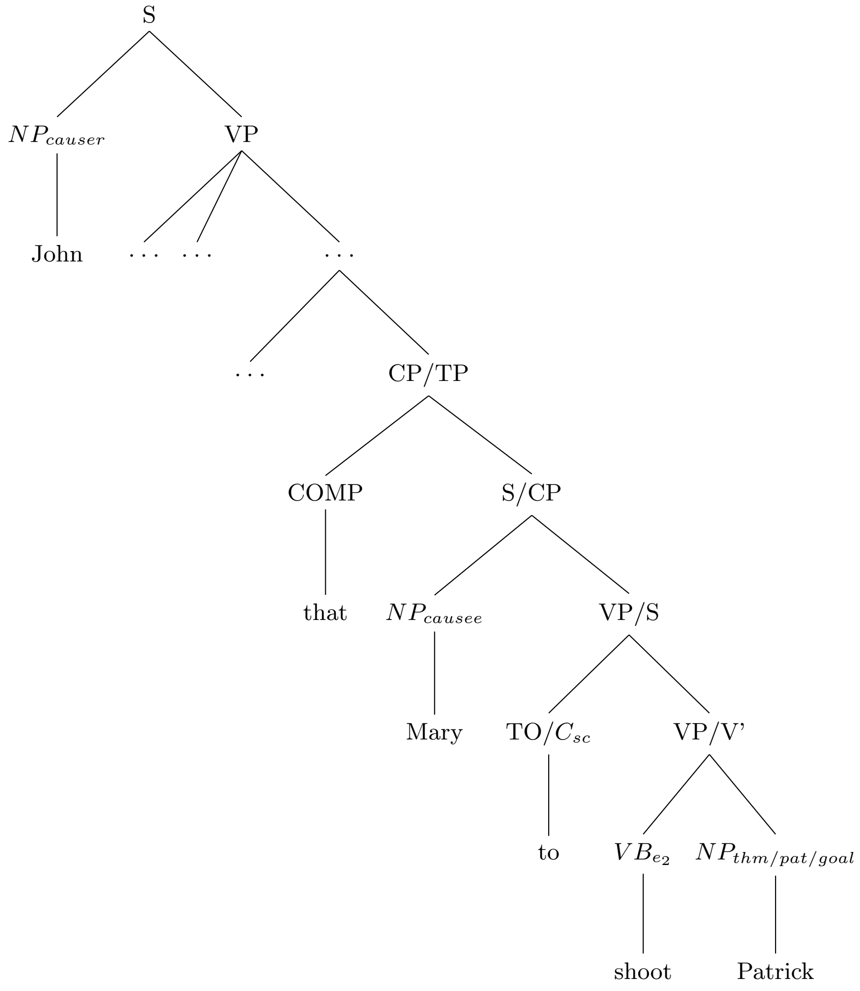
6.3.4 Feasibility of embedded multiple clause structures

Similarly, causality expressed through the use of a pair of embedded clause representing e_1 and e_2 respectively also generally contains a morphosyntactic structure that can be readily recognized in parsed data. These generally correspond to the instances where more than one EDU (Carlson et. al., 2003) is contained within. The general syntactic structure of this class resembles the following in automated parsers, with some variation depending on the specific parser, again with $NP_{theme/patient/goal}$ present only if the caused e_2 is dyadic:



There are other verbs that can take the place of 'cause' in the matrix clause; whereas the use of 'cause' in the surface structure is non-specific with regard to the predicate e_1, other lexical verbs in its place such as *'ask'*, *'require'*, *'command'*, *'force'* imparts more specific semantic content to e_1, many of these fall within PCU verbs discussed earlier, as well as those indicated by Wolff et. al. (2002) and Wolff & Song (2003) as psycholinguistically evidenced to be frequently involved in causal relations among entities.

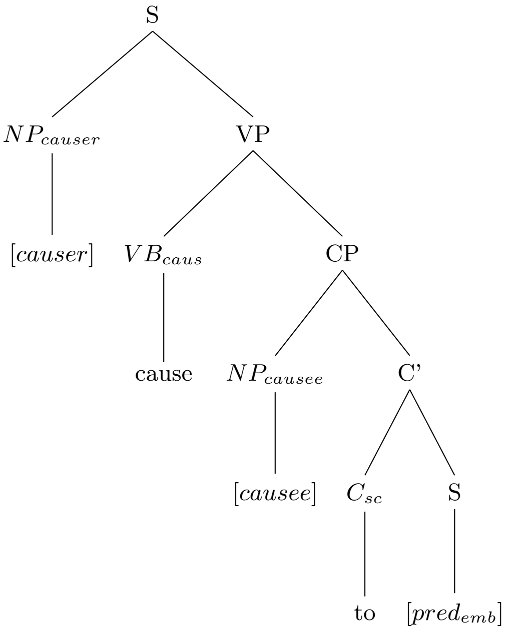
There is somewhat more complication with the modality expressing component of the structure, given that not only lexical items such as 'cause' may play that role, but also linear structures such as *'brought it about that'*, which has internal syntactic structure may also serve the same purpose semantically. Such structures can potentially exhibit varying amount of syntactic complexity with unbounded levels of CP embedding, the additional syntactic structure lengthen the distance between the entity in the external position of the matrix clause, and the embedded predicate, and in part serves to increase the rhetorical distance between them. Some of these intervening clauses may not contribute to the compositional semantics of the overall structure, which occurs when one intervening predicate simply return one of its input parameter, which would the contained embedded clause. As we have seen in Section 6.2.3, these have a tendency to increase the limit on the number of permitted links in a represented causal chain. This is similar to the cross-cultural and cross-linguistic function of indirectness in language (Ogiermann, 2009), and indirectness frequently is used to make the expression increasingly illocutionary (Leech, 1983; Brown & Levinson, 1983), which would allow both more politeness and a more fuzzy notion of causation. Here is a typical example:

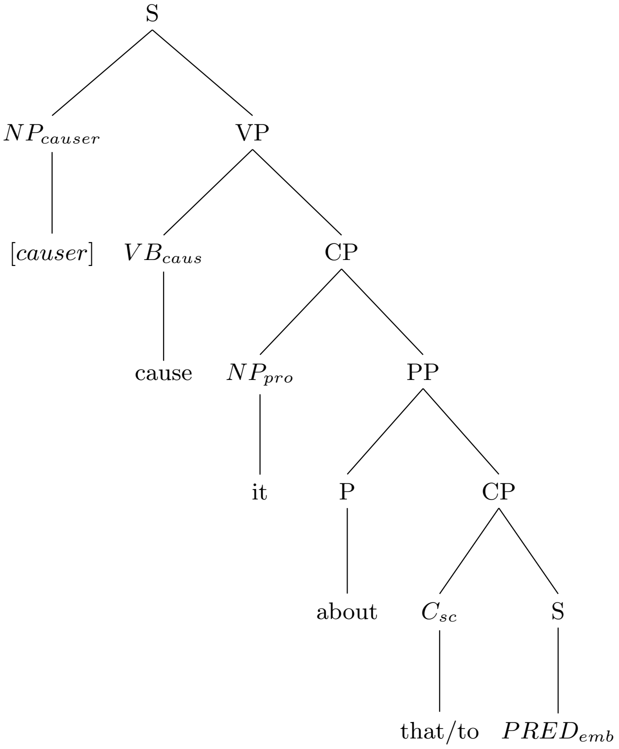


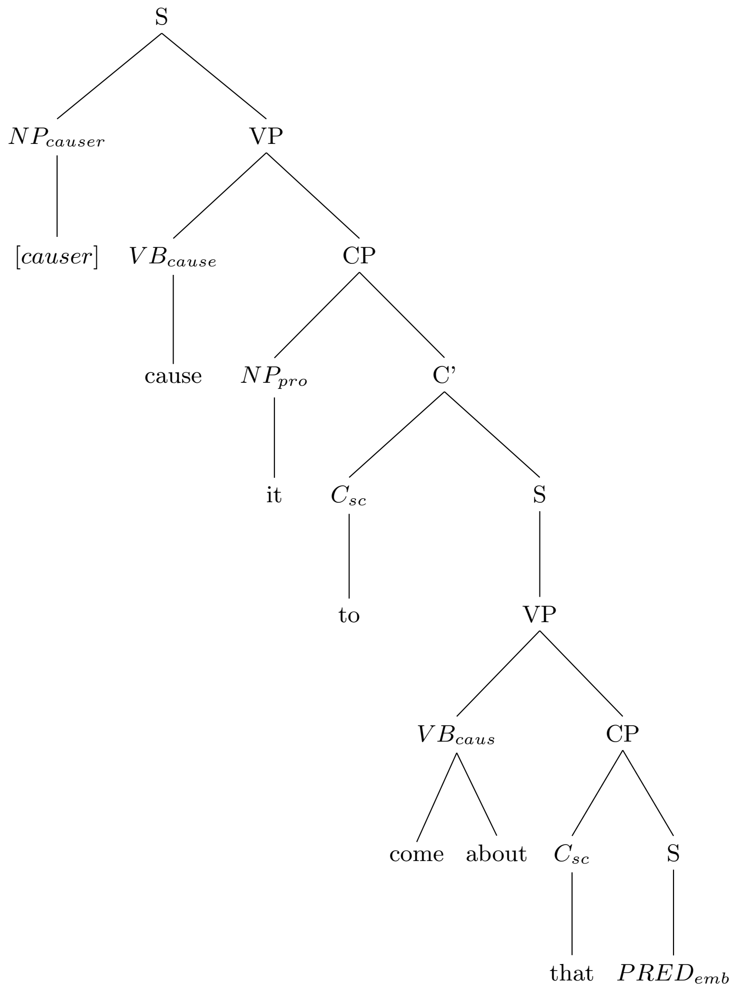
The omitted part in the syntactic tree above has a large amount of lexical and syntactic variation permitted, each of these variants expresses the meaning of the same multi-clausal causative construction, and sometimes also expresses meaning about the predicate referring to e_1 depending on the choice of the lexical verb for the matrix clause. Some examples structures, ranged in order of syntactic complexity are as follows, where the term $SPEC_{mtx}$ refers to the external argument of the matrix clause (causer), the term $cause_{emb}$ refers to the external argument of the innermost embedded clause (causee), and $PRED_{emb}$ refers to the embedded clause representing e_2 in the structure:

1. *causer*$_{mtx}$ *cause [*CP_{SC} *causee*$_{emb}$ *to [*$PRED_{emb}$ *... ...]]*
2. *causer*$_{mtx}$ *make it so [*CP_{SC} *that [*$PRED_{emb}$ *causee*$_{emb}$ *... ...]]*
3. *causer*$_{mtx}$ *bring it about [*CP_{SC} *that [*$PRED_{emb}$ *causee*$_{emb}$ *... ...]]*
4. *causer*$_{mtx}$ *cause it to come about [*CP_{SC} *that [*$PRED_{emb}$ *causee*$_{emb}$ *... ...]]*
5. *causer*$_{mtx}$ *cause it to arrive at the outcome [*CP_{SC} *that [*$PRED_{emb}$ *causee*$_{emb}$ *... ...]]*
6. *causer*$_{mtx}$ *arranged the affairs in a way for [*CP_{SC} *causee*$_{emb}$ *to [*$PRED_{emb}$ *... ...]]*
7. *causer*$_{mtx}$ *predetermined the circumstances in such a way as to lead [*CP_{SC} *causee*$_{emb}$ *to [*$PRED_{emb}$ *... ...]]*

Each of the above examples have structurally identical parts of the syntactic representation from the parser from the matrix VP and above, as well as from the embedded S and below, sometimes with the exception of some movement of the NP of the causee argument to SPEC of the embedded CP when the embedded clause is non-finite. The variation lies in the structure between these two tree nodes of the outer matrix and innermost embedded clauses. The wide range of possibility also leads us to believe the complexity of the intervening portion between the matrix VP and the embedded S is unbounded, and a structure of arbitrary level of embedding in natural language can represent a single causal relation. Examples of the syntactic structures are below:


S
NP_{causer}
VP
[$causer$]
VB_{caus}
CP
cause
NP_{causee}
C'
[$causee$]
C_{sc}
S
to
[$pred_{emb}$]


S
NP_{causer}
VP
[$causer$]
VB_{caus}
CP
cause
NP_{pro}
PP
it
P
CP
about
C_{sc}
S
that/to
$PRED_{emb}$



These come at minimum with a full CP embedded in the matrix clause, and the more complex forms in the range also may come with an additional levels of embedded clauses as in the last example. As stated in Section 6.2.3.1, the more complex the causative construction is, the more it generally tends to license longer chains of causation. But there does not appear to be large variations among different multi-clausal causative constructions with respect to this property, all of which license chains of considerable number of links.

In terms of lexical variation of embedded multi-clausal constructions, especially with regard to the form [*causer* [vb_{cause} [*causee* [$PRED_{emb}$]]]], in addition to semantically neutral causative lexical items such as 'cause', 'bring about', the place of vb_{cause} can also be filled with many of the verbs ∈ *manner-of-causation* mentioned in Section 6.1.5. The pure causative forms like 'cause' in vb_{cause} position we will term *base multi-clausal causatives*, or $\mathcal{BMC}$. The usage of these $\in$ *manner-of-causation* $\backslash\mathcal{BMC}$ in place of $\mathcal{BMC}$ constructions, provide a richer meaning to the predicate referring to e_1, giving a distinct mode of action in the causing event, with the same effect on the causee. Some examples are in the following:

1. *John **roused** Mary to win the contest*
2. *John **prompted** Mary to finish her homework*
3. *John **inspired** Mary to learn a second language*
4. *John **incited** Mary to shoot Patrick*
5. *John **compelled** Mary to rob the bank*
6. *John **bullied** Mary to rob the bank*
7. *John **bribed** Mary to rob the bank*

But we can also observe that not all complements of this class of verbs refer to an e_2 that is causally linked to the original event indicated by the matrix verb. Example (6) and (7) not only has ramifications for the nature of the causing event e_1, but also imposes selection restriction on the intermediate agent 'Mary', constraining it to be an animate entity. Many such constructions actually has a reading that is classified as an *explanatory* relation, as in the sense of *'in order to'*, rather than a causal relation. When such a reading is active, the event represented by the embedded predicate e_2 need not have a real world occurrence, but can be regarded as a proposition that may not be realized, such as in the following:

1. *John **roused** Mary (in order) to prepare for breakfast*
2. *John **inspired** Mary (in order) to make her future brighter* (in the sense of John making Mary's future brighter by inspiring her to do greater things)
3. *John **compelled** Mary (in order) to lead her into a life of crime*

In sum, the feasibility of extracting multi-clausal constructions from SN data requires the solution of two discrimination problems. One is discerning, for those candidates containing some $\in \mathcal{BMC}$, what are the tell-tale structural distinctions between the matrix VP and the innermost embedded S nodes that indicates the likely presence of a multi-clausal causative construction. The other is discriminating, for those candidates containing some $\in$ *manner-of-causation* $\backslash\mathcal{BMC}$ as the matrix verb, what types of complements indicate causal relations to the predicate of the matrix verb, as opposed to an explanatory relation.

6.3.5 Feasibility of causal discourse connectives

Causality expressed through the use of multiple disjoint clauses linked by discourse connectives and other complex discourse structures rarely occur in social media data. since the linguistic input of users into the SN database is mostly fragmented, and many social media services such as Twitter impose rigid limits on the length of any coherent linguistic input. So this class of causal constructions is not practical to pursue within the context of social media.

6.3.6 Feasibility of adjunctive structures that are not explicitly causal

In principle, SN data could contain copious amounts of non-explicit causal structures in pairs of adjoined clauses. Since these do not contain explicit linguistic structure that force causality, and could potentially express other types of relations than causal, so we need a robust and probabilistic definition of causality for such cases. We will need some concept that functions in a similar manner as *causal potential* (Beamer & Girju, 2009), but that is also accurate can flexible enough to not rely on cues in the larger context beyond the target clauses. This can be done taking advantage of SN data representing events in a varied and redundant fashion, and the the ability of its users to express causality beyond simple declarative statements, as we discuss in a later section.

Chapter 7

Adjoined Causal Structures:

In this section, we perform a pilot study on the automatic extraction of adjoined frames causal constructions, which is one of the two types (the other in Chapter 8) of structurally complex causal construction that we earlier identified. This module will treat all extracted frames as a linear sequence in the corpus, and attempt to detect any causal relation between those pairs that are defined to be adjacent, primarily by examining the their linear structures outside the core frame components.

7.1 Adjunctive Causality

There are a number of types of parallel structures, most which are composed of adjacent pair of clauses representing the frames e_i, e_{i+1} which have some potential to express causality, but in of themselves have no explicitly causal linguistic features, some of these have been mentioned in Section 7.2.1.

7.1.1 Basic description

Any pair of adjacent clauses in a text corpus $\langle c_i, c_{i+1} \rangle$, where i is some index of clause's linear position within the corpus, could potentially be causal. For the Penn Discourse Tree Bank (PDTB)'s annotation scheme, in which causality is a prominent class of annotated relations

(Webber & Joshi, 1998; Joshi et. al. 2006), there is always some *connective* between the two adjacent clauses for the causal relation to be manifest, such as: *'both side have agreed that the talks will be most successful* <u>*if*</u> *negotiators start by focusing on areas that can be most easily changed'*. Since for our study, we do not want to be restricted by the presence of specific lexical cues in order to locate adjunctive causal structure, the cases where no explicit causal lexical item is present would be more interesting, such as *'John turn on the faucet, water filled up the sink after 10 minutes'*. Our procedure described in this section was able to provide a ranking system that recognizes many adjoined pairs to be highly likely to be causal, even when there is no such lexical cue in the form of a *connective*; the following are examples extracted by our adjunctive causal procedure from the testing set of BNC:

1. (a) the design is open ended
   (b) it could be adapted to separate passengers from their cars
2. (a) **the government** has shied away from forcing unions to discipline members
   (b) [it] has put proposals to curb strikes in essential services on the back burner

A well designed general procedure for adjoined causality should be able to detect under a variety of types of data. For (1) above, an *'open ended design'* could be easily *'adapted'*; and for (2), the decision of the government to *'shy away'* from any coercive actions on unions provides the arrangement of events *'that put on back burner'* some specific proposal that would restrict union activities. Although our adjoined causal structure module is capable of locating some causal relations with explicit connectives or with statistically significant co-occurrences between two predicates, the above examples show that our algorithm is not restricted to them.

7.1.2 Issues specific to adjoined causal structures

For adjoined causal pairs of clauses, which may or may not have any causal connective in between, the necessary information that allows a speaker to determine whether such a pair is causal (whenever such determination is possible at all), it is through the structure that surrounds this pair of clauses, as well as information in the wider context. In such non-explicitly causal cases, there may or may not be any structural relation between e_i, e_{i+1}, as in one may not be embedded in another, nor are they necessarily embedded in a common matrix clause. So in order to detect such pairs that express causality, it is ideal to arrive at a method that does not depend on whether or not there is any hierarchical structural cue between e_i, e_{i+1}. We simply treat the corpus and all of the frames as a linearly ordered sequence of frames (with some caveats discussed later), thus some of the detected causalities through this module should intersect with those from the embedded causal module in Chapter 8, while others should be unique to this module. In order to treat the cases where the pair has some hierarchical relation and those cases where there is not, the most sensible approach is to attempt to elicit information from a linearly sequential structure that contains the pair $\langle e_i, e_{i+1} \rangle$, since such a pair can always be structured sequentially.

Many of these pairs have certain connectives that link the individual frames into the first level of discourse semantics, which is one type of cue in the stream of text that contributes to the determination of causality in the pair. For instance, a $\langle e_i, e_{i+1} \rangle$ could have *"whenever* $[_{Clause1}$ *Entity*$_x$ *...] , it occurs that* $[_{Clause2}$ *Entity*$_y$ *...]"*, which would be regarded as a type of conditional expression mediated by the extra-frame information conveyed. There is a limited set of such connective pairs as cues for potential causality. However, the coverage of such local discourse markers are not always present, and there is also a significant probability that pairs containing them are not logically causal. A few of such pairs, e.g. *"given the fact that .... , it necessarily follows that .... "* do seem always convey causality, but it is generally not the case with most connectives. Consider the following examples, showing that such local discourse connectives are neither sufficient nor necessary for conveying causality in adjoined clausal pairs:

1. *whenever Ian traveled overseas for an expensive vacation on the Riviera, it occurs that Jane takes a summer course to improve her qualifications at future jobs*

2. *in the case where the sparrows feed on the leftovers by the tourists, then we expect to also see swarms of ants at that location*

3. *On Saturday, Ian bought some tequila, nectar, orange liqueur, and a bag of limes; that night, Jane made some margarita for the house party*

4. *The Russian military supplied a BUK SAM launch system to the separatists in the Ukraine; this week, a commercial airliner flying from the Netherlands to Malaysia was shot down over rebel controlled territory*

For the first two examples, even given fairly clear pairs of discourse connectives in the structure around the two clauses of $\langle e_i, e_{i+1} \rangle$, we observe that there is still no causal relation in either case. For (1), the relation between the events is mostly likely one of contrast, where Ian is filling his summer with leisurely activities, Jane is working toward her future, there is no plausible logical connection. For (2), the most probably relation here is to a common cause, that of the dining of the tourists, the leftover food of which attracted sparrows and ants alike, the former is merely an indication of the latter. For the latter two examples, the adjoined pairs are very likely causal, without the presence of any connectives (although the temporal qualifications played some role). For (3), the ingredients purchased was those called for in the recipe of the drink; and for (4), the supplied SAM system is capable of downing a airliner. Thus some real-world knowledge and ontology with regard to the lexical items involved would aid in detection of causality, as well as the structures within and around the frames. Thus, an implementation with a closed class of local discourse markers is insufficient and often ineffective at detecting causal relations among adjunctive structures, so we need a more sophisticated analysis of the sequential structure within and around the frames of $\langle e_i, e_{i+1} \rangle$.

7.1.3 Importance of adjoined structures in causality

Adjoined pairs of clauses could be causal whether or not there is some explicit lexico-syntactic cue, such as causal discourse connective, to aid in its cognitive recognition as causality. There are traditional methodologies focused on locating causality of two clauses through *discourse connectives* (such as PDTB's approach). There are also those that detect the causal relation between **type**s of events (such as Beamer & Girju 2009; Riaz & Girju 2010, or Do et. al. 2011) which focuses on the relations between *predicate*s. But these methods are not appropriate for location causal relations between pairs that have no explicit connectives, have predicates that occur sparsely in a genre, or that are related through a more complex way than simple event types indicated by the predicates.

1. (a) for an informal dinner party, do not hide your guests from one another with a giant display
   (b) rather use a series of small container groups that suit the shape of your tables
2. (a) your boyfriend is finding it hard to come to terms with the prospect of fatherhood
   (b) [he] is taking his resentment out on you

Neither of these example has a explicit discourse connective that is indicative of causal relations. For (1) above, hiding guests from one another with a display is not normally related to container groups or the shapes of tables, except for very specific circumstances that consists of a dinner part (where people sit around tables), with specific layout of the room (where individual tables are on different sides of rooms), and given a specific type of center piece (here display has a specific sense in meaning applicable to the situation). So it is generally unlikely that predicates *hide* and *use* without all of these caveats and configurations would be identified as causal, and there is very small likelihood that this situation would be repeated multiple times in a corpus.

For (2) above, likewise *finding ... hard* and *take ... out on ...* would not be likely to be found as causal, if we only take into account whether these predicates types generally have a causal relation between them, without the specific circumstances. Our system was able to identify these as highly likely to be causal, because of the specific sequential structures detected, beyond any connective or the core predicates of the two adjacent clauses. For both examples, some deep semantic understanding of the events and their settings are required for humans to recognize their causal relation. So this type of study is an important component to the detection of a large variety of causal relations between pairs of clauses, in achieve good coverage over sub-types that previous method do not normally cover.

7.2 Theoretical foundations

In order to further understand adjoined expressions in general, and to formulate a viable way toward identifying causal pairs among such structures, we first need to discuss a few important related concepts in theory.

7.2.1 Role of probability

Traditional methods in defining and determining such adjoined causal pairs, with or without explicit lexical cues in between, heavily involves the use of probability theory. A strong version of this statistical causal relation (Cartwright, 1979) is sometimes stipulated, where $\mathcal{C}$ is the set of potentially causal factors for some event e, and let C_{-i} be the entire set less some factor c_i; also let the set functions $\mathcal{F}_{-i} = \{f : [n]_{-i} \rightarrow \{0,1\}^{n-1}_{-i}\}$ (given all subsets ignoring the ith position) give out all possible state description that picks whether each factor $\in C_{-i}$ is modeled into the scenario. So in this version, $c_i \xrightarrow{caus} e \quad iff \quad \prod_{f \in \mathcal{F}} P(e|\{c_i\} \cup f(C_i)) > P(e|f(C_i))$; so this extraordinarily strong definition requires that for all combinatorial possibilities of whether each of the the other factors $\in \mathcal{C}$ is present (except the target factor c_i), the presence of the the target factor must raise the probability. This has been regarded by many as an unrealistic way to link statistical metrics with the definition of causality (e.g. Hardcastle, 1991). If we use the Cartwright statistical definition for causal relation, then detecting causation between andy c_i and e would be straightforward and unquestionable, but would also put a severe restriction on what can be considered causative, and would run counter to many intuitions of humans for what constitutes causation; thus we likely need a more flexible version.

The most obvious and least contested property of linguistic causative is that it requires a precedence relation between the two events such that $e_1 \prec e_2$ (Beamer & Girju, 2009; Burks, 1951), which is a fundamental property of space and time. Another observation made by many is that when a speaker utters a causative construction, he/she expects that e_1's occurrence will raise the chance of e_2's occurrence in the chance theory of Mellor (1995; Suppes, 1970); such that there is a change of probability $P(e_2|e_1) > P(e_2|\bar{e_1})$, $e_1 \prec e_2$

(Mellor uses the notation $Ch_C(E)$ and $Ch_{\neg C}(E)$, where C is the cause, and E the effect). Sometimes contains the additional operator of $d(\cdot)$ that determinizes the causal condition, as in fixing an independent variable in an experiment, $P(e_2|d(e_1)) > P(e_2|d(\bar{e_1}))$

This is based on several predication about the relation generally recognized by speakers, including the aforementioned precedence relation, also that e_1 and e_2 are mutual evidence for one another, that the observation of e_1 in part explains the occurrence of e_2, as well as that e_1 creates some partial precondition for e_2. This is fitting for our use in the description of an SN, since Mellor's (1995; Edgington, 1997) framework of defining causality relies on the fact that cause and effect in a causal relation are propositions that can be verified or falsified, and can entertain probabilities of verity, which fits well with the concepts of events as relations between entities. The formulation of $P(e_2|e_1) > P(e_2|\bar{e_1})$ is in fact a quantified version of the expression of *sufficiency*, which is the deterministic $e_1 \Rightarrow e_2$. Also seems to be relevant for causality is the concept of *necessity*, in its deterministic form is such that $\bar{e_1} \Rightarrow \bar{e_2}$, or $e_2 \Rightarrow e_1$. This also has a corresponding probabilistic expression of $P(e_1|e_2) > P(e_1|\bar{e_2}), e_1 \prec e_2$.

Another relevant type of relation closely corresponding to our notion of causality is the counterfactual relation, (Lewis 1979 / 2000, Schaffer, 2004; Menzies, 2009) which can be paraphrased as *"had $X = x$, then $Y = y$"*, holding X as an independent variable by treating it as a constant. Following Pearl (2000) and Tian & Pearl (2000), we will express the counterfactual statement above as $Y_x = y$. This, like the aforementioned $d(\cdot)$ operator presumes some fixed independent variable $X = x$. So this may not correspond to the conditional distribution in the real world, could be thought of as some other possible world scenario with something normally variable in our universe to be fixed. .

7.2.2 Counterfactual causality

Counterfactuals have well developed formal systems of representation and analysis (Stalnacker, 1975; Lewis, 1973 / 1979; Veltman, 2005), and there have been elaborate systems that make computing counterfactuals fully mathematical such as Starr (2010). The instances in the SN data that clearly represent predictions that correspond to counterfactuals, of which we are specifically interested in counterfactual conditionals, since they are by far more likely to represent a would-be causal relation in an entity's assessment. As the irrealis and hypothetical differ from simple proposition in terms of possible worlds with a different set of conditions, and different outcomes for events, clear indications of modality are reliable means for detecting irrealis construction.

7.2.2.1 Common counterfactual expressions

In English, these are normally expressed in the subjunctive or optative (in the sense of such structures that have equivalent modal semantics as those languages that have optative), and any such involving two events e_1, e_2 in a causal relation would require two adjoined propositions. All of them different in terms of the exact modality that each expresses, but generally have pairings of modal expressions in adjacent clauses to indicate a form of irrealis, such as below: :

1. ***if*** *John is walking to school right now, he* ***will*** *be in the rain*
2. ***if*** *John were walking to school right now, he* ***would*** *be in the rain*
3. ***if*** *Mary had studied all night, she* ***would*** *have gotten an A on the exam*
4. ***had*** *Mary studied all night, she* ***would*** *have gotten an A on the exam*
5. ***in case*** *Ian is taking part in the hijack, he* ***will*** *run the risk of shot by snipers*
6. ***should*** *Ian be taking part in the hijack, he* ***would*** *run the risk of shot by snipers*
7. ***in case*** *Jane did not rob the bank, she* ***may*** *not be on the run from the authorities today*
8. ***would*** *Jane not have robbed the bank, she* ***might*** *not be on the run from the authorities today*

For each of the examples (1), (3), (5), and (7) above, the first predication (corresponding to e_1) 's factual occurrence in the real world is not known, but the predication of e_2 is expressed in such a way as to assume the occurrence of e_1. These examples indicates a description of some possible world that may or may not correspond to reality, such as in (1), where the speaker posits a possible universe where John is walking to school, while not knowing if that is the case. In these cases, the possible worlds are potentially, but not necessarily counterfactual, so if they are detected in the data, and let U' be the possible world in question, U be reality, and D be the domain of relevant variables, they at best contribute to counterfactuals with some $P(e_1|U' =_D U)$ with probability of e_2 approximately the same.

For each of the examples (2), (4), (6), and (8), the first predication's factual occurrence is in fact assumed to be at or close to 0, while the predication of e_2 is regardless expressed with the assumption that e_1 occurred. These indicate a description of some possible world that almost certainly does not correspond to reality, such as in (4) where it is assumed that Mary likely had not studied all night. These cases are obligatorily counterfactual, and should contribute to any counterfactual probability of $\approxeq 1.0$ for e_2. Some example such as (7) and (8) adds another layer of complexity, where the probability of $P(e_2|e_1) \not\approxeq 1.0$. So for (7), the contribution to real world probability would be $P^U(e_2) = P(e_2|e_1)P(e_1|U' =_D U)P(U =_D U)$, which can only be used if we have good evidence elsewhere what $P(e_1|U' =_D U)$ is. The contribution to counterfactual probability would be $P(e_2|e_1)P(e_1|U' {\neq}_D U)P(U' {\neq}_D U)$. where neither $P(e_2|e_1)$ nor $P(U' {\neq}_D U)$ is $\approxeq 1.0$; And in (8), we see that even though it is highly likely a counterfactual where ($P(U' {\neq}_D U) \approxeq 1.0$) , we still need to take $P(e_2|e_1) < 1.0$ into account.

7.2.2.2 Logical precedence in counterfactuals

A good example of causal relation that is often mentioned in an SN setting that lends itself to counterfactual assertions and contains copious amounts of causal relations is the set of events surrounding the 2008-2009 global financial crisis. There, a series of events occurring in the financial world depended on a densely connected web of causation, and numerous professional and armature analysts made predictions about what would have happened had certain decisions been made differently at key junctures during the process, which involves counterfactual statements. One of the most appropriate examples would be the causal relation between the meeting between the Federal Reserve and a group of bankers, including potential buyers Bank of America and Barklays (e_1), and the bankruptcy filing of Lehman on the following day with the subsequent consequences (e_2). To most observers, one of the immediately causes of the bankruptcy and later liquidation was the failure of that meeting to produce a viable buyer for the company. For instance, similar statements as below have been said at the time on different SNs:

1. ***if*** *the Fed meeting had produced a buyer, Lehman* ***would*** *not have failed*
2. ***in the case*** *the Fed meeting had succeeded, Lehman* ***would*** *not have to file for bankruptcy*
3. ***had*** *the Fed meeting had produced a buyer, Lehman* ***would*** *not have failed*
4. ***should*** *the Fed meeting had succeeded, Lehman* ***would*** *not have to file for bankruptcy*

The first two examples (and similar statements in SN) make a partial contribution to the counterfactual probability of e_2, makes the same contribution $P(e_2|e_1)P(e_1|U' \neq_D U)P(U' \neq_D U)$ as mentioned before, for those potential counterfactuals for which $P(U' =_D U \ncong 0)$. So these would only fractionally contribute to the final probability, by some weight $\lambda_\partial < 1.0$. The examples (3) and (4) contribute in a straightforward way to the counterfactual probability $P(e_{2_{\bar{e_2}}})$.

7.2.2.3 Adjoined counterfactuals

Not all counterfactuals exhibit the aforementioned explicit cue. Many counterfactuals, especially those that are composed of a pair of adjoined clauses, may not have anything that sets it apart from non-counterfactual or non-causal pairs. Some types of examples may include:

a *John turned on the overhead light; he saw the pair of earrings on the night-stand*

b *John entered the bathroom, opened both of the faucets, and filled the bath tub with warm water*

c *John threw the frisbee across the field, then Mary leapt in the air to catch it; she landed on the grass with frisbee in hand*

d *Mary retrieve the rifle from the case, mounted the scope on top of the weapon, located Patrick among the crowd below, then pulled the trigger*

As illustrated above, the adjunctive candidates sometimes have connectives or adverbials between any pair, (such connectives would not be conditional), while sometimes have a complete absence of such clues. Sometimes their surface form appear simply be two stand alone clauses (2, 4). Other times parallel clauses whose surface forms reside in coordinate structures (1, 3); this is often the case when either both of the *causer/agent* arguments, or both of the *patient/theme/goal* arguments are indexed to be the same real-world entity (such as when John is one that both entered the bathroom and open the faucets).

Like other pairs of adjoined clauses, there are frequently different types of sentential connectives or adverbials that occur between the main sequences of the candidate *protasis* and *apodosis*. Below is a variety of different connectives / adverbials, or the absence of any such, occurring sequentially between a pair that is potentially counterfactual and causal.

i *the Euromaidan revolution established a pro-western government. Russia sent troops to incite a separatist movement in the Crimea*

ii *the Euromaidan revolution established a pro-western government, then Russia sent troops to incite a separatist movement in the Crimea*

iii *the Euromaidan revolution established a pro-western government. a week later, Russia sent troops to incite a separatist movement in the Crimea*

iv *the Euromaidan revolution established a pro-western government. without any warning, Russia sent troops to incite a separatist movement in the Crimea*

Even though these are not logical operators in the way that a subset of conjunctions such as *if*, *then*, *either* are, they still subtly alter the perception of the logical relation between the clauses. For any of the above examples (i) - (iv), if there is another declarative statement such as *"the Euromaidan revolution did not establish a pro-wester government"*, or *"the Euromaidan revolution left the pro-Russian establishment in power in Kiev"*, then the entire construction becomes very likely to be a counterfactual causal. Here the actual occurrence of the real world did not match the claim within the antecedent of the pair, but the claim of the subsequent nevertheless follows from that of the antecedent. Thus by definition this pair is counterfactual, and they are also very likely causal, since there isn't any real world correspondence between these two occurrences, so the connection must be one that is logically sequential.

Similarly, certain connectives corresponding to functions of logical operators have an especially strong effect on selecting which pairs are likely to be counterfactuals, and which ones cannot be at all. We can see it in the following, where a pair in a coordinate structure is followed by a declarative statement negating the would be protasis. While some of these operator connectives bias the reader toward the counterfactual reading, others do not permit such a reading at all:

1. *Many made the earlier conjecture about the nature of Russian involvement in the Crimea. Russia both sent additional troops to protect its base in the Crimea, and incited a separatist movement there. We now have reports that Russia in fact did not send any additional troops there*
2. *Many made the earlier conjecture about the nature of Russian involvement in the Crimea. Russia either sent additional troops to protect its base in the Crimea, or incited a separatist movement there. We now have reports that Russia in fact did not send any additional troops there*

In the above, (1) can easily be interpreted as counterfactual causality, given the operator connective *"and"*. Whereas (2) cannot be interpreted as a counterfactual at all, given the operator connective *"or"*. It is, most likely given the coordinate structure and the negation declarative, to not portray a purely hypothetical scenario, but one that *"Russia incited a separatist movement in the Crimea"* actually occurred, since the would be e_1 *"Russia sent additional troops to protect its bas in the Crimea* as been deselected from this disjunctive logical structure. This phenomenon with operator connectives occurs with all types of adjunctive causalities, but has further interactions in the counterfactual sub-type, when the potential protasis is negated.

7.2.3 Representation of events

Another important factor in formulating a procedure able to discover adjoined causals is an appropriate representation of its component events. When we perform analysis on the results of the module (and the adjunctive causality module), we need to have a good notion of *event* that can be derived from the frame-structures. To say that $Event_x \xrightarrow{caus} Event_y$ first requires that there be a consistent definition for both $Event_x, Event_y$, having a constant level of specificity, in order to have a consistent definition of relations among events.

It is useful to differentiate the concepts of *event type* and an *event token.* Ultimately, a precise definition of causality refer to relations among event tokens. (Scheffler, 1992) We are only able to speak of causation between some actual events e_i, e_j with real-world participants in real space and time. It cannot be assumed that some other events e'_i, e'_j will have the same causal relation, even if they share event types $\tau(e_i), \tau(e_j)$ with e_i, e_j. If e_i is John setting up a trip wire, and e_j is Mary falling due to that trip wire, it is not true that $\forall e'_i \hat{\in} \tau(e_i), e'_j \hat{\in} \tau(e_j), [e'_i \xrightarrow{caus} e'_j]$, such as when Patrick sets a trip wire in Dublin during the Renaissance, while Quintessa trips up in New York in 2012. So when we consider of causal relation between two event types, the only precise way is based on inference using some instances of causal relations of event tokens.

This entails that such a precise computation entails event tokens defined by the predicate with its arguments and the relevant adjuncts to preserve its specificity. Omitting some essential component entails making some generalization over causal relations. While we can compute a probability of two classes of events being causal, such as the following, where $\tau(e_i)$ is the predicate type of e_i:

$$P(\tau_i \xrightarrow{caus} \tau_j) \hat{=} \frac{\displaystyle\sum_{\langle e_i, e_j \rangle \in \mathcal{E} \times \mathcal{E}} 1\Big(\tau(e_i) = \tau_i, \tau(e_j) = \tau_j, e_i \xrightarrow{caus} e_j\Big)}{\Big|\Big\{\langle e_i, e_j \rangle \in \mathcal{E} \times \mathcal{E}, \tau(e_i) = \tau_i, \tau(e_j) = \tau_j\Big\}\Big|} \tag{7.1}$$

By the same token, there is no a priori reason to disregard the causal probability of any two event tokens involving two predetermined entities, such as the following, where $\nu(e_i)$ gives the external argument of the predicate of e_i:

$$P(n_s \xrightarrow{caus} n_t) \hat{=} \frac{\displaystyle\sum_{\langle e_i, e_j \rangle \in \mathcal{E} \times \mathcal{E}} 1\Big(\tau(e_i) = \nu(e_i), \nu(e_j) = \tau_j, e_i \xrightarrow{caus} e_j\Big)}{\Big|\Big\{\langle e_i, e_j \rangle \in \mathcal{E} \times \mathcal{E}, \nu(e_i) = n_s, \nu(e_j) = n_t\Big\}\Big|} \tag{7.2}$$

While the *event token* is the most precise specification, computing causal metric with this definition is unrealistic for most datasets. Combinatorial possibilities given the components of both e_1 and e_2 are too numerous to arrive at a reasonable sample size for any *event token.* Thus it is necessary to prune the components of e_1, e_2, and possibly adopt a fuzzier notion of equivalence between events.

The aforementioned conception of semantic frame (Davidson, 1967, Parsons 1985 / 1990) affords a good foundation for a useful definition for dealing with this combinatorial issue. The use of $\phi(A_0,\ A_1,\ ...\ A_k, e_i) \wedge X_1(e_i) \wedge X_2(e_i) \wedge\ \cdots$ as representations strikes a middle ground in terms of the number of combinatorial possibilities between that of single *predicate*, which is generalized to the exclusion of any entity information; and that of full list of arguments, which is over-specified to the point, such that for any frame, another frame of the same identity is unlikely to be observed in the corpus of reasonable size.

Hence this representation of *event type* and *event token* follows the previously extracted *semantic-frames* during Section 5 in terms of structure and complexity. It follows frames' essential and non-essential components as defined by the extraction mechanism, which is informed by the adicity structure of each predicate whenever available. In the implementation itself, the *event token* is simply implemented as a wrapper layer on top of the frame structure, and the *event type* is simply those tokens that are deemed to be identical with respect to relevant components. The definition for *event type*, in fact, is variable according to the need of the analysis; it could be defined, for example, as those event tokens with identical *predicate + external argument* of the frame. This flexibility will be crucial for a future phase, when the information of events and relations are used to construct or supplement graph theoretic structure of social networks of entities from the corpus. Using this system preserve sufficient information about the principle entities involved in the event, and allow sufficient generalization so that providing large enough sample size become much less problematic.

7.3 Adjoined causal structure extraction and ranking

The ranking procedure described here is an unsupervised procedure with no labeled data input. We could have potentially begun with annotated adjunctive causal examples in formulating our procedure. In that case, the learning and ranking procedures would be similar to that of unsupervised method described below, but the automatic extraction of counterfactuals would be omitted from the procedure and replaced with a small annotated training set. It is preferable to utilize a completely automated method to extract the training samples if feasible; this would lower the reliability of the training data significantly, but also would make this entire procedure much easier to perform, and much less time-consuming to replicate for different datasets in the future.

The procedure has its goal in examining all possible adjacent clause pairs as defined by the syntactic parsing and semantic frame extraction steps. We aim to have an unsupervised methodology in locating the most likely pairs. Unlike *embedded causality*, there is no guarantee that a hierarchical structure that encompasses the *protasis* and the *apodosis* of the causality, so we need to rely on patterns in the linearized sequences. We have seen that *counterfactual causality* offers one of the best guarantees that the underlying structure is causal. Given that some counterfactuals can be verified automatically in the text (as we will see next), these verified samples can be used as training samples to learn patterns in causal surface sequences.

The procedure uses this automatically extracted, relatively small sample of causality (verified counterfactuals), and then provides a composite representation of such sequences in **hidden markov model**s (HMM here on). With sets of HMMs trained on these extracted data, the procedure takes into account of a number of modifications and extensions of the basic HMM algorithms to accommodate for the data type and our goal. The end result is an integrated procedure that is capable of ranking all possible adjacent pairs with respect to their likelihood of being a causal pair that contains a *protasis* and an *apodosis*.

7.3.1 Extraction of counterfactuals

As we have seen, causality in language bears logical attributes that can be expressed in axioms described using predicate logic. We can utilize the natural outcomes of this property in language to extract the likely counterfactual forms. Given that counterfactuals of the form $e_i \xrightarrow{\rho} e_{i+1}$, by definition, are event sequences that have not taken place, so the relation ρ that is observed in the text cannot be attributed to any real world association. Since the relation cannot be attributed to some incidental temporal/spatial juxtaposition of the pair of events in the reference world (here referring to the world that the speaker/writer is referring to) any ρ that is postulated for $e_i \rightarrow e_{i+1}$ is likely attributed to some sequential relation between the events with some logical connection. When we encounter counterfactual expressions occurring in an adjoined form, which is similar to all adjoined expressions (expressions without explicit logical connectives or lexico-syntactic cues of logical sequencing), a similar surface form must also be able to express some logical connection between the event types, when it is not counterfactual. Thus we can reasonably see that, if the hypothesis holds that the linear sequencing outside of the core components of the frames of e_i, e_{i+1} has a discrimination effect on determining whether there is a logical connection between the two, then using adjoined counterfactual forms should allow us to train for all adjoined pairs (not necessarily counterfactual) with that logical connection; this includes using adjoined counterfactual causals.

7.3.1.1 Axiomatic property of counterfactual

As we have seen in Section 7.2.2.3, extraction of counterfactual among adjoined pairs has a special tendency. The general form of causality has no additional requirements in its context; even in its most strict definition (Cartwright, 1979), $\mathcal{U}(\tau_1, \tau_2) \vdash \Big[\forall e_i[e_i \in \mathcal{E}, \overline{\tau_1(e_i)}, \overline{\tau_2(e_i)}] \longrightarrow [\tau_2|\tau_1 \wedge e_i] \wedge [\tau_2|\tau_1 \wedge \bar{e_i}]\Big]$, ($\mathcal{U}$ represents a causal relation between event types, and $\overline{\tau_j(e)}$ represents the event e is not of the type τ_j). The representation above basically says that for any event types pair $\langle\tau_1, \tau_2\rangle$, a causality $\mathcal{U}(\tau_1, \tau_2)$ is present entails that for all events in the event-set considered, $\mathcal{E}$, both the presence of that non-τ_1, non-τ_2 event e_i does not affect the logical consequence $\tau_1 \rightarrow \tau_2$. Also take not that an enumeration of all events $e_i \in \mathcal{E}$ must be considered, and no individual event has any unique effect. Counterfactual causality entails an additional requirement on the *protasis* of the logical form, where it must be the case at some point in time that $\mathcal{U}_{\mathcal{CF}}(\tau_1, \tau_2) \vdash \Big[\exists e_i \in \mathcal{E}[\tau(e_i) = \tau_1] \wedge [\bar{\tau_2}|\bar{e_i} \wedge \bar{e_i}]\Big]$.

This implies that if something is meant to be expressed as unambiguously counterfactual, there should be some declarative expression that approximates the negation of the meaning of the *protasis* of the causality. Those declarative forms that have the *negated* form of a semantic frame of some protasis candidate, with the same predicate and essential argument list, and in close linear proximity to the candidate pair, would be considered a very likely indicate of counterfactual causality. An pair such as *"John turned on the stove in the kitchen* *before dinner. Then the water in the pot boiled"*, along with *"John did not turn on the stove"* within a few clauses, would likely indicate a counterfactual.

The adjoined pair alone seems like a pair of factual assertions indicating actual occurrence of the events, but the negated protasis makes it likely that this pair does not indicate factual occurrence, and these pairs are often nestled within descriptions of thoughts, desires, assumptions, quotations, etc. Compared to the types of counterfactual causals that are indicated by lexico-syntactic cues, this type is less certain and more prone to noise. The advantage here lies in the fact that the candidate pairs are of the identical types in terms of their surface sequence as the general adjunctive causal constructions. If we use this type of potential counterfactual as training samples, and our training procedures can tolerate certain amount of false positives, then it is far more suitable for discovery of patterns in adjunctive causal structures.

7.3.1.2 Extraction procedure

The candidate pairs examined in this part of the study are pairs of clauses $\langle i, i+1\rangle$ in the corpus. While some canonical counterfactuals such as the form *"should SPEC-ELEM ... CORE-VP ... , SPEC-ELEM would CORE-VP ... " "in the case SPEC-ELEM ... CORE-VP ... , SPEC-ELEM will CORE-VP "* can be extracted with relative ease, these forms of counterfactuals, in terms of their surface sequences, are quite unique to the counterfactual logical form, and cannot be easily extended to surface sequences of other types of causal structures. These forms do not correspond well to the structures of general forms of adjunctive causals in the corpus.

So we utilize a type of contextual property that is dictated by the logical implications of counterfactuals. For each of these pairs, the first is the assumed candidate protasis e_i, while the second assumed candidate apodosis e_{i+1}. We examine all other declarative clauses within a reasonable context (could be the entire document, in practice the maximum distance is determined by computational requirements) to find potential contradiction with e_1. The contradiction score between each possible $\langle e_i,\ e_j | i \neq j \neq i+1\rangle$ is determined by three component factors, and the product of the three provides the contradiction potential of the

e_i, e_j pair.

7.3.1.3 Counterfactual scoring

Each of these is based on the semantic frame structures that have been extracted for each of the surface clauses. One component is the sign metric between these two frames, where the negation component(s) for each of these frames e_i, e_j must have opposite signs from each other to be 1, and 0 if they agree. A second component is the distance metric of the pair $\langle i, j\rangle$; this is done using technique similar to inverse rank, where the this component decays proportional to $\frac{1.0}{|i-j|}$. The third component is a similarity measure between the individual components of the semantic frames e_i, e_j; this is a metric which is a product of the similarity between the predicates predicates as well as the similarity between the argument sets. The argument sets are measured by similarities of the essential arguments, and the permutation of their order between the sets. The similarity between the pair of predicates $PRED^{e_i}, PRED^{e_j}$ or between some pair of individual arguments $ARG_m^{e_i} \in ARGLIST^{e_i}$, $ARG_n^{e_j} \in ARGLIST^{e_j}$ is computed using either PMI, a WordNet similarity measure (explained in an earlier section), or the product of both. These augmentations allow for some gradation and flexibility in the way we determine whether a predicate or argument is similar to another in a different frame.

7.3.1.3.1 Mutual information modifications to the similarity function Part of the augmentation of the similarity measure between predicates and arguments is to observe their occurrence on a per clause basis in the corpus. A frequently employed information theoretic way of computing contextual association of tokens is a simple metric, the *pointwise mutual information* (PMI). Here, we simply define co-occurrence of two token types as appearing in the same clause, and thus simply treat the set of terminals in each clause as a bag in context, with $\mathcal{X}$ as the set of bags of co-occurrences; and the set $\mathcal{W}$ as the set of all types of tokens that occur in the corpus. We can define the PMI $\mathfrak{I}(w_i, w_j)$ in our case as:

$$\begin{cases} \Im(w_i, w_j | w_i, w_j \in \mathcal{W}) \ = \\ \qquad log \ \frac{P(w_i, w_j \in X_k | X_k \in \mathcal{X})}{P(w_i \in X_k | X_k \in \mathcal{X}) \cdot P(w_j \in X_k | X_k \in \mathcal{X})} \end{cases} \tag{7.3}$$

This allows us to measure, for the pair $\langle w_i, w_j \rangle$, the discrepancy between their co-occurrence by chance in the corpus and their actual co-occurrence observed in individual clauses.

7.3.1.3.2 Lexico-semantic modifications to the similarity function The similarity measure between the target frame and any potentially counterfactual frame can be further modified based on lexico-semantic means to be a more accurate metric for tree edit. The second cost function η will be based on the relational distance in a structure knowledge resource of lexical entries, WordNet here is one of the logical choices. The tokens $\in \Sigma_T$ need to be lemmatized into their lexical forms. It does not make sense to use WordNet for certain classes of lexical items, especially those of the closed class such as modals, auxiliaries, complementizers, pronouns, and prepositions; for any such class the measurement used is pure identity. For the content verbs and nominals, it is sensible to use lexico-semantic resources to provide a better substitution cost.

Since WordNet is primarily built on *hypernymy* and *hyponymy* relation, there exists a robust IS-A backbone for classification of lexical entries, especially for entities and events. Certain classes of lexical entries among entities and events are likely to occur in the same types of intermediate structures in these embedded multi-clausal causatives, such as *"... **arranged** the **affairs** such that ... "*, *"**aligned** the **circumstances** such that ..."*. Thus, finding the highest level categories in WordNet whose hyponyms are all likely to occur in these causal structures would provide a much more robust scoring system for substitution in the tree edit algorithm.

Least common ancestor (LCA) problem is a well investigated area of graph theory with well understood mechanics (Harel & Tarjan, 1984; Dietz, 1991; Berman & Vishkin, 1994; Bender & Farach-Colton, 2002/2004; Moufatich, 2008; Ben-Amram, 2009; among others). In many of the works in this area, the concept of Euler tour of trees (Tarjan & Vishkin, 1984) is used to facilitate the discovery of the the LCA; while our goal is substantially different, we can utilize a similar approach to obtain a set of optimized ancestors that is reasonably small to afford a good amount of generalization, but at the same time remains representative of the lexical entries. In order to find that, we employ a path intersection algorithm by treating the WordNet structure as a graph. Assume that T^w be a tree structure consists of a top-level node such as 'entity' or 'event', and all of its IS-A descendents. Using T^w, we turn this into a graph traversing problem.

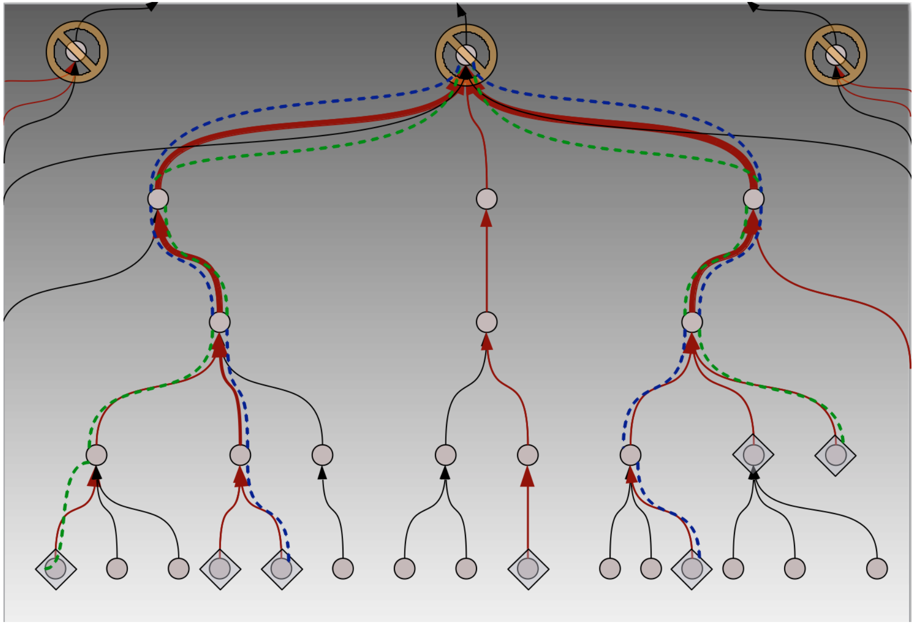
For each class of open class content lemmas found in the intervening parts of embedded multi-clausal causatives (e.g. nominals, content verbs), we find the corresponding tree T^w rooted at the corresponding top level lexical semantic category (entity, event). Let the V^w be the set of corresponding nodes $\in V(T^w)$. The method to obtain the candidates for the set of optimal ancestors $\mathcal{Y}$ follows these simple steps of the algorithm ($p_w(v_i)$ again is the parent of v_i in the topology of the tree T^w, and $\bar{\mathcal{A}}$ is the set of generalization caps for any potential ancestry):

$$\begin{cases} \mathcal{P} = \Big\{ P = \langle v_h,\ \dots\ ,\ v_i, \dots,\ v_j \rangle \Big| v_h, v_j \in V(T^w) \wedge \\ \qquad \forall v_i \in V(P)\ [N^-(v_i) \cap \bar{\mathcal{A}} = \varnothing] \Big\} \\ \mathcal{X} = \Big\{ x_l \in V(P_m) \cap V(P_n) \ \Big|\ P_m, P_n \in \mathcal{P}, P_m \neq P_n \Big\} \\ \mathbb{N}^{\mathcal{X}} \begin{cases} \forall x_l \in V(T^w) \setminus \mathcal{X}, \quad \mathbb{N}^{\mathcal{X}}_{x_l} = 0 \\ \forall x_l \in \mathcal{X}, \quad \mathbb{N}^{\mathcal{X}}_{x_l} = \Big| \Big\{ P_m | P_m \in \mathcal{P} \wedge x_l \in V(P_m) \Big\} \Big| \end{cases} \\ \mathcal{Y} = \ \Big\{ y_k \ \Big|\ y_k \in \mathcal{X}, \mathbb{N}^{\mathcal{X}}_{p_w(y_k)} \leq \mathbb{N}^{X}_{y_k} \Big\} \end{cases} \tag{7.4}$$

The algorithm basically starts with a subset of nodes in T^w that are detected as lexical items of a particular POS in the intervening structure, and finds the set of all paths between them $\mathcal{P}$, each of which should be unique due to the graph property of a tree. All the nodes $\in V(T^w)$ are tested to see whether any path $\in \mathcal{P}$ passes through it, if so then it is $\in \mathcal{X}$.

The number of paths that passes through each node $x_l \in \mathcal{X}$ is $\mathbb{N}^{\mathcal{X}}_{x_l}$, which is the *'traffic'* detected at each node. If the traffic of any node is larger than its parent, then it is placed into the optimized set of ancestors $\mathcal{Y}$, which then can be used as a form of generalization that we can use in further processing.

Figure 7.1: Here, once we have caps on the generalization that we want, and the target nodes in diamonds, we can discover paths between pairs of relevant nodes, two of which are illustrated in green and blue

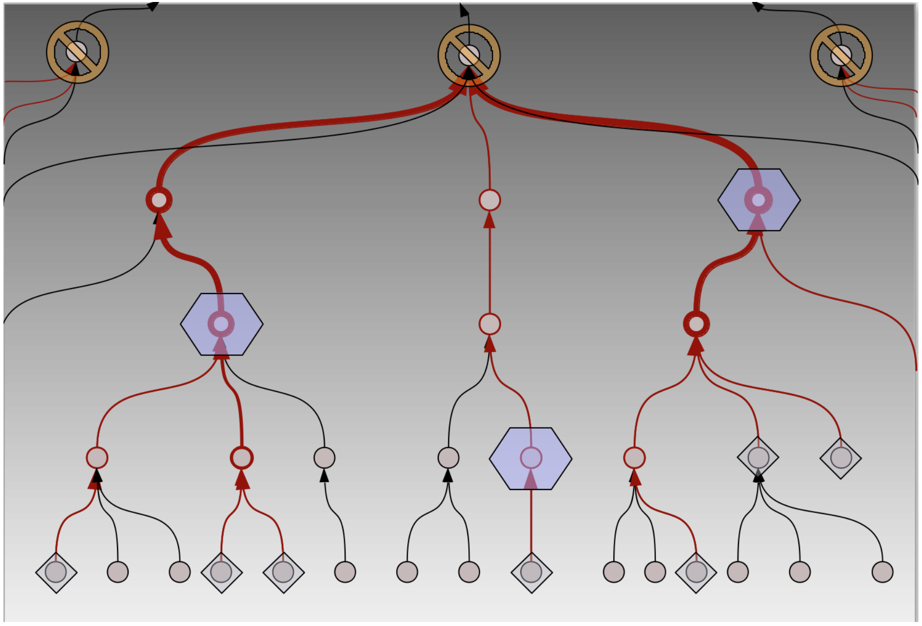


The forms of generalization would be mediated through the set of LCA, with pair of its descendents having some measure of similarity that can be taken into account when comparing the original frame with a potentially counterfactual frame to it. This similarity can also be graded, that is proportional to the length of $P_{i,j} \in \mathcal{P} | v_i, v_j$ are the WN node in the target frame, and the WN node in the potential counterfactual frame. Any pair $\langle v_i, v_j \rangle$ that cannot be reach with such as $P_{i,j} \in \mathcal{P}$ can be ruled as dissimilar and assessed a much larger penalty when determining the similarities between argument lists.

7.3.2 Adjoined sequences and HMM

The procedure to measure and rank the potential for each adjacent pair of frames to be causal is designed so that it relies almost exclusively on the information that are contained within those two context free trees of the candidate *protasis* and *apodosis* structures, outside any frame essential information. So this excludes both any contextual information outside of the structures containing the two frames; and also excludes information from the *predicate*

Figure 7.2: There after, we proceed to find the likely set of LCAs (in hexagonal) as high trafficked nodes with no ancestor (below the generalization cap that are diagonally stricken) that is higher trafficked



and essential list of arguments from each candidate frame. Both contextual information, as well as frame essential information, are widely used to rank such causal candidate pairs, but the aim of this study is to develop a method that is separate from, and complementary to the traditional parametric machine learning, or purely logical frame based methods.

7.3.2.1 Surface sequence of adjoined causality

Since in *adjunctive causal structures*, the majority of the potential pairing of frames have no shared hierarchy in syntax, that is they usually reside in completely separate but consecutive clauses (unlike in embedded causal structures), it is sensible to use model that is able to represent linear sequence at the surface well. Since in this part of the study we aim to discover causal relation without regard to any hierarchical structure, the candidate semantic frames corresponding to each $\langle \mathit{protasis}, \mathit{apodosis} \rangle$ pair are semantic frames that are in linear sequence, disregarding whether they reside in any larger hierarchical structure. (This also means that there may be some overlap in the results of this procedure and that of the embedded-causal procedure.)

HMM is a very widely used model to represent such sequential patterns. HMM is also well adapted for this purpose because we seek to identify a hidden layer of sequence that has some correspondence to the surface linear structure, the hidden layer has at the minimum a distinct structure of

$$INTERSEQ_p\ PROTASIS\ INTERSEQ_q\ APODOSIS\ INTERSEQ_r$$

where the transitions among the set of hidden states occur at the transitions between any pair of subsequence, (1) not having seen either protasis or apodosis, (2) already seen protasis but not having seen apodosis, and (3) having seen both protasis and apodosis.

In practice, we have found that some of the sequential information within the protasis and apodosis sub-sequences also may be useful in determining the appropriate HMM representation, in particular the components that reside in between the SPEC-ELEM and the CORE-VP containing the main content VERB of each clause. The SPEC-ELEM is the head-constituent residing in the external position of the matrix clause, and the CORE-VP is the VP of the clause that the main content verb is heading. The SPEC-ELEM and the CORE-VP themselves, since they are present in most adjunctive causal pairs, can essentially be treated as variables, and these constituents will be assigned a single emission symbol in the sequence. Thus the model of sequence that concerns us is $Q_0\ SPEC\_ELEM_{prot}\ Q_1\ CORE\_VP_{prot}\ Q_2\ SPEC\_ELEM_{apod}\ Q_3\ CORE\_VP_{apod}\ Q_4$, where each $Q_.$ is the intervening elements in the linear sequence.

7.3.2.2 Standard HMM algorithms

The HMM is a parametric description of a Markov process, and is expressed as a 5-tuple:

$$\theta = \left\{\begin{array}{l} S = \{s_1, ..., s_n\} \\ \Sigma = \{\sigma_a, ..., \sigma_z\} \\ \Pi = \{\pi^{s_\varkappa} \mid s_\varkappa \in S\} \\ A = \Big\{a_{\zeta,\eta} = \\ \qquad P(s_\eta, t \mid s_\zeta, t-1) \Big| \\ \qquad 1 \leq \zeta \leq n, 1 \leq \eta \leq n, \\ \qquad 1 \leq t \leq |O|\Big\} \\ B = \{b_{\iota,\kappa} = P(o_\kappa \mid s_\iota) \Big| \\ \qquad 1 \leq \iota \leq n, \sigma_\kappa \in \Sigma\} \end{array}\right. \Bigg| \left\{\begin{array}{l} \displaystyle \forall\zeta \sum_{\eta=1}^{n} a_{\zeta,\eta} = 1 \\ \displaystyle \forall\iota \sum_{\sigma_\kappa}^{\sigma_\kappa \in \Sigma} b_{\iota,\kappa} = 1 \\ \displaystyle \sum_{\zeta=1}^{n} \sum_{\eta=1}^{n} a_{\zeta,\eta} b_{\iota,\eta} = 1 \\ \displaystyle \sum^{s_\varkappa \in S} \pi_\varkappa = 1 \end{array}\right. \tag{7.5}$$

The basic HMM algorithms used in this part of the study includes the training of the Π, A, B components using a form of Expectation-n (Baum-Welch algorithm) (Baum & Petrie, 1966; Baum & Eagon, 1967; Baum et. al. 1970, Welch, 2003) based on sequences of observations in the training data, the discovery of the most likely observation sequence(s) O given a trained set of parameters θ (Forward algorithm), and the discovery of the most likely sequence of hidden states given θ and O (Viterbi algorithm).

Given that a baseline Baum-Welch trains on one sequence of emission symbols, in order to utilize the entire corpus to parameterize a single HMM, we need to treat the entire corpus as a single sequence, which is clearly infeasible considering the size of the corpus and the and the way the algorithm treat the α and β (forward & backward) components of algorithm. So in order to take the whole dataset into account, we need to train multiple HMMs, or use some modified form of Baum-Welch. We will use a combination of these strategies to resolve this issue, to be discussed in Sections 7.3.3.2 and 7.3.3.6 .

For the observation sequences, not all syntactic POS need to be treated equally, some classes are much more likely to have an impact on the model's ability to discriminate when treated as token-types, while others can be treated as general classes. The syntactic constituents (outside of any extraposed parts) of the SPEC-NP and CORE-VP, can be treated as singular components. Certain other constituents, such as nominal entities, or adjectivals that modify specific heads of NP, can be treated as that class, so we can roughly treat them as their syntactic tags in the surface sequence.

7.3.2.3 A concise description of Baum-Welch

The standard Baum-Welch algorithm, used to re-estimate the each component of θ for each iteration. The forward and backward procedures α and β are described as, given that $o_{[t]}$ is the symbol in the observation sequence, and N is the number of states in the machine:

$$\begin{cases} \alpha_i(t) = \pi^i b_i(o_{[t]}) \ |t = 0 \\ \alpha_j(t) = b_j(o_{[t]}) \displaystyle\sum_{i=1}^{N} \alpha_i(t-1) \ a_{i\to j} \ |t \geq 1 \end{cases} \tag{7.6}$$

$$\begin{cases} \beta_i(t) = 1 \ |t = T \\ \beta_i(t-1) = \displaystyle\sum_{j=1}^{N} \beta_j(t) \ a_{i\to j} \ b_j(o_{[t]}) \ |t \leq T \end{cases} \tag{7.7}$$

With these defined, the auxiliary parameters γ and ξ then can be estimated as the following:

$$\gamma_i(t) = \frac{\alpha_i(t) \ \beta_i(t)}{\displaystyle\sum_{j=1}^{N} \alpha_j(t) \ \beta_j^t} \tag{7.8}$$

$$\xi_{i\to j}(t) = \frac{\alpha_i(t) \ a_{i\to j} \ b_j(o_{[t+1]}) \ \beta_j(t+1)}{\displaystyle\sum_{h=1}^{N} \alpha_h(t) \ \beta_h(t)} \tag{7.9}$$

And the estimation of the Π, A and B components of the following iteration, which we denote the re-estimated parameters within the next set of parameters for the machine $\hat{\theta}$ with the $\hat{a}_{i\to j}$ and $\hat{b}_j(t)$, and $\varsigma(s)$ produces the symbol σ_s in the position s in the set σ, while $\imath(\cdot)$ returns 1.0 if the statement within is true, and 0.0 otherwise:

$$\hat{\pi}_i = \gamma_i(0) \tag{7.10}$$

$$\hat{a}_{i \to j} = \frac{\displaystyle\sum_{t=0}^{T-2} \xi_{i \to j}(t)}{\displaystyle\sum_{t=0}^{T-2} \gamma_i(t)} \tag{7.11}$$

$$\hat{b}_j(s) = \frac{\displaystyle\sum_{t=0}^{T-1} \imath(o_{[t]} = \varsigma(s))\ \gamma_i(t)}{\displaystyle\sum_{t=0}^{T-1} \gamma_i(t)} \tag{7.12}$$

7.3.3 HMM modifications

This specific problem requires several significant modifications to the basic HMM model and algorithms. The training of data is done through the E-M algorithm Baum-Welch, and each potential surface sequence that contains a potential $\langle$ *protasis*, *apodosis* $\rangle$ pair is treated as an individual observation sequence. These are necessitated by the nature of the data type, computational complexity, as well as the integration of additional source of information into the model, explained in detail below.

7.3.3.1 Emission backoff

For each HMM trained using Baum-Welch, in order to cut down on the size of each distribution, we for each emission distribution in θ_m, trained on the $l-$th subset of observation sequences, we do not compute emission of all possible observation symbols, but only those symbols that are present in the subset of observations that are used to train θ_m. Given the way O is partitioned (to be discussed next) into each training subset O_m, probability is high that some testing observation sequence, on which we must run Viterbi algorithm, would contain some emission symbol not present in the original O_m

It is possible to adopt a simple smoothing technique, and treat all observation tokens unseen in the training data as equivalent to *hapax legomenon*; this results in a large number of ineffectual emission probabilities. There is a natural way for the backoff mechanism to have an intermediate level of specificity, given that the surface sequence here essentially has two levels of observational symbols of $\langle$ *token*, *tag*$\rangle$. This allows us to employ the backoff of $token \longrightarrow tag \longrightarrow hapax\_legomenon$. This entails that we must also provide a set of emission distribution for tag data for each θ_m, trained in the same way as the token sequence.

7.3.3.2 Multi-observation HMM

As earlier mentioned, a valid strategy would be to proliferate the number of HMMs with the number of training sequences; but this would make the training of parameter sets and the reintegration of information from HMM algorithms on testing sequences computationally much more expensive. Using modified Baum-Welch to train an HMM on multiple emission sequences presents a tradeoff. Training some HMM based on the collection of all sequences regardless of their length has the advantage of greater power in statistical inference, and mitigates much of the risk of over-training. . On the other hand, training each sub-set corresponding to some subset of training sequences of the same length allows us to make finer distinctions among different types of training sequences, as defined in Section 7.3.3.6. We attempted both methods, and found that there is not a large difference in the way it affects subsequence induction (discussed later), and that training aggregate sets without regard to lengths reduces time-complexity in computation, thus allowing us to work with larger testing set, leads us to aggregate the training sequences disregarding length.

In each iteration of the maximization step, we need to recompute Π, A, B components of θ from α, β of each of the member sequences of the set. There are a variety of methods to weigh the contribution from each, such as simple mean, randomized, windsorized method, etc. We chose the weighting method devised by Rabiner et al (1993), which uses $P^{fwd}(o_k, \theta)$ to inform o_k's contribution to $\hat{\theta}$. The idea in the Rabiner method for Baum-Welch is to equilibrate the contributions at each step to make the final set of parameters as much as possible to reflect all sequences in the set. And $P^{fwd}(o_k, \theta)$ is estimated by running the forwards algorithm with θ, o_k as inputs. The Rabiner method was shown to work very well, except in the filtering out of irrelevant sequences to a prototype, due to its effort in attempting to equalize the impact of all member sequences of subset at each step; this issue will be dealt with during scoring, by a mechanism in Section 7.3.3.5.

Given that there are total K observation sequences that we are training a specific HMM on, with $o_k | 1 \leq k \leq K$ being the k^{th} sequence in the set O_m, N is the number of states for this machine, and T^k being the linear size of that particular emission sequence, we can define the parameter $\hat{A}$, which is the updated set of transition probability distributions, with A being the current set:

$$\hat{a}_{i \to j} = \frac{\sum\limits_{k=0}^{K} \frac{1}{P^{fwd}(o_k)} \sum\limits_{t=0}^{T^k-2} \xi_{i \to j}(t)}{\sum\limits_{k=0}^{K} \frac{1}{P^{fwd}(o_k)} \sum\limits_{t=0}^{T^k-2} \gamma_i(t)} \Bigg| \, o_k \in O_m \tag{7.13}$$

And the following is the multi-observation sequence using Rabiner method for $\hat{B}$, the next set of emission distributions.

$$\hat{b}_j(s) = \frac{\sum\limits_{k=0}^{K} \frac{1}{P^{fwd}(o_k)} \sum\limits_{t=0}^{T^k-1} \mathbb{1}(o_{[t]} = \varsigma(s)) \; \gamma_i(t)}{\sum\limits_{k=0}^{K} \frac{1}{P^{fwd}(o_k)} \sum\limits_{t=0}^{T^k-1} \gamma_i(t)} \Bigg| \, o_k \in O_m \tag{7.14}$$

Each multi-observation HMM will be trained with the subset O_m as input sequences; the result θ_m is a single HMM to be used according to the description in Section 7.3.3.6 .

7.3.3.3 Numerical stability

Another issue that is frequently encountered with training HMMs with emission sequences that are sufficiently long, and number of possible emission symbols sufficiently large is numerical stability, when computing α, β, and the auxiliary variables. Some of the conditional probabilities computed can have an extremely small value, and lead to float-point over/under-flow, and the presence of *'nan'* values in the computation, which must be dealt with inelegantly. With the default implementation of Baum-Welch, the sequence set extracted causes a large proportion of all computations to result in the underflow of values computed for A and B of each θ_m.

Rescaling the probabilities each iteration can mitigate some of the problems, but we opted for the more robust solution of reformulating the Baum-Welch and Viterbi parts in log-arithmetic (Mann, 2006). This eliminates all but the most extreme forms (values that scale hyper-exponentially with length of sequence) of numerical instability, and is generally sufficient for our purposes. We will use the symbol '$\mathfrak{nan}$' to represent the potential domain error resulting from $ln(0)$, and define a modified natural logarithm function $\widetilde{ln}(\cdot)$:

$$\widetilde{ln} := \begin{cases} log_e(x) \mid x > 0.0 \\ \mathfrak{nan} \mid x = 0.0 \end{cases} \tag{7.15}$$

The corresponding summation and product functions to be used in log-arithmetic version of the parameter estimation functions, $\widetilde{+}$, $\widetilde{\times}$ then can be reformulated as the following:

$$\widetilde{ln}(x) \mathbin{\widetilde{+}} \widetilde{ln}(y) := \begin{cases} \widetilde{ln}(x+y) \mid x > 0.0 \wedge y > 0.0 \\ \widetilde{ln}(y) \quad \mid x = 0.0 \\ \widetilde{ln}(x) \quad \mid y = 0.0 \end{cases} \tag{7.16}$$

$$\widetilde{ln}(x) \mathbin{\widetilde{\times}} \widetilde{ln}(y) := \begin{cases} \widetilde{ln}(x) + \widetilde{ln}(y) \mid x > 0.0 \wedge y > 0.0 \\ \mathfrak{nan} \quad \mid x = 0.0 \vee y = 0.0 \end{cases} \tag{7.17}$$

With the utilization of logarithmic arithmetic in place, individual sequences are able to scale to much longer lengths, without the concern of returning some invalid θ_m.

7.3.3.4 Adjunctive causality scoring mechanisms

The set of HMMs trained then would be used to score each adjunctive causal candidate pair. Each individual HMM contributes to the scoring of the pairs by the application of both Forward and Viterbi algorithms. Scoring by Forward algorithm is natural, given the the purpose of Forward algorithm is to determine the probability of some training sequence of emission symbols o_k given some set of parameters θ_m of an HMM. This scoring is reflective of the how well each testing sequence conforms to θ_m in a majority of the circumstances.

Some scenarios do prove a challenge, such as when most of the likely hidden sequence has high transition probabilities, and the corresponding observation sequence has high emission probability from each hidden state; but a single ultra-low transition probability $A^l_{i,j}$ required for the necessary hidden state, or a single ultra-low emission probability $B^l_{j,\sigma}$ required for a specific symbol in the sequence, can bring the scoring of the entire sequence to a much lower level. This scenario will result in low score for the testing sequence even though the majority of its required transitions and emissions have high probability in A^l and B^l. Sometimes the requirement of some ultra-low probability transition is ameliorated by mechanism in Section 7.3.3.6, when the this type of transition occurs between natural constituents in the CF-structures; but it still presents a problem when such transition is within the terminal sequence of a sufficiently small constituent.

This can be circumvented by using Viterbi algorithm to produce the most probable ordered set of hidden states given the entire set of sequences $\in O_m$, and then compare the sequence(s) of hidden states to the expected sequence of $Q_0 \longrightarrow Q_1 \longrightarrow Q_2 \longrightarrow Q_3 \longrightarrow Q_4$ by edit distance. The cost matrix for the edit distance matrix is designed to make the cost of deletion of a state from the end of a sequence less costly, such as given the original sequence $q_1 \rightarrow q_2 \rightarrow q_3 \rightarrow q_4$, it would be less costly to edit it to become $q_2 \rightarrow q_3 \rightarrow q_4$, or $q_1 \rightarrow q_2 \rightarrow q_3$, than it is to become $q_1 \rightarrow q_3 \rightarrow q_4$, or $q_1 \rightarrow q_2 \rightarrow q_4$. This design is to make $S^{vit}(o_k, \theta_m)$ more favorable toward likely sub-constituents examined during the subsequence induction in 7.3.3.6 later, since the subsequence induction mechanism tends to form truncated but contiguous parts of the observation sequences.

The better conforming sequences should produce hidden state sequences using Viterbi algorithm is, the more similar the sequence is to a Bakis model of the canonical hidden state sequence (a model where the transitions from state $q_i \rightarrow (q_i|q_{i+1})$ is highly favored). We represent the Forward derived score as $S^{fwd}(o_k, \theta_m)$, and the Viterbi-derived score as

$S^{vit}(o_k, \theta_m)$, these will continue to be modified through the following mechanisms. The scoring from the two methods then need to be balanced, by scaling their respective means to an equivalent value, and then the composite score for each o_k, θ_m pair is computed additively.

7.3.3.5 Sequence similarity and obliquity

We consider that a training sequence o_k as having low likelihood of producing θ of an HMM with respect to a testing sequence o_l, if o_k and o_l share very few emission symbols in common. Given some $O_m \subset O$ where O_m is used to train θ_m; if there is no $o_k \in O_m$ such that o_k shares significant number of emission symbols with the testing sequence o_l, then any high score $P^{fwd}(o_l, \theta_m)$ is likely to be incidental, and should be discounted. So, this auxiliary weighting system for the contributions of each HMM_m on the scoring of the testing o_l is basically a filtering mechanism that lowers the scores of potentially incidental contributions.

This principle can be applied recursively on all the sub-sequences of each pairs of sequences. The partition of the sequence can either be defined as sub-tree according to the processed context free structure, of $T_0(T_1, T_2)$; or some relatively equal length division of the linear sequence, such that $w = u.v \Big| -e < |u| - |v| < e$, with some fractional differential limit to the lengths between the two sub-sequences u and v. Each has its advantage. The sub-tree division would always produce morphosyntactically relevant constituent, thus would correspond to the linguistic analysis well, but tend to have less balance in branching. The length based division produces a more arbitrary and non-syntactic structure, but produces structures of more comparable sized sub-sequences, and has the additional benefit of catching associations that are not based on syntactic constituents. We experimented with both, and selected the length based sub-sequencing, since there is already another inductive sub-module in Section 7.3.3.6 that is largely informed by syntactically based substructures.

Assuming that the context-free form of the parsed and transformed data is in CNF, for the pair o_k, o_l, each has a pair of sub-sequences, $\langle o_k^1, o_k^2 \rangle$, $\langle o_l^1, o_l^2 \rangle$, looking at the similarities $o_k^1 \leftrightarrow o_l^1$ and $o_k^2 \leftrightarrow o_l^2$. This is assuming that the sub-sequences have the same alignment between o_k and o_l; which is the *cis* direction. Moreover, these comparisons can also be made in the opposite, *trans* direction, where we look at the similarities of $o_k^1 \leftrightarrow o_l^2$ and $o_l^2 \leftrightarrow o_l^1$. We can then use a discount function over the recursive components, and bias it toward one direction (normally *cis*). This weighting mechanism can be applied to any use of Forward or Viterbi on some pair $\langle o_l, \theta_m \rangle$. Thus the aforementioned mutual scaling between the Forward-derived and Viterbi-derived scores is performed after the application of this form of weighting on these pairs.

7.3.3.6 Subsequences and induction

In a similar way, we can utilize the internal structures of the observation sequences, with respect to their sub-constituents, to enhance the scoring of $\langle o_l, \theta_m \rangle$, and to allow us to catch patterns otherwise would be missed. The $S^{fwd}(o_l, \theta_m)$ and $S^{vit}(o_l, \theta_m)$ Forward-derived and Viterbi derived scores can be measured more than merely in terms of $P^{fwd}(o_l, \theta_m)$ and simple $Dist(Path_{vit}(o_l, \theta), Path_c) \; |Path_c = q_0 \rightarrow q_1 \rightarrow q_2 \rightarrow q_3 \rightarrow q_4$. We can take into account an inductive definition for both $S^{fwd}(o_l, \theta_m)$ and $S^{vit}(o_l, \theta_m)$, inducting on the subsequences of o_l and $o_k | o_k \in O_m$, corresponding to each's sub-constituents.

First, the multi-observation Baum-Welch algorithm is modified to accommodate this change; the set O is partitioned into $\{O_m^\tau\} \subset 2^O$, where m is the length of that subset of training sequences, which we term $\widehat{2^O}$. Here, τ is the starting symbol (constituent tag) for that part of the transformed tree. All possible subtrees that produce a terminal sequence longer than some minimal length is also added to O to form the meta-observation set $\widehat{2^O}$. For the $S_{fwd}^d(o_l, \theta_m^\tau) \; \Big| \theta_m^\tau = BW(O_m^\tau | \; O_m^\tau \in \widehat{2^O})$ at level d, some discounted contribution of its sub-constituents, such that $disc(c^{(d+1)})$ could become a passed parameter as a coefficient that discounts the level d, $S_{fwd}^{d+1}(o_l^1, \, \theta_p^\upsilon, disc(c^{(d+1)}))$ and $S_{fwd}^{d+1}(o_l^2, \, \theta_q^\phi, disc(c^{(d+1)}))$, where υ and ϕ are starting symbols of the corresponding subtrees. A process occurs for the $S^{vit}(\cdot)$ component of the score, and $disc(\cdot)$ is a discount function that inverse logarithmically scales with d, are also taken into account. So the composite score can be expressed as:

$$S_{\cdot}^{d}(o_l, \theta_m^{\tau}, c^{[d]}) := \begin{cases} c \times \Bigg(\; n^{f} \cdot S_{fwd}^{d+1}(o_l, \theta_m^{\tau}) + \\ n^{v} \cdot S_{vit}^{d+1}(o_l, \theta_m^{\tau}) + S_1 + S_2 \Bigg) \\ \quad \Bigg| \; S_1 = S_{\cdot}^{d+1}(o_l^1, \theta_p^{v}, disc(c^{[d]})) \\ \quad \bigwedge S_2 = S_{\cdot}^{d+1}(o_l^2, \theta_q^{\phi}, disc(c^{[d]})) \\ \quad \bigwedge o_l^1 . o_l^2 = o_l \ , \ \theta_m^{\tau} = \vartheta(o_l) \ , \\ \quad \theta_p^{v} = \vartheta(o_l^1) \ , \ \theta_q^{\phi} = \vartheta(o_l^2) \end{cases} \tag{7.18}$$

The expression $S_{\cdot}^{d}(o_l, \theta_m^{\tau}, c^{[d]})$ is now the composite scoring mechanism. Where n^f, n^v are the normalization factors for the forward and Viterbi components, which are scales that ensure that over the entire structure from the root node of the tree structure, the contribution between these two components are relatively equivalent in contribution. In fact, the scoring for both are done, and kept in separate data-structures, and the normalization is performed after the score computations are complete, but here they are represented in an equivalent single process form. Also, $\vartheta(o_k)$ is a function that looks up the machine with the parameter θ, that corresponds to the root symbol of the subtree that produces the sequence of terminals o_k .

7.3.4 Frame indexation

For any two frames the relative indices of the frames of the pair $\langle e_i, e_j \rangle$ is a prerequisite that qualifies the pair to be scored as potential adjunctive causal pairs, because the basic definition of an adjoined pair would be their adjacency in the sequential ordering of the frames within the corpus. The simple restriction of examining only $\langle e_i, e_{i+1} \rangle$ works in most of the cases. There are however, a significant population of pairs that might not have this indexical quality; these usually involve coordinate structures where parallel clauses are conjoined in a single morphosyntactic structure, as in the following examples, assuming that in each case (a) and (b) are immediately adjacent in the corpus:

1. (a) Patricia completed all of her course work, received an exceptional GRE score, and then wrote a well acclaimed senior thesis
   (b) So she is able to attend the best graduate program in the country
2. (a) John set up the trip wire, connected it to the alarm around the property, and went off to sleep, so that some foxes that are ruining his gardens at night were trapped in the pit
   (b) Mary unsuspectingly tripped over the wire falling to the ground
3. (a) Ian put the plan for the bank robbery on the kitchen table
   (b) Jane shopped for groceries at the market, drove home intending to make dinner, but then saw the bank blue-print laying on the table, decided to join Ian in the scheme

Each of these has at one of the two clauses representing some parallel structure of several different frames (sharing the external argument, as common in narratives), and at least one of these frames within a single composite clause has some causal relation with the following (1/2) or the following clause (3) in sequential order. For (1), we observe that all three of the contained frames from (1a) represents an event that fulfills some prerequisite of entrance into a graduate program, thus in some sense each of these three events (course-work, GRE exam, senior thesis) have some causal bearing on her eventually entering grad school; so there is some causal relation in each of $\langle e_i^1, e_{i+1}\rangle, \langle e_i^2, e_{i+1}\rangle, \langle e_i^3, e_{i+1}\rangle$ (where the superscript represents individual events within the conjoined (1a)). For (2), the first of these frames in the parallel structure contained in (2a), namely setting up the trip wire, has a causal relation with the event in (2b), Mary tripping on it; so while there is a relation between $\langle e_i^1, e_{i+1}$, as well as potentially within (2a) itself $\langle e_i^1, e_i^4\rangle$, where e_i^4 is the frame that describe catching foxes. For (3), it is similar that $\langle e_i, e_{i+1}^3\rangle, \langle e_i, e_{i+1}^4$ are causal relations, and the frames within (3b) also have some causal relation in that sequence, such as $\langle e_{i+1}^1, e_{i+1}^2\rangle$.

So we can observe that there is often the case where there is potentially causal relation between some frame within a parallel structure of a clause with some adjacent clause, or between some pair of frames within the parallel structure itself, and this is often the result of syntactic coordination. If we index each parallel structure with a single index as above, and only allow such pairs as $\langle e_i, e_{i+1}\rangle$, then there is no chance that any causal relation within the parallel structure is detected, and any potential causal relation with adjacent clause becomes less specific. Where as if we index each frame within the parallel structure with a different index, we may miss some of the causal relations between certain sub-frame within the parallel structure with an adjacent clause's frame, such as $\langle e_i^1, e_{i+1} \rangle$ in (2), in which case e_i^1, e_{i+1} themselves are not judged to be adjacent. So we use the scheme that adopts the above double indexation of e_i^d, where when any frame, or sub-frame within parallel structure has adjacent indices $i, i+1$, then this is a valid input for examination of adjunctive causality. For any pair of sub-frame within e_i, we allow for an $e_i^d, e_i^{d+c} | c > 0$ to be a valid input. Thus we are able to examine all of the appropriate potential pairs for causality, when there is parallel structure represented in syntax.

Many potential $\langle e_i, e_{i+1} \rangle$ are contained in clauses which are substantial in length, and would provide a good amount of immediate context for examining with HMM. Some of the cases where there is little or no additional structure outside of the core frame components of e_i, e_{i+1}, the existing linear structure may be insufficient of an immediate context. So in the cases of the short clausal pairs, additional material from e_{i-1}, e_{i+2} are added to the input for that particular observation sequence of the appropriate HMM for the $\langle e_i, e_{i+1} \rangle$ pair, this is simply done by measuring the total length of e_i, e_{i+1} minus the terminals that correspond to each frame's essential predicate and arguments.

7.4 Results

The procedure described above produces a relative ranking among all possible pairs of frame-transformed clauses in the corpus.

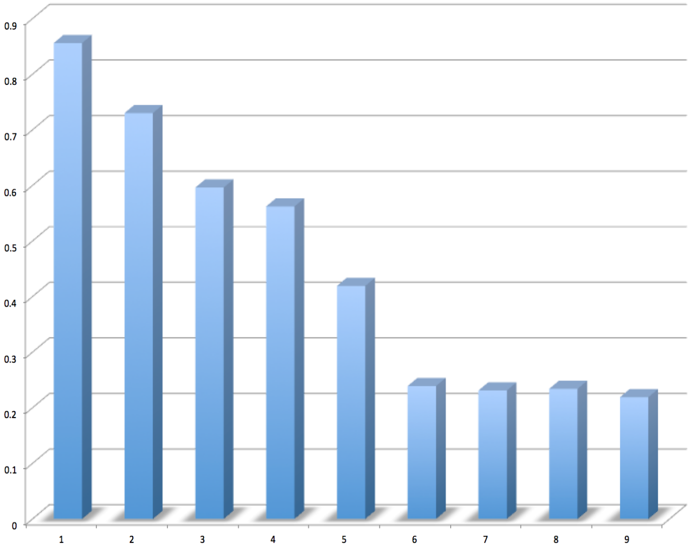
7.4.1 Quantile ranking

For evaluation, since adjunctive causal structures are likely a very small proportion of candidate adjacent pairs, and potentially has a long-tailed distribution, as described in Section 5.1.3, we used a sparse *quantile*-based annotation. We annotated three sets of $k = 100+$ (actually about 115 each, because we want to ensure that there are at least 100 determinable samples in each) samples. We can implement the scoring from the techniques described in the previous section alone, or in concert with other methods for ranking candidate pairs. Using only the procedure above entails that the extraction procedure utilizes only the information contained within the pair of clauses themselves, excluding any information from the SPEC-ELEM (likely animate *cause/agent* of each eventuality), and excluding information from the CORE-VP (the actual predicate and other obligatory internal arguments of the frames). Most attempts at extracting causal relations focus on either the *conditional* and *joint* probabilities of $e_1 \longrightarrow e_2$ (the core predicates and their respective essential argument list), or on the contextual information surrounding the two eventualities. But the largely structural information that is non-contextual and not within the primary parameters of the semantic frames is typically ignored. This is the primary area that we will explore in this case.

The first set of results includes only mechanisms and scoring that stems from unsupervised extraction of likely counterfactual-causals and training HMMs with this data. The annotation of adjunctive causal pairs is even more labor-intensive, given that the context is often needed for humans to determine causal structures, we annotated each until a clear pattern emerges. The total number of frames that have potential embedding within (having some other verbal predicate or gerund within) is 50160. Out of these, we have annotated the top 6 quantiles for BNC. We annotated top 1035 total annotated samples from the BNC set that have been annotated from that testing set; at which time we see a clear pattern from the results emerging. The top quantile results is at 85.5% precision, then dropping quickly to below 50% by the 5th quantile, and seems to have a long-tailed distribution. The following charts describe quantiles of the BNC and its top 6 quantiles:

For the novels testing set; there is a similar amount of attention to detail necessary for a reasonably accurate annotation. Novels as a genre tend to have complex set of individuals and events intertwined over a considerable portion of its plot, thus the temporal sequence of elements are strong from one part to another. The challenge differs slightly in that the immediate context may not be quite as important as in news stories or parliamentary proceedings, but the character development and long distance logical links between events

Figure 7.3: BNC results in its top 9 quantiles ranked according to our algorithm, each bar is the fraction positive within 115 samples of the quantile, the hollow bars are the partially annotated quantiles



became more prominent. Thus there is more work required in searching out the actions of characters in other parts of the plot-line of the story, or in events that are connected in logical sequence earlier in the novel. The total number of frames with potential embedding within is 41894. Out of these, we have annotated a total of top 1265 samples for this testing set. The top quantile results is at 85.4% precision, but seems to a more gradual descent; where it is below 50% by the 7th quantile, and we annotated until the 11th quantile at just above 30%; so this seems to have a even longer tail than the BNC testing set.

7.4.1.1 Cumulative quantiles:

It is also sometimes useful to see the binary discrimination power of an algorithm with respect to a task. So we also show the cumulative quantiles, such that each of the data-points is the precision of the subset of data from the 1^{st} (top) to the k^{th}, where k^{th} is the current quantile in question. This shows the predictable power of each potential boundary, if we turn some division between quantiles k and $k+1$ into the division in a binary classification.

For both BNC and novels sets, in terms of cumulative precision of down-to a certain

Figure 7.4: Novels results in its top 11 quantiles ranked according to our algorithm; each bar is likewise the positive fraction within a quantile

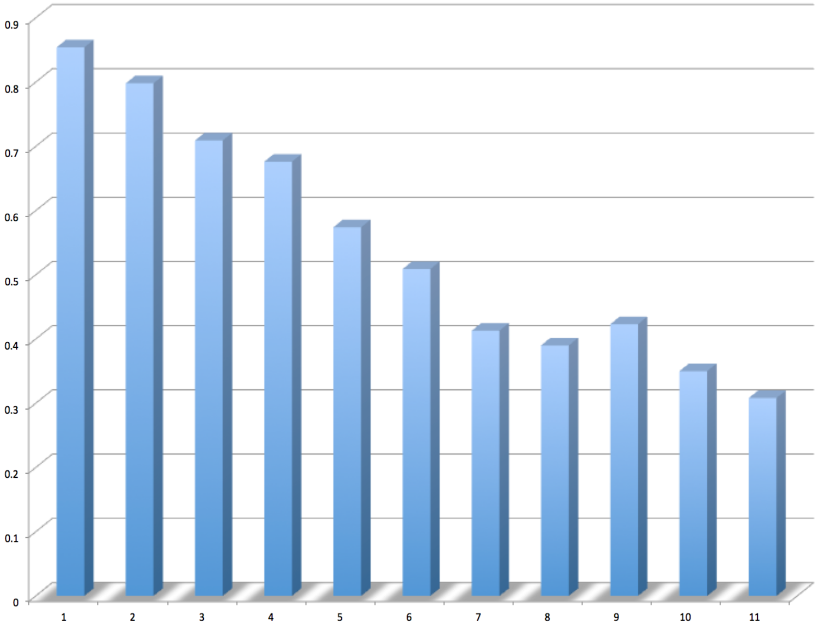


Figure 7.5: **Cumulative** results from BNC, where the k^{th} bar represents cumulative precision from 1^{st} through k^{th} quantile

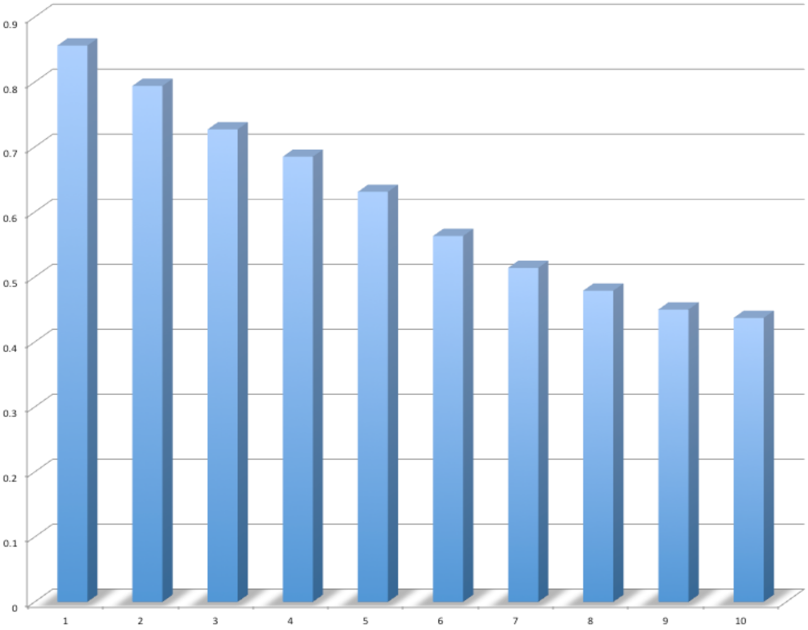
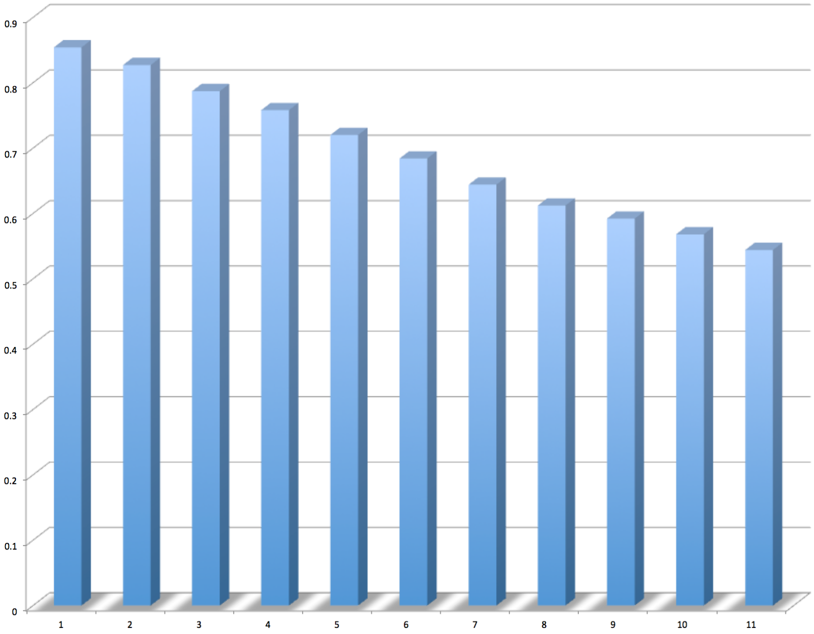


Figure 7.6: **Cumulative** results from novels, where the k^{th} bar represents cumulative precision from 1^{st} through k^{th} quantile



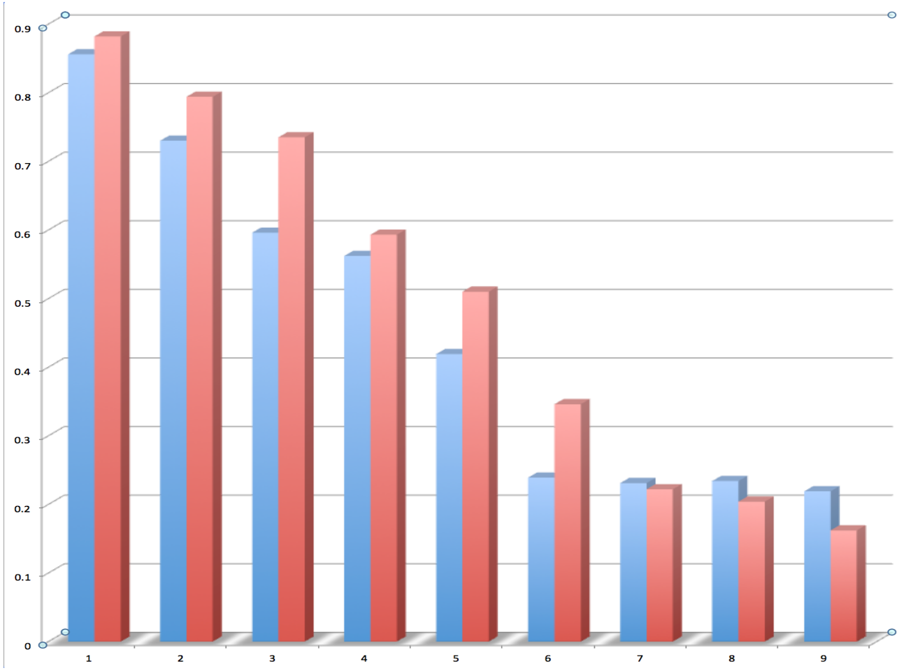
quantile, these are above 50% for the top 4 cumulative quantiles. If a simple threshold was designed for these types of data-sets, there should also be a convenient place between some two top quantiles to make a binary decision of whether some structure is embedded causal.

7.4.2 Composite with argument-agreement

We also performed a test to supplement our HMM-based algorithm with some additional argument-agreement metric between the candidate protasis and the candidate apodosis, which is a traditionally use metric (Do et. al. 2011). The metric contributes to to the scoring multiplicatively, where α_{min} is the base multiplier with no agreement between the two clausal forms. And $\alpha_{sim} = \frac{c \cdot |agreements|}{min(|arglist_p|, |arglist_a|)}$ is the measure of similarity between to argument lists. The final score $\hat{S}$ on each sample pair is obtained by $\hat{S} = (\alpha_{min} + \alpha_{sim}) \cdot S$ We produced a ranking based on this composite ADJ-CAUS + arg-agreement method, and the result of the top quantiles are compared to just the ADJ-CAUS method in Figure 7.4.2.

We observed that near the top of the ranking, this additional metric was able to raise

Figure 7.7: BNC results in top quantiles from our algorithm alone, compared against same results from our algorithm + argument agreement metric

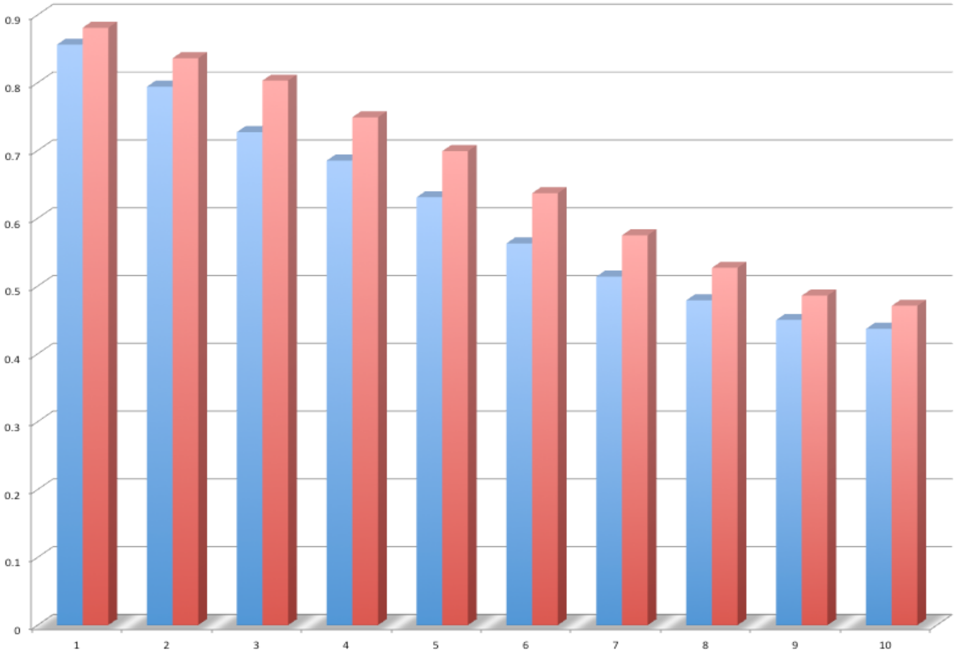


the precision of each quantiles. (This trend is no longer consistent after the 6^{th} quantile in the ranking.) So we can see that our ranking system is at least partially orthogonal to the ranking based on argument lists, and thus it could be improved on accuracy by more traditional metrics of detecting relations between pairs of frames.

7.4.3 Baseline comparisons

We are not aware of any comparable systems that are available for testing. Textual entailment (TE) systems are the most similar to complex causality. Thus, we used the textual entailment VENSES system (Delmonte et. al. 2007/2009). This test against annotations of adjunctive causality is not appropriate for the original purpose of VENSES, but is done with our data and annotation to see any correlation to our results, in order to compare to a reasonably close semantic task. For any given sample of testing set, we determine if pair of clauses, is identified as entailment by VENSES. The *protasis* in each case is considered *"text"* for the VENSES system, and *apodosis* considered the *"hypothesis"*. We compared

Figure 7.8: cumulative BNC results in top quantiles from our algorithm alone, with the k^{th} bar representing cumulative result of 1^{st} through k^{th} quantile, compared against same results from our algorithm + argument agreement metric

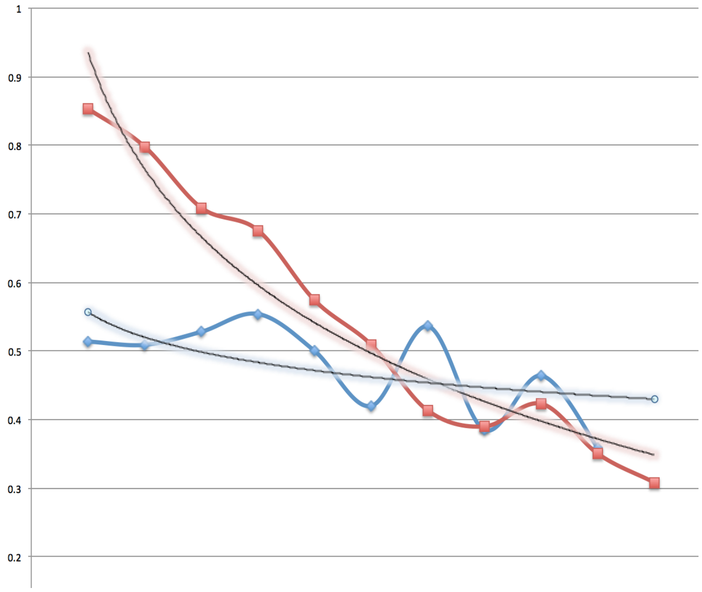


the results against our gold standard (for adjoined causality). The samples are the top 10 quantiles of the novels data-set (which is the set with more annotations), ranked according to our algorithm.

The samples are taken from the top 10 quantiles (those annotated for their causality), and tested to see whether VENSES considers the $\langle protasis, apodosis\rangle$ to be an entailed pair. We observe a weak but clear trend, of whether fraction in each quantile VENSES identifies as positive, in the same direction as our algorithm. This is what we expected, given that VENSES is designed for task similar to but distinct from complex causality, so it should be moderately correlated with the trend in our result. Our system is able to obtain significantly better discrimination and more consistently monotonic trend across the ranking.

Since our task is discovering causality in sequences involving two adjacent clauses, we also used n-gram model, a widely used generic algorithm on sequences of tokens (e.g. Brown et. al. 1992), to compare to our result here. There is no other unsupervised method

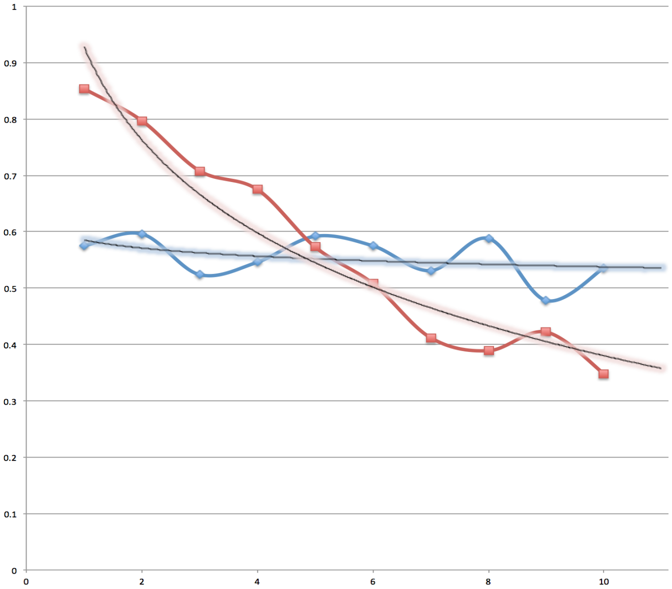
Figure 7.9: Novels results according to VENSES in its top 10 quantiles, samples drawn from top 10 quantiles according to our algorithm: each data-point is a quantile; the blue line shows the fraction these VENSES ranked quantiles is annotated by humans as adjunctively causal; the green line show the performance of our algorithm; there are also logarithmic trend-lines for both VENSES and our algorithm



built for the purpose of complex causality, and since our method in the adjunctive causal module is unsupervised, there is no training data for a supervised learning comparison with the identical training data. So we utilized RTE-anaphore_resolved corpus from LDC, which is another data-set that contains a two adjacent clause structure, but built for TE purposes (Pakray et. al. 2010). Even though the n-gram method has the advantage of being supervised, it has two inherent disadvantages compared to our method, ❶ it is less specifically built than ours for complex causal structures, and ❷ its training corpus is not specifically purposed for causality, but something else close to causal structures, which is itself very small for this type of training (only 130 usable positive examples). The training is standard n-gram with backoff and smoothing, with a length n padding at the beginning of sequence.

We re-ranked this subset using a 5-gram model trained on the RTE-anaphore_resolved data-set, taking the same subset of samples as the top 10 quantiles ranked according to our

Figure 7.10: Novels results according to 5-gram model in its top 10 quantiles, samples draw from top 10 quantiles according to our algorithm; each data-point is a quantile; the blue line show the fraction these 5-gram ranked quantiles is annotated as adjunctively causal; the red line shows the performance of our algorithm; there are also logarithmic trend-lines for both VENSES and our algorithm



algorithm as the testing data. The model produces a score on the testing data by taking the harmonic mean over the tokens in the sequence, so given the trained model N^5 and a sequence of length m, with padding of $\langle w_{1-n}, ..., w_0\rangle$, the score is: $\mathcal{H}(M^n(w_1|w_0, ..., w_{1-n})$, $M^n(w_2|w_1, ..., w_{2-n})$, ..., $N^n(w_m, |w_{m-1}, ..., w_{m-n}))$ Several models were tested, and found that beyond 5-gram model there is little change to the result. A slight correlation is found between the the n-gram result and our result.

7.4.4 Brief discussion of results

We will have a brief discussion of the adjunctive causal results here, with a fuller discussion in Chapter 9. Some of the top ranked examples do have high likelihood of causality just by looking at the predicates, such as *'you broke some rules, you deserve some punishment'*; a small number of them are aided by some discourse connectives, such as: *'when you feel discontented, think over your blessings. and you will be grateful'*. However, a majority of the samples in the top quantiles of each data-set contains that contained no causal connectives, and where the pair of predicates alone are not normally predictive of a causal relation, such as in the following:

1. (a) he is trapped in a body [that] is severely disabled
   (b) [he] will spend the rest of his life in a wheel chair
2. (a) the club plans to sell its present grounds in east london
   (b) [it plans] to move half mile down the road to create a leisure and community center
3. (a) ten thousand pounds will build you the highest column in the world
   (b) [the column] will produce an astonishing effect
4. (a) Half a dozen jovial lads were talking about skates in another part of the room
   (b) she longed to go to join them, for skating was one of the joys of her life

None of the above underlined pairs of predicates in of themselves would be considered sufficient to indicate causality, and all of the pairs of frames are have no connectives / adverbials between them at all, or are parallel sentences in coordinate structures. There are also a number of highly ranked pairs that require highly specialist knowledge to recognize the causal relation, which is recognize even though our system is does not cater to domain-specific relations:

1. (a) advance in the field of cosmetics means that, today, superfluous ingredients and allergens can be identified and substituted
   (b) this is great news indeed for anyone with sensitive skin
2. (a) regular exercise lowers ldl cholesterol level, yet raises hdl cholesterol level
   (b) it reduces the risk of heart disease

We will see that at a deep semantic level, adjunctive and embedded causal structures share similar characteristics and categorizations; and a comparison with embedded causal samples would allow for elucidation of any semantic distinctiveness of each; so further analysis is better left to Chapter 9 with the output of both modules.

Chapter 8

Embedded Causal Structures

Here, we pilot a study on the automatic extraction of one of the two types of structurally complex causal constructions in language (the other being Chapter 7). In this module, the frames which have some mutual embedding are examined, to see whether their structural details yield evidence as to whether relations among them can be classified as causal. We will utilize a representation *diffuse prototype* to take into account both lexico-syntactic, as well as structural information present in embedded causals; which is both flexible enough to account for a many-modal distribution in feature space, as well as being generalizable for formation of patterns. The algorithm used to produce this model from a small amount of training data is version of *genetic programming* (Cramer, 1985) adapted for our purposes.

8.1 Embedded Causality

Embedded causality is a complex form of causality that appears as deeply embedded structure that expresses the causality contained within through lexico-syntactically structural means.

8.1.1 Brief characterization

The causal nature of the expression is in part conveyed through the configuration of the structure. An example from our data-set found to be embedded causal would be:

1. a smart shower at eleven had evidently quenched the enthusiasm of the young ladies who were to arrive at twelve for nobody came and at two the exhausted family sat down in a blaze of sunshine to consume the perishable portions of the feast (prepared in anticipation of the guests) that nothing might be lost (Alcott, 1868)
    (a) a smart shower at eleven had evidently quenched the enthusiasm of the young ladies who were to arrive at twelve
    (b) $\xrightarrow{cause}$ nobody came
    (c) $\xrightarrow{cause}$ the exhausted family sat down in a blaze of sunshine
    (d) $\xrightarrow{cause}$ consume the perishable portions of the feast
    (e) $\xrightarrow{cause}$ nothing might be lost

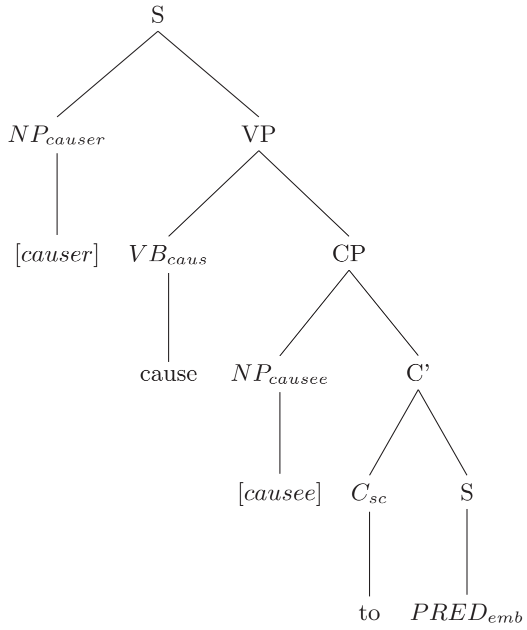
which in morphosyntactic terms is a single matrix clause that have multiple nested embedded clauses within. a Not that some small constituents such as *'for'*, *'to'*, *'that'*, etc also play a role in conveying causality of the entire structure to the reader.

The causal relation is primarily conveyed through the structural components that intervene between and through the frame components of each link in the causal chain. Although the semantics within the frames also contribute to causality, our task here is to examine the feasibility of determining causality from structural information, and develop a methodology for it. In order to study and to incorporate this class of structure into an automated system of causal detection and extraction, we must depend on the availability of positive examples of this class. The outer most matrix clause is relatively reliable to detect, since the vast majority of which contain an entity in the external position (the likely original causal entity), and a verb of the *manner-of-causation*-class that in some way indicate the causal force of e_0.

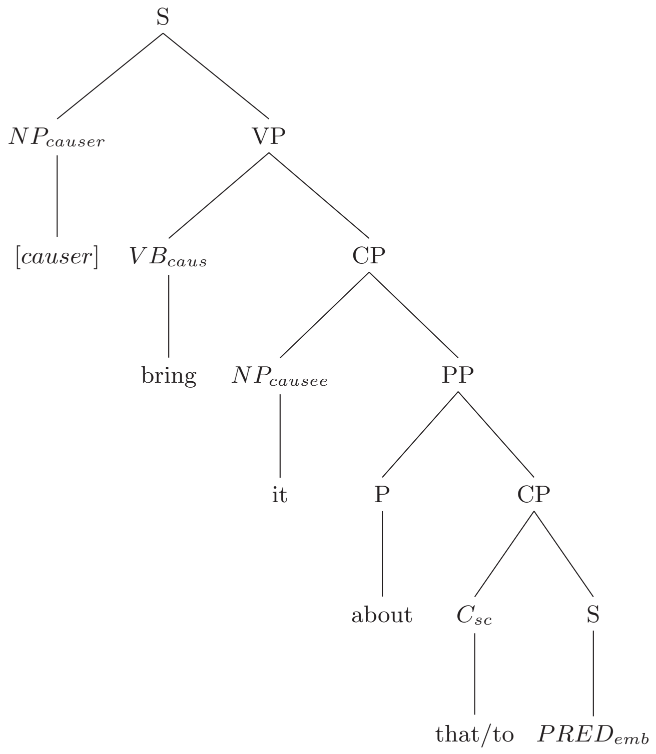
The presence of the inner embedded clause is much more difficult to detect for two reasons. One is that the classes of likely content verbs that head their VPs is much broader, the other is that they themselves may contain non-causally linked embedded clauses that contain separate events; such as *"the fireworks in the park brought it about such that kindles around the barn ignited so as to set the garage on fire where the antique vehicles are parked"*, where the ultimate e_n is headed by *set....on fire*, and not *park*, which is obvious when analyzing the temporal sequence of events. So the annotation of these examples needs to be focused on the location of the embedded clause; and the remainder of the structure will be implicitly indicated as part of an embedded causal structure. And for the purposes of training a model for lexico-syntactic pattern, there is no need for labeling detailed structure of each of the frames involved in the causal chain. We have annotated a sample of so far 500 clauses of positive examples of embedded causal constructions. The inner embedded clause of each is marked on their S/SBAR/SINV node to be ES/ESBAR/ESINV.

8.1.2 Issues specific to embedded causal constructions

Embedded causal structures, like complex hierarchical structures in language, have distinctive tree form; and a combination of their morphosyntactic and lexico-semantic characteristic give them the semantic property of frequently representing complex causal chains. For explaining embedded causal structures, we will denote the original causing event as e_1, the ultimate caused event as e_n, while the overall event of causation as e_0. The most generalized form of such constructions can be exemplified as such:

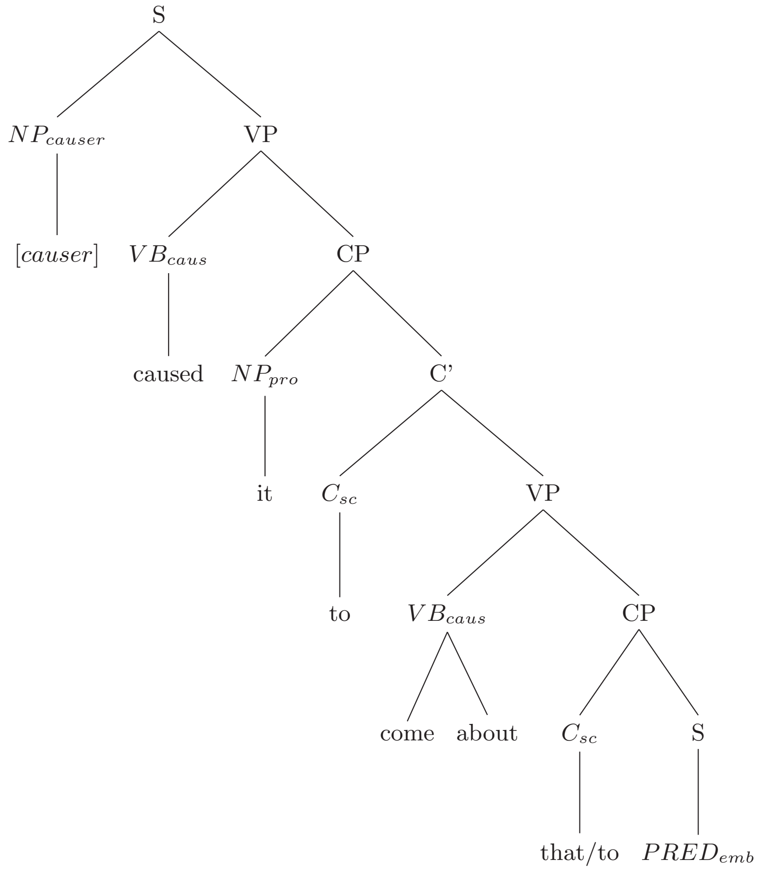


This class of causatives always contain at least one outer matrix clause that contains some representation of the entity or concept that is regarded as the original causer, and an innermost embedded (full or small) clause that represents the ultimate caused event. Within such, there are several defining components present in nearly all instances. There is present an entity represented in the SPEC position of the matrix clause, which is always the origin of the causal event (or a chain of causal events).



8.1.3 Distinctive elements of embedded causal constructions

There is frequently a distinctive verb of a certain semantic class present, that mediates the force of causation, and colors the causing event with additional semantic content, ranging from the meaning of *suggest* to that of *force*; this class of verbs we will term *manner-of-causation*-verbs. There is also an innermost embedded clause that represents some event at the end of the chain of causation. Additionally, there is a continuum of the depth of the semantic content of such *manner-of-causation*-verbs. The most vacuous of which consist the likes of *cause*, *bring-about*, which tells us little more than the presence of e_0 in the construction, where as the most meaningful spell out the specific event as the original e_1



Depending on its complexity, there may be one or more intermediate clausal structures that represent links in the chains of causation, along with intermediate causal agents. The presence of such intermediate structures at times have relatively small semantic contribution, such as in *"... caused the circumstances to line up in such a way as to ..."*, which mostly contributes the readers' understanding of the causal distance between the original cause and the ultimate caused event, but informs us little of the nature of the intermediate agents and events. At other times, such structures make significant contribution in referential semantics, pointing out specific entities and events involved in the chain of causation.

8.1.4 Semantic importance of embedded causal structures

Whereas the chief difficulty of adjunctive causal forms is the probabilistic nature of such construction, the main issue here is that embedded causals can be composed of arbitrarily complex nested structures, which along with the choice of strategic lexical items, mediate the causal force in their meanings. This class of causal constructions is especially important for the discovery of high probability long chains of causation within corpora.

Long chains of causal relations are frequently denoted by a complex embedding of multiple clauses through lexico-syntactic structures, structures which are causally linked. Following previous approaches (Menzies 2009, Beamer & Girju 2009), we define a causal relation as $e_1 \xrightarrow{cause} e_2$, where e_1 precedes e_2 temporally and, had e_1 failed to take place, e_2 would also not have taken place, or more generally, $P(e_2|e_1) > P(e_2|\neg e_1)$. This is a general and agreed upon definition of causality which encompasses various classes of causal types of interest (if one chooses to go deeper into this problem). Our unit of representation (for cause and effect) is a *semantic frame*, given by a predicate and a list of arguments in the form $\phi(ARG_i, ARG_j, ARG_k, ...)$. This corresponds to a clausal structure in morphosyntax which is usually embedded to express a causal chain, as in the following example (from *"Lily of the Nile"*):

1. I (Cleopatra Selene II) was a Ptolemy princess (meaning descended from Hellenic-pharonic blood-line), a queen in exile who must bide her time until she could think of some plot, some plan to [some plot/plan] return her to her throne

- *I was a Ptolemy princess*
- $\xrightarrow{caus}$ *[I was] a queen in exile*
- $\xrightarrow{caus}$ *who must bide her time*
- $\xrightarrow{caus}$ *she could think of some plot, some plan*
- $\xrightarrow{caus}$ *return her to her throne*

8.2 Diffuse prototypes representation

We need to encompass available lexico-semantic (symbolic) and morphosyntactic (structural) information into a single representation that can be compared and transformed. And since our goal is to extract causal chains from complex structures, the representation needs to generalize the information over the member frames/clauses. We mostly focus on the intervening information and structural configuration between clausal subtrees. The ideal product would be a set of maximally complex sub-structures in the reflection of their causality, which would not compromise their ability to generalize over all embedded causal structures. This model combines the strength of each to provide a representation that is both sufficiently flexible to represent a multi-modal distribution within feature space, sufficiently concise to be processed efficiently, as well as sufficiently generalizable to be able to easy represent common patterns in a semantic class. We will term this model *diffuse prototype*.

Due to the complexity of these elements and the context-free structure that knit them together for an embedded causal structure, there is combinatorially a very large number of possible embedded causal constructions, for embedded causal structures of a given depth. And the depth of the tree structures are unbounded. So it is very difficult to find all possible individual structure exemplars for each distinct type of these structures; and the required annotated dataset for supervised training of such would also be exceptionally large and impractical to compile.

This task involves finding some distinctive characteristics that are common in embedded causal constructions, and each of these abstract characteristics can be concretely extracted and stored as a set of closely related subtrees of causal structures. Expanding on a previous example, if we had a group of examples as:

i ENT_{causer} *caused it to come about that* ENT_{causee} *[*$PRED_{emb}$*....]*

ii ENT_{causer} *made it come about that* ENT_{causee} *[*$PRED_{emb}$ *....]*

iii ENT_{causer} *arranged the events so that it comes about that* ENT_{causee} *[*$PRED_{emb}$ *....]*

iv ENT_{causer} *had the foresight to prepare the circumstances so that it comes about that* ENT_{causee} *[*$PRED_{emb}$*....]*

All of the above express substantially the same causal relation between the *causer*, *causee*, and the event indicated by the embedded predication, possibly with some pragmatic variation on the indicated length of the causal chain. For all of the examples above, we can see that a subtree producing the terminals (when we treat the innermost predication $[PRED_{emb}]$ and each entity as a single terminal, and treat each morphologically inflected string as its lemma with $\lambda(\cdot)$) would be *"$\lambda(come{+}INFL)$ about that ENT_{causee} $[PRED_{emb}....]$ "*. A subtree like this can be used to further identify larger embedded structures as causal, and each embedded causative construction thus identified would contain one or more such subtrees. Consider the following:

a *ENT_{causer} arranged the events so that it comes about that ENT_{causee} [$PRED_{emb}$ ....]*

b *ENT_{causer} arranged the events so that it happens that ENT_{causee} [$PRED_{emb}$ ....]*

c *ENT_{causer} arranged the events so it results in ENT_{causee} [$PRED_{emb}$ ....]*

d *ENT_{causer} arranged the events so it brings about the fact that ENT_{causee} [$PRED_{emb}$ ....]*

Where the example (a) is (iii) from before, which also contains a second subtree that is common among embedded causatives, which is mirrored in the examples (b - d) here. These sub-structures can be considered partial prototypes, a set of which can allow us to stitch together prototype-like context free patterns that allow us points of reference for assessing whether a complex embedded structure is likely causal or not.

8.2.1 Computing over diffuse prototype

The algorithm we have chosen is one developed from genetic algorithm. The algorithm simulates the growth of subtrees that are shared between any two reference trees T, T', that we want to produce a map in between. In principle, the algorithm is based on simulation of a evolutionary process of a population of organisms in nature, also known as a *genetic algorithm*, based on Darwin's original work, and first conceived to be in an algorithmic form by Alan Turing. (Darwin, 1859; Turing, 1950; Barricelli, 1962; Rechenberg, 1973; Holland, 1975; Brindle, 1981; Baker, 1985 / 1989; Goldberg, 1989; Goldberg & Deb, 1993; Fogel, 1998) This process in principle occurs iteratively in generations, each generation contains new substructures that are potentially added to the population, and the appropriate components of this representation is 'grown' from the simplest possible structures. The makeup and the growth procedure is to be discussed in the next sections.

8.2.1.1 Sensible representation for adaptable extraction

The ideal product of such a process would be a set of structures that are at their maximal complexity which reflect their causality, but not compromise their ability to generalize over all embedded causal structures. So it eventually aims (but not guarantees) to produce a population of locally maximal subtrees for these two reference trees, each of which represents some shared of region of both. The eventual production of such a population of subtrees from many examples would allow us to have some type of diffuse prototype for us to compare any future unobserved example with, to tell how likely it is to be an embedded causal structure.

The initial generation consists of subtrees of single nodes that have equivalent labels. The notion of label-equivalence varies depending on the type of node, with syntactic tags (for non-terminals), and classes of surface tokens (for terminals). The token equivalence-classes range from broad, such as *manner-of-causation*-verbs or classes of detected named entities (Organization, or Location, e.g.), while others are as of individual lexical items (individual types of prepositions and Complementizers, e.g.). The edges of trees in this incarnation are identical in label, thus are disregarded except for recognizing the local topology. Given that a syntactic notion of a tree is rooted, we will use the directed notion of $N_T^{+/-}(v_i)$ to indicate children or ancestor of v_i with respect to T, whenever such an distinction is deemed necessary.

In this case, a purely parametric approach will not work for any tree structure of sufficient size, given the number of binary parameters that would need to represent the presence or absence of an edge $\langle v_i, v_j \rangle$ is $\mathcal{O}(n(T)^2)$, and number of possible configurations comes to $\mathcal{O}(2^{n(T)^2})$ without taking into account labels or other sources of complexity. In terms of cognitive models of categorization that we can draw on, *prototype* and *exemplar*s are the primary theories for consideration for most problems. A *prototype* representation has the advantage of simplicity, and requiring only the addition of a similarity function (edit distance) in order to effectively produce a model to recognize a property, if the problem can be reduced to such. A single prototype is ideal for representing a set of similar objects that can be unimodally represented in some feature space, which we will see soon, is not the case. A set of *exemplar*s has the advantage of allowing problems where the data is distributed in many modes in feature-space, each cluster in feature space may be represented by a single exemplar. This is also not ideal for this problem, at least in natural way where we would describe the feature space (based on tree topology), in which case the number of exemplars may approach the number of samples observed in the training set.

8.2.1.2 Diffuse prototype definition

Given the above considerations, we can provide a new categorial model that combines the strength of both prototype and exemplar theories, and specifically targeted toward the problem at hand. This model, like prototype theory, provides a relatively small number of individual structures as representatives of the class, thus allows a high degree of generalization over the class, and is far more concise than the list of samples in training. It also, like exemplar theory, allows for a high degree of adaptability in terms of a multi-modal distribution of the class over some naturally defined feature space, and provides the wide coverage of all different subtypes under the class. This representation can also be readily trained and modeled on tree structures, and is a natural derivative of the complex embedded structures that is the hallmark of this class of causative expressions.

This concept is a set of sub-structures, which are potential composite characteristics that are common to a subset of multiple exemplars, but has a notion of prototype edit distance from this central set of substructures, instead from a single prototype. This we will term a *diffuse prototype* of the class in data. This concept denotes that, given a feature space $X = [x_{[1]}, x_{[2]},\ ....\ x_{[n]}] \in \{0,1\}^n$, a substructure that could be considered a component within a *diffuse prototype* is $X_s = \{x_{[\kappa_j]}\} \mid j \in \kappa \sqsubset [1, 2, ..., n]$ such that $\exists Y^p, Y^q \in Y\ \forall j \in \kappa\ [Y^p_{[\kappa_j]} = Y^q_{[\kappa_j]}]$, where Y is the set of all positive samples for that semantic class. In other words, the samples Y^p, Y^q agree on some (usually substantial sized) substructure within the feature space.

This was the diffuse prototype conceived in an unordered and unstructured feature space, which is not what our problem presents. When the feature space itself is structured in some way, we need to introduce an additional constraint, that the substructures that can be members of the diffuse prototype must be *contiguous* by some definition. Take the simple example above, but where we have the contiguity constraint that κ must follow given some linear ordering, which X_s must follow for its contiguity definition. This requires that $\forall i, j \in \kappa \wedge X_{[\kappa_i]}, X_{[\kappa_j]} \in X_s$ and where $P_{i \to j} := \lhd i,\ \ldots,\ j \rhd$ is some consecutive sequence $\sqsubset \mathbb{N}$, we have that $\forall k \in P_{i \to j}\ [\kappa_k \in X_s]$ ($\sqsubset$ here symbolizes sub-sequence relation). So now in this example, any substructure that can serve in a diffuse prototype can no longer by arbitrarily maximized to the greatest common denominator between Y^p, Y^q, but must be restricted by some linearly contiguous region of X, with some notion of strict linear ordering, which in this case would be by the notion of ordering within $\mathbb{N}$

Given that the target structure for this representation are trees, we have a more complex structure of the feature space, such that the notion of *contiguity* in this case now refers to $N_T^+(v_i)$ and $N_T^-(v_i)$. And given that no universal notion of ordering such as in $\mathbb{N}$ applies for tree structures, the contiguity function is defined for individual tree substructures. In a scenario where T_s, T_t are subtrees of T, and $\{v_i\} = V(T_t) \setminus V(T_s) \wedge \langle v_i, v_j \rangle \in E(T_t)$, T_t is allowable in T as a contiguous extension of T_s, and thus can be considered a X_s if T_s is considered so. So we can reformulate the types of allowed substructures in the diffuse prototype as:

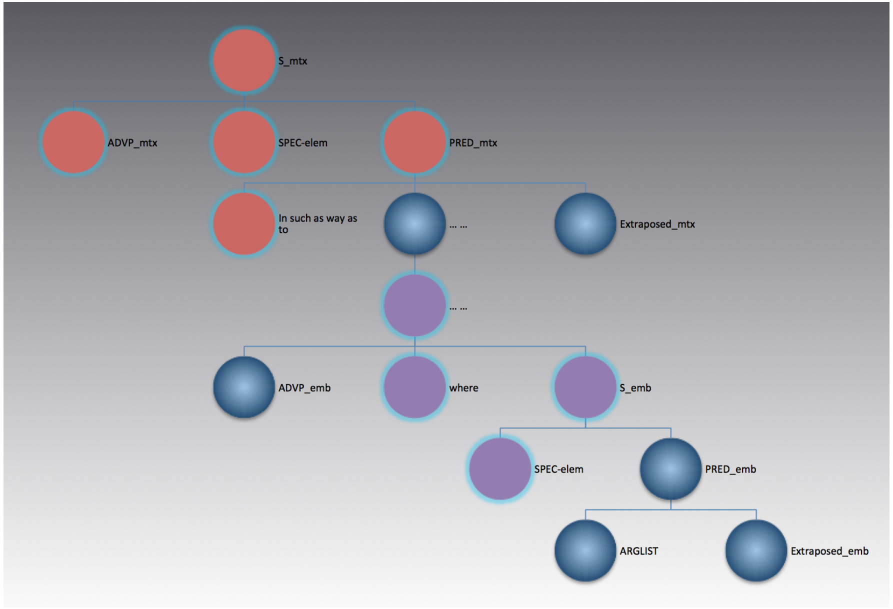
$$T_t = G(X_s) \begin{cases} \forall v_p, v_q \in V(T_t) \\ \exists P_{p \to q} := \ \lhd i,\ \ldots, k, \ldots\ j \rhd\ \sqsubset \kappa^P \\ \forall \kappa_k^P, \kappa_{k+1}^P \in \kappa^P \left[v_{[\kappa_{k+1}^P]} \in N_T^+(v_{[\kappa_k^P]}) \right] \end{cases} \tag{8.1}$$

Where κ^P is a specific ordering of $V(T)$ that conforms to the path P. In other words, the only type of X_s we are looking for, are the ones where we can, with its set of parameters, form a proper subtree T_t of the original tree. This is a natural way to allow generalization into members of the diffuse prototype, and thus some fragmented forest subgraph of T is not desirable. This notion of *contiguity* applied here is also important for the next phase of adapting this problem to *genetic algorithm*.

8.2.1.3 Diffuse prototype illustration

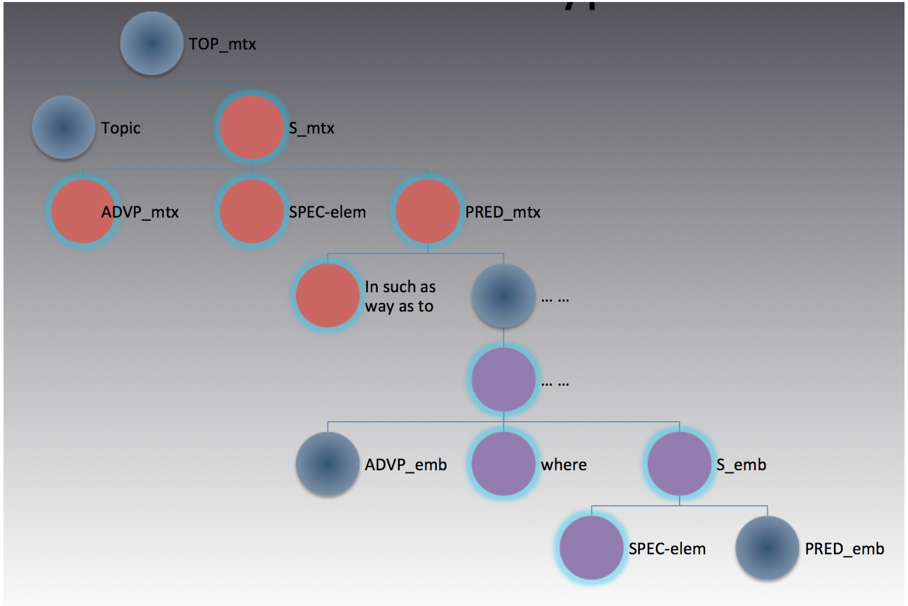
In terms of actual application of embedded causal structures, we can see that the trees T and T' in Figures 8.1 and 8.2, both contain some pair of substructures T_s and T_t corresponding to the red and violet regions, that can be used to predict that a structure is causal. For these and following pictorial illustrations, we only show a reduced structure, where some individual node in the graphic may represent several connected nodes in the real context-free tree; this is done to save space. The space in the paper does not allow a full presentation of all individual nodes, while maintaining each to be readable. Each shared sub-graph becomes a potential member in the diffuse prototype set, and the best such sub-graphs to serve in the diffuse prototype would be those that are contiguously maximized over some non-trivial (at least 2) set of positive samples. The shared subgraphs of both then can be used to further determine yet some other tree T'', where the variable regions (in blue-gray) can be quite different from either that of T or T'.

Figure 8.1:



The same applies to the pair of examples in Figures 8.3 and 8.4, with two sub-structures that correspond to the orange and green regions. The topologies in the remainder of the

Figure 8.2:



two trees vary considerably from one another, but the maximal shared regions in these can be used as indication for the desired property.

8.3 Embedded causative extraction procedure

The extraction of generalized patterns from trees is a computationally challenging problem, the key difficulty lies in the comparison of any two trees, and the mapping of their nodes with some defined notion of *isomorphism*. These has proven to be NP-complete when solved deterministically to a global minimum. Also, since a forest is a structure of unbounded complexity, where the existence of each node is orthogonal to the existence of any other, and existence of one edge is independent of most others, tackling it with learning using high dimensional space representation is not realistic either.

8.3.1 Baseline genetic algorithm

Inspired by the *On the Origin of Species* (Darwin, 1859), the diverse class of *genetic algorithms* is a wide array of adaptive algorithms (Rechenberg, 1973; Holland, 1975; de Jong,

Figure 8.3:

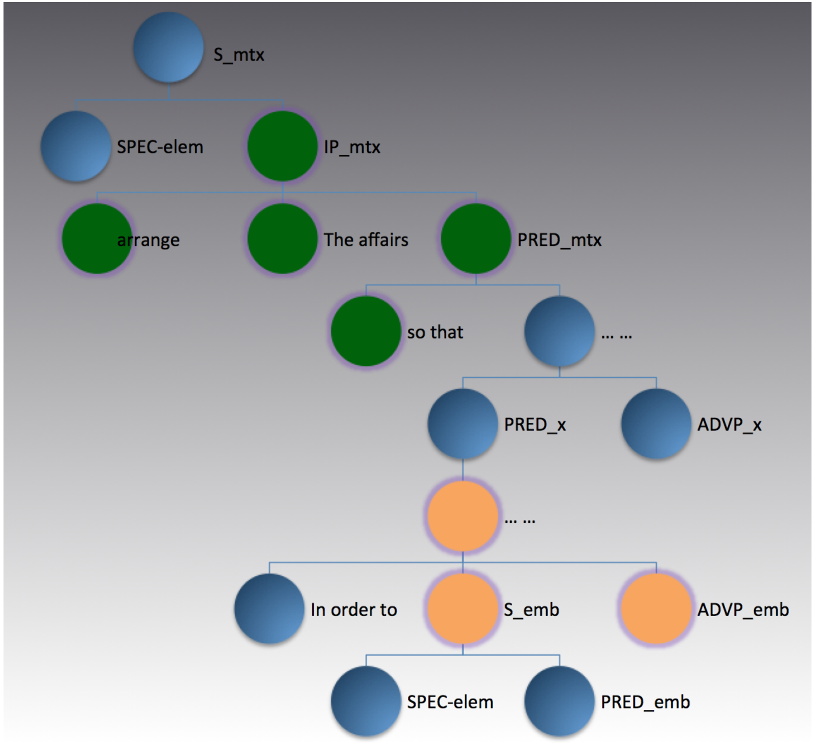
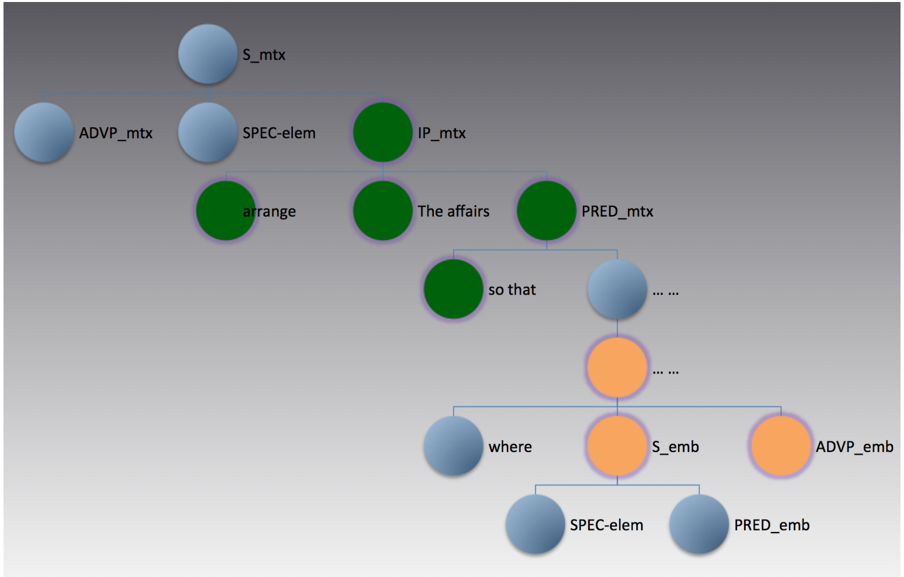


Figure 8.4:



1975) with biological origin. Here we will draw on traditional genetic algorithm practices, and change and adapt its concepts to fit the problem of discriminating embedded causal structures' diffuse prototypes.

8.3.1.1 Previous GA-related work

The form of evolutionary algorithm developed for our purposes has similarities to *genetic programming* (Cramer, 1985; Schmidhuber, 1987), which is closely related to GA. There also has been applications of GA to other areas of computational linguistics, in text-alignment problem in statistical machine translation (Otto & Rojas 2007; Rodriguez et. al. 2008; Bungum & Gamback 2010); formation of syntactic grammar from annotated text using *genetic programming* (de Pauw 2003); and the use of baseline-GA in word sense disambiguation (Decadt et. al. 2004). The design of the algorithm takes cues from the most general forms of GA, but also reaches back further into the inspirations of GA in the biological systems themselves, by emulating processes in biological systems in the operations within each iteration, in order to arrive at an effective and efficient method for discovery of substructures for *diffuse prototype*. Much of this is necessitated by the fact that the degrees of freedom, unlike in the typical GA, is variable in our case, where the growth of each substructure potentially affords more variability each generation. This process has a biological analog in macro-evolution, where the number of gene loci and their spatial organization on the chromosome potentially also vary over long periods of time.

8.3.1.2 Basic concept

The *genetic algorithm*s is central to the concept of evolutionary computing, and arrive at a solution mimicking, to varying degrees, the evolutionary process of organisms in nature. This is a multi-step alternating iterative metaheuristic, which bases the process of arriving at a desired solution heavily on *natural selection*, over multiple generations, where the members of one generation are in some manner variations of the previous. This stochastic process generally starts with an appropriate and efficient representation of the problem in some form of genetic encoding. Then some method for introducing variability is added for each generation, to produce sufficient variety for the selection mechanism to operate on. The *fitness function* then provides some evaluation of the suitability of each member, and selects the members for the formation of the next generation based on their fitness, and most of the time, also on some randomness in the process.

This class of algorithms has a very wide set of variations, with wide variety of forms for each of its basic components. There are numerous domains of applications for GA, which range from sociological modeling, to markets and economic simulations, to microprocessor circuit design and process scheduling, among many others. GA generally has no guarantee of finding the globally optimum solution of any problem, but provides a relatively efficient method of arriving at a good solution for problems large and high-dimensional search spaces, with noisy heuristics. It is also highly applicable for problems expected to have multi-modal solutions, where the number of clusters for a purely parametric solution is expected to be very high, which is the case with our embedded causal structures.

8.3.1.3 Generic elements of genetic algorithm

A number of elements and concepts are common to nearly all forms of GA. Generally, some way of expressing the problem's static representation, and dynamical properties in concepts of *chromosome/genome*, *allele*, *non-homogenizing operator*, *homogenizing operator*, *culling*, *carrying capacity*, and of course *selection pressure*, would be necessary to reformulate the problem in a way that is amenable to a GA solution. All of these are generally required in some form in order to complete the alternating multi-step process of the basic algorithm, in order to arrive at a solution of some (not necessarily global) optimum. Here, we will describe the generic strategies in formulating a problem according to GA criteria, and as a baseline formulation that we can perform further adaptation on, for out problem at hand.

The *genome* of an organism in the entire encoded set of information that provides the the blue-print for the morphogenesis process, which determines the organism's set of phenotypes, which is its set of heritable traits which can be expressed in a form that affects its ability for survival and reproduction. The set of phenotypes in turn affects the output of the *fitness function* with the environment also being an input. The *genome* is sometimes divided into discrete subsets of genephoric structures, where the genes within a single *chromosome*, a unit of inheritance of multiple genes, which have a higher probability of being passed onto the next generation together; while sometimes the entire genome would be composed of a single unit (as in Bacteria and Achaea). An *allele* is a possible variation of values within a single gene locus, where each individual organism has up to one or two (depending on whether one is within a haploid or diploid generation) distinct types; multiple allelic types of a gene is the primary mechanism for a population of organisms to have genetic diversity. In a specific computational problem, the entire set of variabilities of solutions need to be mapped in some way into a set of genes in the genome of a solution (individual organism), which may be further divided into discrete units, within each there is stronger associations of individual parameters. The possible values for each parameter is modeled as an allelic variation, and the morphogenesis process needs to be modeled in such a way as to translate these allelic variations in different survival and reproduction strategies.

Each ecosystem in nature has some specific environmental niche that an organism is adapting toward, and when that ecosystem is modeled to be static, adaptation gradually converges something close to optimal through a directed random walk process. For any environmental niche, there is a limited amount of resources and habitats for the organism to survive and reproduce in, which in turn limits the long term population limit for that niche. This is an important part of selection, which entails that non-competitive individuals would not (be likely to) survive. This means some *culling* of the population is necessary, which should have a *selection* component mediated by a *fitness function* that takes the organism's current set of phenotype and the environmental conditions as inputs, in order for the evolution process to be directed. The culling process naturally also requires some randomness, which in nature would be termed *genetic drift*, which prevents premature convergence on suboptimal solutions, but also has certain down-sides for many implementations. For a computational problem, it is necessary to model the fitness function and selection carefully, as that is the central directed drive toward a desirable solution. The individual alleles need to have some manifestation in each generation (phenotypes) that are relevant to the selection process; and either a hard or soft population threshold need to be modeled in order to provide the role of carrying capacity.

Ultimately, the genetic variation in a natural population is provided by the process of *mutation*, which is able to alter the allele of some gene locus of an individual to another allele, or to a previously unseen allele altogether. The allelic variation is not only essential for the selection pressure to operate on, it also furnishes the variability within a single gene locus for some crossover "mixing" to occur in reproduction, for those organisms with two copies (potentially two allelic variations) of the same locus. This crossover process is *recombination* of two chromosomes that contain the same set of gene loci. Due to the fact that *mutation* is capable of generating alleles de novo, it is considered the primary example of *non-homogenizing operator* on the population, as it promotes new variations of genetic make up. *Recombination* on the other hand, over multiple generations of reproduction, tends to make the genetic makeup of any chromosome similar throughout the population, hence is considered the primary example of *homogenizing operator*. For any computational problem to be modeled correctly, both the non-homogenizing and homogenizing operators need to be selected carefully, so that each generation of individuals reproduced would have the capabilities of generating new values for certain traits that may not be present in the previous generation, and some capability of integrating new values for these parameters into larger proportions of the population easily, should these new traits be evolutionarily favorable.

8.3.1.4 Formulating our problem in evolutionary terms

Seeded with these minimal subtrees in the initialization step, each successive generation is produced through three alternation steps. The first of which is a growth stage, where as each pair of parent subtrees (identical, but each subtree of a different sample $\in \mathbb{T}$) reproduces some sub-population of next generation subtrees, each of which is a defined region of both reference trees T and T'; this corresponds to the process of *mutation* in genetic algorithm, and allows each type of subtree to grow in one or more random directions. If some pair of subtrees cannot grow in a mutually agreed way, their descendents do not participate in the gene pools of future generations.

The second is a stage that, rather than create genetic features de novo, allows for the existing features to mix, thus creating new combinations of existing genes that would not occur given only mutation type operator. This is a process that is implemented as a type of *"recombination"* process of chromosomes, but requires considerable alteration for it to be applicable in a structured tree context.

The third is an elimination stage that culls a part of the population with respect to some notion of *fitness*, which corresponds to *directed selection*. This eliminates part of the produced generation of subtrees, based on some property that is related to their ability to survive and reproduce in future generations. A relatively simple metric that can be used is to take the $\frac{fertility}{fecundity}$ of the parents as their probable fitness. Much more complex metrics can be developed by studying the nature of syntactic trees in general, and the specific characteristics of embedded causal syntax.

There is additional component of elimination that culls a part of the population through the use of a random variable. This allows for some ability of the model to escape some local minima (simulated annealing (Kirkpatrick et. al., 1983)), and corresponds to the process of *genetic drift*. The magnitude of each of the cull processes is determined by a notion of *carrying capacity* (which is often important for producing the right type of convergence behavior (Goldberg et. al., 1991)). Our conception of carrying capacity is fixed with respect to the generations, but is variable with respect to the choice of the reference trees T, T'. This is preliminarily determined by a function with the orders of the reference trees, as well as the computational throughput of the machines that we are doing the computation on (primarily physical memory capacity).

8.3.2 Modifications to genetic algorithm

For our evolutionary algorithm, we have thoroughly reformulated the three primary operators, *non-homogenizing*, *homogenizing*, and *culling*, as well as how gene loci are structured, from the baseline GA. Each pass of this machinery allows for the generation of some prototype population of subtrees with respect to some substructure T_s that works as a common characteristic of set of embedded causal training samples. In a run of the GA, a series of $\frac{n^2}{2}$

passes would produce all of the populations of the subtrees that we need, each pass examines the possible population of substructures detectable for a pair of positive samples $\langle T, T' \rangle$. Some high confidence portion of these populations could serve as a diffuse prototype of the structures that we are looking for.

8.3.2.1 Objective function

To begin, we have a small number of positive training samples of complex embedded trees that convey causal chains. The environment, which to a large degree determines the objective function for a genome, in this case, needs to be modeled to facilitate the recognition of the traits that the positive samples have in common. For a minimal ecological niche, we need to find some lexico-syntactic structure that is shared by $h \geq 2$ positive samples. So the simplest case would be taking two positive samples $\langle T, T' \rangle$, and use that as a minimal niche where the evolutionary process may play out, to find the maximal sub-structures that may be shared between T, T'. So the process's object is to maximize the potential complexity of these substructures, so as to minimize the number of possible T configurations $\in \mathbb{T}$ that could contain such a sub-structure, thus maximizing the amount of specificity of each member of the diffuse prototype ($\mathbb{T}$ is the set of positive samples). These can be used later to construct more complex and higher confidence cases, to be discussed in the following sections.

The manifestation of the phenotypes, which is the morphogenesis process, can be very simply modeled in this case. Our genotype is cast as a piece of structural information within some induced subtree T_s of $\langle T, T' \rangle$ that convey causality, so an entire chromosome can be modeled as the set of parameters necessary to encode T_s, which we denote as ξ^{T_s}. Thus, the phenotype is simply whether ξ^{T_s}, once decoded into the structure T_s fits inside the environment as induced subgraph. Whether such a *"phenotype"* is well adapted for the *"environment"* can simply be a subgraph isomorphism test, which here-on we will denote as $\mathcal{I}^{\mathcal{S}}(T_s, T)$; even this step of computation, we will see shortly, is unnecessary since the algorithm is imbued with some properties of dynamic programming which makes isomorphism tests $\mathcal{I}^{\mathcal{S}}(T_s, T)$ on the entire substructure each time unnecessary.

8.3.2.2 Individual and population

In the cases where there is a need for several chromosomes to be preferentially heritable together, or where there is a need for packages of genetic material to be strongly associated and modeled as exchange of unit genephores of some time, such as *homologous chromatids*, then it is necessary to distinguish between the individual (a package of chromosomes) and a population (containing the entire pool of genetic material under simulation). Here it is sufficient to model individuals simply as single chromosomes. The representation of ξ^{T_s} as a chromosome/individual entails that the entire set of such sub-structures of $\langle T, T' \rangle$ becomes the population in question.

The various representations of all of the chromosomes in an evolving population can take up an inordinant amount of space. But given this particular type of structure T_s, subgraph of trees, we can leverage the original data-structures $\langle T, T' \rangle$ to provide most of the information of each chromosome. The information that ξ^{T_s} must contain are the locations of the boundary nodes of the substructure within T; and in order to facilitate the computational process, such boundaries of both T and T' are contained within ξ^{T_s}, where each point in the boundary is implemented as a pointer to a tree node. This can be formulated as the structure:

$$\xi^{T_s} = \begin{cases} \langle v_r, v'_r \rangle \, \Bigg| \, \Big[v_r \in V(T) \wedge \nexists v_s \in V(T_s)[v_s \in N^-_T(v_s)] \Big] \\ \qquad \bigwedge \Big[v'_r \in V(T') \wedge \nexists v'_s \in V(T'_s)[v'_s \in N^-_{T'}(v'_s)] \Big] \\ \\ \Big\langle V_l = \Big\{ v_l \Big| \exists v_m \in N^+_T(v_l) \\ \qquad [v_m \notin N^+_{T_s}(v_l)] \ \vee \quad |N^+_T(v_l)| = 0 \Big\} \\ \quad V'_l = \Big\{ v'_l \Big| \exists v'_m \in N^+_{T'}(v'_l) \\ \qquad [v'_m \notin N^+_{T_s}(v'_l)] \ \vee \quad |N^+_{T'}(v'_l)| = 0 \Big\} \Big\rangle \end{cases} \tag{8.2}$$

In other words, the structure ξ^{T_s} is a free-tree structure which is bounded by some root (v_r and v'_r are very likely the same node) with respect to the rooted-topology of $\langle T, T' \rangle$, and some set of other nodes that are leaf of T_s, or nodes that has at least one child not in

$V(T_s)$. So ξ^{T_s} basically contains a collection of pointers with respect to $\langle T, T' \rangle$, such that by moving these pointers around $V(T), V(T')$, we can precisely decode the current state of evolution for T_s. We will use the following functions to return the different components of ξ^{T_s}:

$$\begin{cases} v_r = \rho_{\check{T}}(\xi^{T_s}) \mid v_r \in V(\check{T}) \\ \mathfrak{o} = \rho_{\check{T}}(\xi^{T_s}) \mid v_r \notin V(\check{T}) \\ v'_r = \rho'_{\check{T}}(\xi^{T_s}) \mid v'_r \in V(\check{T}) \\ \mathfrak{o} = \rho'_{\check{T}}(\xi^{T_s}) \mid v'_r \notin V(\check{T}) \\ V_l = \lambda_{\check{T}}(\xi^{T_s}) \mid \forall v_l \in V_l[v_l \in V(\check{T})] \\ \mathfrak{o} = \lambda_{\check{T}}(\xi^{T_s}) \mid \exists v_l \in V_l[v_l \notin V(\check{T})] \\ V'_l = \lambda'_{\check{T}}(\xi^{T_s}) \mid \forall v'_l \in V'_l[v'_l \in V(\check{T})] \\ \mathfrak{o} = \lambda'_{\check{T}}(\xi^{T_s}) \mid \exists v'_l \in V'_l[v'_l \notin V(\check{T})] \end{cases} \tag{8.3}$$

For initialization of the process, we can take all of the nodes between $\langle T, T' \rangle$ that have the same symbol, and create the generation of $\mathcal{G}^0$ at $g = 0$, where $\forall \xi^{T_s} \in \mathcal{G}^0$, and G is the maximum number of generations, $\{\rho_{\check{T}}(\xi^{T_s})\} = \lambda_{\check{T}(\xi^{T_s})} \wedge \{\rho'_{\check{T}}(\xi^{T_s})\} = \lambda'_{\check{T}}(\xi^{T_s})$. Again $\varsigma(v_i)$ produces the relevant symbol at v_i

$$\mathcal{G}^0 := \begin{cases} \Big\{ \Big\langle \ \langle v_i, v'_i \rangle, \ \langle \{v_i,\}, \{v'_i,\} \rangle \ \Big\rangle \\ \quad \Big| \ v_i \in V(T), v'_i \in V(T'), \varsigma(v_i) = \varsigma(v'_i) \Big\} \end{cases} \tag{8.4}$$

8.3.2.3 Non-homogenizing operator

The non-homogenizing operator should be designed to create new variations in the genes that are previously absent from the population. Given that the starting point of the algorithm is a forest of trees where $V(T_s) = \varnothing$, and the goals of the algorithm should be ideally a set of maximal shared substructures of $\langle T, T' \rangle$, the natural direction of the iterative algorithm is to grow the substructures one step from one generation to the next. And given that larger substructures are constantly being created, such new structures are unlikely to have been seen before in the population, and thus is by nature non-homogenizing. The non-homogenizing operator of the system should be normally modeled on *mutation*, as simply flipping bits of information in a parameter vector, with the presence of each $v_i \in V(T)$ and $\langle v_i, v_j \rangle \in E(T)$. The initialization of such a representation would be a vector of zeros for the parameter set. But given the fact that we cannot efficiently encode all possible subgraphs of T, and use the far more efficient ξ^{T_s}, we perform basic operations of $T_s \xrightarrow{\cdots} T'_s$ by manipulating the pointers.

We can easily define an operation that might add a new vertex $v_i \in V(T) \setminus V(T_s)$ and edge $\langle v_i, v_j \rangle$ or $\langle v_j, v_i \rangle \in E(T), v_j \in V(T_s)$. This is easiest realized in two subtypes, because of the directed nature of T, to be $T'_s = \dot{+}_r(T_s, T)$ and $T'_s = \dot{+}_l(T_s, v_i)$. The following are the precursors of our non-homogenizing operator:

$$\dot{+}_r(T_s, T) = \begin{cases} G\Big(V(T_s) \cup \{v_i\}, E(T_s) \cup \{\langle v_i, v_j \rangle\}\Big) \\ \Big|\; v_j = \rho_T(\xi^{T_s}),\ v_i \in N_T^-(v_j) \end{cases} \tag{8.5}$$

The formulation above is with regard to T in the pair $\langle T, T' \rangle$, the corresponding $\dot{+}'_r(T_s, T')$ for T' has the same form with T replaced with T', $\rho_T(\cdot)$ replaced with $\rho'_{T'}(\cdot)$, and $\lambda_T(\cdot)$ replaced with $\lambda'_{T'}(\cdot)$; and the same symmetry also applies to $\dot{+}'_l(T_s, v_i)$ in the following formulation of $\dot{+}^l(T_s, v_j)$ in the following:

$$\dot{+}_l(T_s, v_i) = \begin{cases} G\Big(V(T_s) \cup \{v_i\}, E(T_s) \cup \{\langle v_j, v_i \rangle\}\Big) \\ \Big|\; v_i \notin V(T_s) \wedge \exists v_j \in \lambda_T(\xi^{T_s}) \\ \quad [\ v_i \in N_T^+(v_j)\ ] \end{cases} \tag{8.6}$$

In formulating our non-homogenizing operator, we consider that since both trees of the pair $\langle T, T' \rangle$ need to properly contain the substructure in order for it to be a meaningful indicator for embedded causality, we need to consider that those potential *"mutations"* that are not subgraphs of both to be evolutionary dead-ends. These should be analogous to those individuals that die before a reproductive age or have zero fecundity, and have no effect on the genetic makeup of the following generation. Thus the non-homogenizing operator needs to be constructed so as to produce viable off-springs. Again, for simplicity because of the directed graph nature of $\langle T, T' \rangle$, we formulate two non-homogenizing operators $\xi^{T'_s} =$

$\mu_r(\xi^{T_s}, \langle T, T'\rangle)$ and $\xi^{T_s'} = \mu_l(\xi^{T_s}, \langle T, T'\rangle)$, for both directions:

$$\mu_r(\xi^{T_s}, \langle T, T'\rangle) = \begin{cases} \Big\langle \langle v_j, v_j' \rangle, \langle \lambda_T(\xi^{T_s}), \lambda'_{T'}(\xi^{T_s}) \rangle \Big\rangle \\ \quad \Big| \Big(v_j = \rho_T(\xi^{T_s}),\ v_i \in N_T^-(v_j) \Big) \\ \quad \bigwedge \Big(v_j' = \rho'_{T'}(\xi^{T_s}),\ v_i \in N_{T'}^-(v_j) \Big) \\ \quad \bigwedge \Big(\ \varsigma(v_j) = \varsigma(v_j')\ \Big) \\ \\ \xi^{T_s} \quad \Big| \ otherwise \end{cases} \tag{8.7}$$

The mutation in the r-direction occurs only when both $\dotplus_r(\cdots)$ and $\dotplus'_r(\cdots)$ return valid structures that mutually agree. When the agreement is there, it returns $\xi^{T_s'}$ that has grown from the structure of T_s in the r-direction. For the mutation operator in the l direction:

$$\mu_l(\xi^{T_s}, \langle T, T'\rangle) = \begin{cases} \Big\langle \langle \rho_T(\xi^{T_s}),\ \rho'_{T'}(\xi^{T_s}) \rangle, \\ \quad \langle \lambda_T(\xi^{T_s}) \cup \{v_i\},\ \lambda'_{T'}(\xi^{T_s}) \cup \{v_i'\} \rangle \Big\rangle \\ \quad \Big| \ \exists v_i \in V(T), v_i' \in V(T'), v_i =_{T_s} v_i' \\ \quad \Big(v_i \notin V(T_s) \wedge \exists v_j \in \lambda_T(\xi^{T_s}) \\ \qquad [\ v_i \in N_T^+(v_j)\ ] \Big) \bigwedge \\ \quad \Big(v_i' \notin V(T_s) \wedge \exists v_j \in \lambda'_{T'}(\xi^{T_s}) \\ \qquad [\ v_i' \in N_T^+(v_j)\ ] \Big) \\ \quad \bigwedge \Big(\ \varsigma(v_i) = \varsigma(v_i')\ \Big) \\ \\ \xi^{T_s} \quad \Big| \ otherwise \end{cases} \tag{8.8}$$

The operator $v_h =_{T_t} v_k$ denotes that they are topologically equivalent with respect to the substructure T_t that is shared within the pair $\langle T, T'\rangle$. We can also design an equivalent μ operator for the direction of reduction in structural complexity of ξ^{T_s}, although that would be largely pointless since the reduced structure would have been produced in some past generation. So the direction of an evolutionary process only driven by the μ operator would be monotonic toward more complexity, and more relevance for constructing the diffuse prototype of embedded causal structures. During the non-homogenizing stage of a generation, each individual ξ^{T_s} within the population has a chance to undergo either $\mu_r(\xi^{T_s}, \langle T, T' \rangle)$ or $\mu_l(\xi^{T_s}, \langle T, T' \rangle)$. The probabilities are mediated by the random variables $R^{\blacktriangle}$ and $R^{\blacktriangledown}$, and the ratio between $R^{\blacktriangle}, R^{\blacktriangledown}$ is governed by the mean branching factor of the directed graphs T, T', so to ensure that the growth in all directions are approximately at the same rate.

8.3.2.4 Homogenizing operator

A homogenizing operator in a biological system or a GA serves to randomize the distribution of alleles and re-distribute new allelic types among the population. It accomplishes this generally by exchange of information between distinct units of inheritance (most of the time chromosome) between homologous gene loci. The most frequently used homogenizing operator among GAs is the process of *recombination*. In real or simulated sexual reproduction, the recombination process generally occurs between *chromatid*s of pairs of homologous chromosomes within the chromosomes of an individuals within a diploid generation (a generation where the functional loci within genome contain two alleles or two copies of the same allele). In haploid organism/generation (those without such duplication in individual genome), the process is performed often between individuals, such as exchange of *plasmid*s (smaller units of genetic information that can be integrated or excised from the bacterial chromosome) between individual bacteria. For our purposes, the recombination process resembles the latter model, given that we too do not make any distinction between individual genomes and chromosomes.

Given that our units of inheritance is the packet of encoding ξ^{T_s} for a substructure of $\langle T, T' \rangle$, the exchange of information is between some ξ^{T_s} and ξ^{T_t}. Homology is a difficult concept to translate from the biological model to a model consisted of graphs, and the concept of *homologous loci* on distinct chromosomes needs to be redefined to fit our mathematical structure. The standard concept for homology of two genes in the biological model is to view the genome for each spieces as a predefined map based on gene expression at individual loci, where each chromosome contains a per-determined set of loci fixed in some specific order, such that the two individual genomes are mapped as $X = [x_1, x_2,\ ....,\ x_n]$ and $X' = [x'_1, x'_2,\ ....,\ x'_n]$, and the homology is defined as $\langle x_i \xleftrightarrow{\mathcal{H}} x'_i \rangle$, and also for subsequences $\langle [x_i, ..., x_j] \xleftrightarrow{\mathcal{H}} [x'_i, ..., x'_j] \rangle$.

There is an alternate concept of homology that is applicable here, which is to view homology linearly contextually, such that for some subsequence homology $\langle [x_i, ..., x_j] \xleftrightarrow{\mathcal{H}} [x'_i, ..., x'_j] \rangle$, when set of pairings exit $\varkappa = \{\langle x_k, x'_k \rangle \mid x_k = x'_k, i - k \leq d \vee k - j \leq d\}$. This is essentially some set of identical pairs of genes in $\langle X, X' \rangle$, such that each pair within $\langle x_k, x'_k \rangle$ is identical distance from the target sequences respectively, and the distance is restricted within some distance d; the size of this identical set $|\varkappa| = c$ varies with the level of confidence we seek. This is based on probability, since for a pair on a homologous locus, there is a higher probability pairs of identical alleles from the target pair at the same distance from the locus.

We can provide an analogous operation to this taking our data-structure type into consideration. The size of each T_s and T_t would be quite small on average, so we can allow set of loci that are examined to be close to the target locus, and stipulate that the target locus needs to be contiguous with the group of context loci to be used with respect to the graph structure of T_s and T_t. The size c would then correspond to the number of nodes in some shared and contiguous portion between T_s, T_t, which is termed here $\varkappa^{T_s \xleftrightarrow{\mathcal{I}} T_t}$, where $T_s \xleftrightarrow{\mathcal{I}} T_t$ indicates partial isomorphism. There are multiple configurations where these can occur, we will focus on the two that are most amenable to formulating a definition of *recombination* according to our requirements.

These two disparate types of configuration are analogous to single-point and two-point cross-overs in linear genomes. The first type is a single contiguous region of shared loci between T_s and T_t, and is denoted as $\varkappa_{\Diamond}^{T_s \xleftrightarrow{\mathcal{I}} T_t}$. Given some minimum size requirement for the shared region $c_{\Diamond}$, we can formulate the construction as the maximum common subgraph of T_s, T_t, or algorithmically:

$$\varkappa_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t} := \begin{cases} V_m(T_s) \;\Big|\; \psi^{T_s,T_t}(V_m), |V_m| \geq c^{\Diamond} \\ \quad \nexists V'_m \subset V(T_s)\ [\psi^{T_s,T_t}(V'_m) \wedge |V'_m| > |V_m|]\ ; \\ \quad \psi^{T_s,T_t}(V_p) = \Big(\ \forall v_i \in V_p[\varsigma(v_i) = \varsigma(v'_i)]\ \Big) \\ \quad \bigwedge \Big(\langle v_i, v_j \rangle \in V_p(T_s) \leftrightarrow \langle v'_i, v'_j \rangle \in V_p(T_t) \Big) \end{cases} \tag{8.9}$$

where $V_i(G)$ here denotes the induced subgraph on G by the vertex set V_i, which is defined for the purpose of this project as:

$$V_i(G) = \begin{cases} V' = V(G) \cap V_i, \\ E' = \{\langle v_j, v_k \rangle \in E(G)\ \wedge\ v_j \in V_i,\ v_k \in V_i\} \end{cases} \tag{8.10}$$

The second type is two discontiguous regions of shared loci between T_s and T_t, denoted as $\varkappa_{\bowtie}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$, which is the pair of maximum common subgraphs of T_s, T_t that are disjoint, or algorithmically:

$$\varkappa_{\bowtie}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t} := \begin{cases} \langle V_m(T_s), V_n(T_s) \rangle \;\Big|\; \psi^{T_s,T_t}(V_m) \wedge \psi^{T_s,T_t}(V_n) \\ \quad \nexists \langle V'_m, V'_n \rangle\ \Big[\ V'_m \subset V(T_s), V'_n \subset V(T_t) \\ \quad \wedge\ (\psi^{T_s,T_t}(V'_m) \wedge\ |V'_m| > |V_m|) \\ \quad \wedge\ (\psi^{T_s,T_t}(V'_n) \wedge\ |V'_n| > |V_n|)\ , \\ \quad V'_m \cap V'_n = \varnothing\ \Big] \\ \quad V_m \cap V_n = \varnothing, |V_m| \geq c^{\bowtie}, |V_n| \geq c^{\bowtie}\ ; \\ \quad \psi^{T_s,T_t}(V_n) = \Big(\ \forall v_i \in V_m[\varsigma(v_i) = \varsigma(v'_i)]\ \Big) \\ \quad \bigwedge \Big(\langle v_i, v_j \rangle \in V_n(T_s) \leftrightarrow \langle v'_i, v'_j \rangle \in V_n(T_t) \Big) \end{cases} \tag{8.11}$$

Where we may denote the elements in the pair as $\varkappa_{\bowtie\,[s]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$, and $\varkappa_{\bowtie\,[t]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$. These shared regions of T_s, T_t of $\varkappa^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$ essentially function as a highly specialized form of a widely used technique *rank elitism* (Chakraborty & Chaudhuri, 2003; Mashohor, 2005; Yang, 2007;

Chudasama, 2011; Yaman & Yilmaz, 2012; Bora et. al. 2012; etc), that ensures that the cross-over mechanism produces highly adapted offspring chromosomes. Here, we have a form of *elitism* that is designed to operate specifically with our sub-structures, where the $\varkappa-$regions function to filter pairs of $\langle T_s, T_t \rangle$ so that only the highly compatible pairs would be able to undergo the homogenizing operation. This we may term *structural elitism*, and preserves the structural integrity of sub-structures that are highly effective in serving in the diffuse prototype of embedded causals, in the next generation.

Provided that we have found acceptable shared regions as in above, which accounts for the context(s) surrounding the target loci, the subsequent step would be to locate the actual target locus or set of loci from T_s, T_t used in recombination. For $\varkappa_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$ there is a group of one or more graph components induced by $V(T_s) \setminus V_m$ and $V(T_t) \setminus V_m$. We may choose two of these components from $(V(T_s) \setminus V_m)(T_s)$ and two components from $(V(T_t) \setminus V_m)(T_t)$, which we will term $\varrho_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$, and $\varrho_{\bowtie}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$. We will denote subgraph relation as $\trianglelefteq$, and will use $\divideontimes(G)$ to denote a set of all connected components of G, which uses the following:

$$\begin{cases} \divideontimes G = \Big\{ G(V_p) \mid V_p \subseteq V(G) \wedge \psi(V_p, T) \Big\} \\ \quad \psi(V_p, T) = \forall v_q, v_r \in V_p \ \exists \ P_{q,r} = [v_q, ..., v_r] \trianglelefteq T \end{cases} \tag{8.12}$$

to test for whether a V-set is a component. We will use $\binom{S}{c}$ for denoting the choosing of c elements from the set S. We will also employ a random variable R^S such that $\binom{S}{c}^{R^S}$ chooses according to a probability distribution so that the members $\in S$ that has the greatest size has the highest probability of being chosen (this in practice is rarely necessary, since S rarely contains more than 2 components). For the $\Diamond$ type regions, these can be formulated algorithmically as in the following:

$$\varrho_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t} := \begin{cases} \Big\langle \ \binom{S}{2}^{R^S}, \ \binom{S'}{2}^{R^S} \ \Big\rangle \quad \Big| \\ \quad S = \divideontimes\Big(\phi\Big(\ (V(T_s) \setminus V(\varkappa_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}))(T_s) \ \Big)\Big) \ , \\ \quad S' = \divideontimes\Big(\phi\Big(\ (V(T_t) \setminus V(\varkappa_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}))(T_t) \ \Big)\Big) \ ; \\ \quad \phi(E_p) = G(\ \{v_q \mid \langle v_q, v_r \rangle \in E_p \ \vee \\ \qquad \langle v_r, v_q \rangle \in E_p\}, \ E_p \ ) \end{cases} \tag{8.13}$$

We can denote the different elements within the target loci range to be: $\varrho_{\Diamond\ [S,0]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$, $\varrho_{\Diamond\ [S,1]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$, $\varrho_{\Diamond\ [T,0]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$, and $\varrho_{\Diamond\ [T,1]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$ respectively. The corresponding ϱ for the $\bowtie$ type regions then can be formulated algorithmically as the following:

$$\varrho_{\bowtie}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t} := \begin{cases} \Big\langle T_u,\ T_v \Big\rangle \ \Big| \ \psi(T_u, S, T_s) \ \wedge \ \psi(T_v, S', T_t) \\ \quad S = \divideontimes\Bigg(\phi\Bigg(\ (V(T_s) \setminus V(\varkappa_{\bowtie\ [s]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}))(T_s)\ \Bigg)\Bigg)\ , \\ \quad S' = \divideontimes\Bigg(\phi\Bigg(\ (V(T_t) \setminus V(\varkappa_{\bowtie\ [t]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}))(T_t)\ \Bigg)\Bigg)\ ; \\ \quad \phi(E_p) = G(\ \{v_q \mid \langle v_q, v_r\rangle \in E_p \ \vee \\ \qquad \langle v_r, v_q\rangle \in E_p\},\ E_p\ )\ ; \\ \quad \psi(T_w, S_x, T) = T_w \in S_x \wedge \exists v_i, v_j \in \\ \qquad \varkappa_{\bowtie}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}\ \Big[\exists P_{i,j} = [v_i,\ \ldots,\ v_j] \trianglelefteq T, \\ \qquad [\exists \langle v_h, v_k\rangle \in E(T_w)\ \langle v_h, v_k\rangle \in E(P_{i,j})]\Big] \end{cases} \tag{8.14}$$

where we denote the two elements as $\varrho_{\bowtie\ [S]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$, $\varrho_{\bowtie\ [T]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$. Given that all of the structures in question are trees, and knowing that for any tree T, $\forall v_i, v_j \in V(T)[\exists! P_{i,j} \trianglelefteq T]$ (exactly one path between any pair of nodes through T), so the new component after the division of T_s by $\varkappa_{\bowtie}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$, there will be exactly one component that lies on the path between the two parts of $\varkappa_{\bowtie}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$.

There is also a pair of random variables $R^{\Diamond}$ and $R^{\bowtie}$ that gives the probability that each of the $\Diamond$ or $\bowtie$ type operator would be conducted on any pair $\langle T_s, T_t\rangle \in \mathcal{G}^g \times \mathcal{G}^g$ of the g^{th} generation's population; these random variables depend on the relative sizes of T_s, T_t. Then we can define the homogenizing operation of the algorithm, with the $\Diamond$ type operation with regard to the $[1]-$component, defined in the following as $\eta_{\Diamond}^{s \rightarrowtail t}(s \rightarrowtail t)$ (t $[1]-$component grafted onto T_s), and $\eta_{\Diamond}^{t \rightarrowtail s}(T_s, T_t)$ (s $[1]-$component grated onto T_t); we will show the first of these processes in detail in the following (the complementary process can be worked out easily):

$$\eta_{\Diamond}^{s\rightarrowtail t}(T_s, T_t) := \begin{cases} V^{s\rightarrowtail t} = (V(T_s) \setminus V(\varrho_{\Diamond\ [S,1]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t})) \\ \qquad \cup\, V(\varrho_{\Diamond\ [T,1]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \\ E^{s\rightarrowtail t} = E(\ V^{s\rightarrowtail t}(T_s)\ ) \cup \\ \qquad E\left(\varrho_{\Diamond\ [T,1]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}\right) \ \cup\ \{\langle v_i, v_j' \rangle\} \\ \qquad \Bigg|\ \Big(\phi_{\Diamond}(\langle v_i, v_j \rangle) \vee \phi_{\Diamond}(\langle v_j, v_i \rangle)\ \Big) \\ \qquad \bigwedge \Big(\psi_{\Diamond}(\langle v_i', v_j' \rangle) \vee \psi_{\Diamond}(\langle v_j', v_i' \rangle) \Big)\ ; \end{cases} \tag{8.15}$$

$$\begin{cases} \phi_{\Diamond}(\langle v_h, v_k \rangle) = \langle v_h, v_j \rangle \in E(T_s)\ \bigwedge \\ \quad \Bigg(\Big(v_h \in V(\varkappa_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \wedge v_k \in V(\varrho_{\Diamond\ [S,1]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \Big) \\ \quad \bigvee \Big(v_k \in V(\varkappa_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \wedge v_h \in V(\varrho_{\Diamond\ [S,1]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \Big) \Bigg) \\ \psi_{\Diamond}(\langle v_h, v_k \rangle) = \langle v_h, v_j \rangle \in E(T_t)\ \bigwedge \\ \quad \Bigg(\Big(v_h \in V(\varkappa_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \wedge v_k \in V(\varrho_{\Diamond\ [T,1]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \Big) \\ \quad \bigvee \Big(v_k \in V(\varkappa_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \wedge v_h \in V(\varrho_{\Diamond\ [T,1]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \Big) \Bigg) \end{cases} \tag{8.16}$$

This process basically takes the necessary nodes and edges from the graft $[1]-$component of T_t, and the remaining components of T_s, and add a new edge to it so that both components are still attached in the same configuration as they were when they resided in T_s and T_t. The auxiliary functions $\phi_{\Diamond}(\cdot)$ and $\phi_{\Diamond}\cdot$ ensures that those attachment configurations are preserved. And the definition of $\eta_{\Diamond}$ of the other configuration, with respect to the $[0]-$component, would be an analogous structure, only with the identity of the graft altered, such that the $[0]-$component of T_s is grafted onto T_t, and the $[0]-$component of T_t is grafted onto T_s. As with before, we will show one of these processes, the complementary process can be easily worked out. The $t \rightarrowtail s$ process is algorithmically represented in the following, which is the mirror image of the previous process of $s \rightarrowtail t$ when the same $\varrho_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t} \Leftrightarrow \varkappa_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}$

attachment is involved:

$$\eta_{\Diamond}^{t\rightarrowtail s}(T_t, T_s) := \left\{ \begin{aligned} & V^{t\rightarrowtail s} = \left(V(T_t) \setminus V(\varrho_{\Diamond\ [T,0]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t})\right) \\ & \qquad \cup V(\varrho_{\Diamond\ [S,0]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \\ & E^{t\rightarrowtail s} = E(\ V^{t\rightarrowtail s}(T_t)\ ) \cup \\ & \qquad E\left(\varrho_{\Diamond\ [S,0]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}\right) \ \cup\ \{\langle v_i, v'_j \rangle\} \\ & \qquad \Bigg| \left(\phi'_{\Diamond}(\langle v_i, v_j \rangle) \vee \phi'_{\Diamond}(\langle v_j, v_i \rangle)\ \right) \\ & \qquad \bigwedge \left(\psi'_{\Diamond}(\langle v'_i, v'_j \rangle) \vee \psi'_{\Diamond}(\langle v'_j, v'_i \rangle)\right)\ ; \end{aligned} \right. \tag{8.17}$$

$$\left\{ \begin{aligned} & \phi'_{\Diamond}(\langle v_h, v_k \rangle) = \langle v_h, v_j \rangle \in E(T_s)\ \bigwedge \\ & \quad \left(\left(v_h \in V(\varkappa_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \wedge v_k \in V(\varrho_{\Diamond\ [T,0]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \right) \right. \\ & \quad \left. \bigvee \left(v_k \in V(\varkappa_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \wedge v_h \in V(\varrho_{\Diamond\ [T,0]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \right) \right) \\ & \psi'_{\Diamond}(\langle v_h, v_k \rangle) = \langle v_h, v_j \rangle \in E(T_t)\ \bigwedge \\ & \quad \left(\left(v_h \in V(\varkappa_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \wedge v_k \in V(\varrho_{\Diamond\ [S,0]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \right) \right. \\ & \quad \left. \bigvee \left(v_k \in V(\varkappa_{\Diamond}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \wedge v_h \in V(\varrho_{\Diamond\ [S,0]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \right) \right) \end{aligned} \right. \tag{8.18}$$

When there are only two components for each of T_s, T_t when discounting the shared nodes, then the above two types of processes are equivalent. We can observe the similarity of the $\Diamond-$type operation to single-point cross-over in linear genetic structures, as only a single new edge $\langle v_i, v'_j \rangle$ is needed for the formation of the new composite sub-structure, the addition of the edge being analogous to a ligase-mediated splicing mechanism.

The $\bowtie$ $-$type operation, on the other hand, is correspondingly similar to a two-point cross-over in its procedure, with two new edges $\langle v_i, v'_j \rangle,\ \langle v_p, v'_q \rangle$ necessary for the formation of the composite. And we can define the $\bowtie$ type operation, with the two recombinations as $\eta_{\bowtie}(T_s, T_t)$. We will demonstrate the direction of grafting a T component onto the remainder of T_s, the complementary process can be worked out accordingly.

$$\eta_{\bowtie}(T_s, T_t) := \left\{ \begin{array}{l} V^{\bowtie} = (V(T_s) \ \setminus V(\varrho_{\bowtie\ [S]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t})) \\ \qquad \cup\, V(\varrho_{\bowtie\ [T]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \\ E^{\bowtie} = E(V^{\bowtie}(T_s)) \cup E(\varrho_{\bowtie\ [T]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \cup \\ \qquad \{\langle v_i, v'_j\rangle, \langle v'_p, v_q\rangle\} \ \Big| \ v_i \neq v_q \ \bigwedge \\ \quad \Bigg(\Bigg(\phi_{\bowtie}(\langle v_i, v_j\rangle) \vee \phi_{\bowtie}(\langle v_j, v_i\rangle)\Bigg) \bigwedge \\ \quad \Bigg(\phi_{\bowtie}(\langle v_p, v_q\rangle) \vee \phi_{\bowtie}(\langle v_q, v_p\rangle)\Bigg)\Bigg) \bigwedge \\ \quad \Bigg(\Bigg(\psi_{\bowtie}(\langle v'_i, v'_j\rangle) \vee \psi_{\bowtie}(\langle v'_j, v'_i\rangle)\Bigg) \bigwedge \\ \quad \Bigg(\psi_{\bowtie}(\langle v'_p, v'_q\rangle) \vee \psi_{\bowtie}(\langle v'_q, v'_p\rangle)\Bigg)\Bigg) \end{array} \right. \tag{8.19}$$

$$\left\{ \begin{array}{l} \phi_{\bowtie}(\langle v_h, v_k\rangle) = \langle v_h, v_k\rangle \in E(T_s) \ \bigwedge \\ \qquad \Bigg(\Bigg(v_h \in V(\varkappa_{\bowtie}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \wedge v_k \in V(\varrho_{\bowtie\ [S]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t})\Bigg) \\ \qquad \bigvee \Bigg(v_k \in V(\varkappa_{\bowtie}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \wedge v_h \in V(\varrho_{\bowtie\ [S]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t})\Bigg)\Bigg) \\ \psi_{\bowtie}(\langle v_h, v_k\rangle) = \langle v_h, v_k\rangle \in E(T_t) \ \bigwedge \\ \qquad \Bigg(\Bigg(v_h \in V(\varkappa_{\bowtie}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \wedge v_k \in V(\varrho_{\bowtie\ [T]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t})\Bigg) \\ \qquad \bigvee \Bigg(v_k \in V(\varkappa_{\bowtie}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t}) \wedge v_h \in V(\varrho_{\bowtie\ [T]}^{T_s \overset{\mathcal{I}}{\longleftrightarrow} T_t})\Bigg)\Bigg) \end{array} \right. \tag{8.20}$$

Similar to the previous $\Diamond-$type operator, the $\bowtie-$type operator takes the necessary nodes from the two shared regions between T_s and T_t and any non-shared regions that does not connect the two shared regions. It then includes all of the nodes in the graft component (the region between the two shared regions, on the other substructure). It also includes all of the edges where both the origin and the terminus are in the above mentioned regions. Finally, it includes two new edges, making the connection between these regions, while preserving the local configurations at the attachment points.

While it is complex to describe algorithmically the process of this *"recombination"*, it is considerably easier to show pictorially, which are illustrated in figures 8.5 - 8.8. In each of these illustrations, we have some samples of $\langle T, T'\rangle$, in which are embedded a pair of

Figure 8.5:

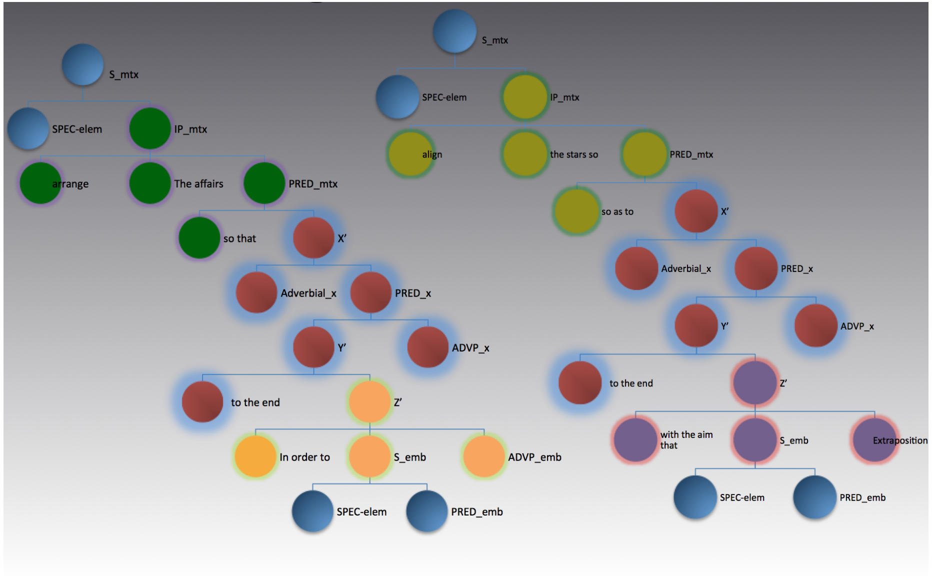
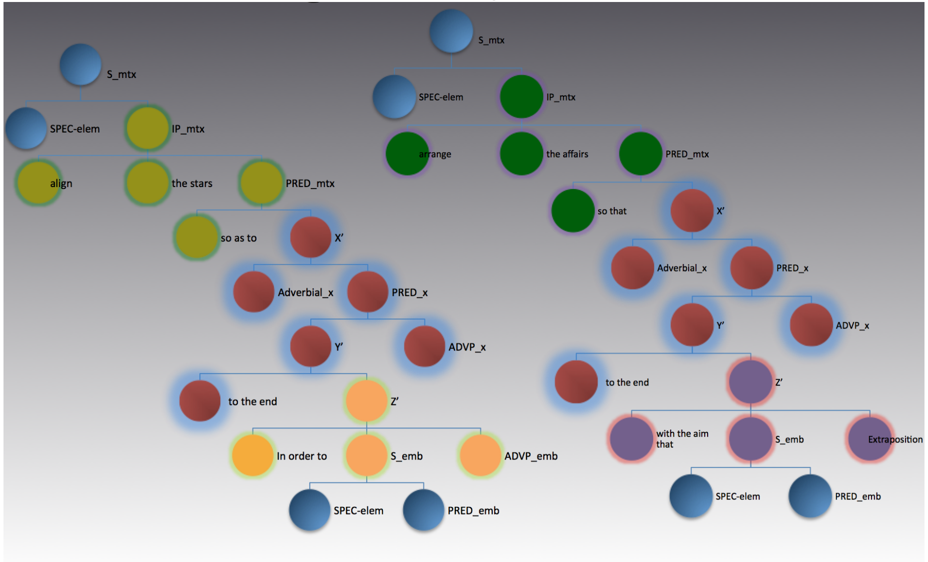


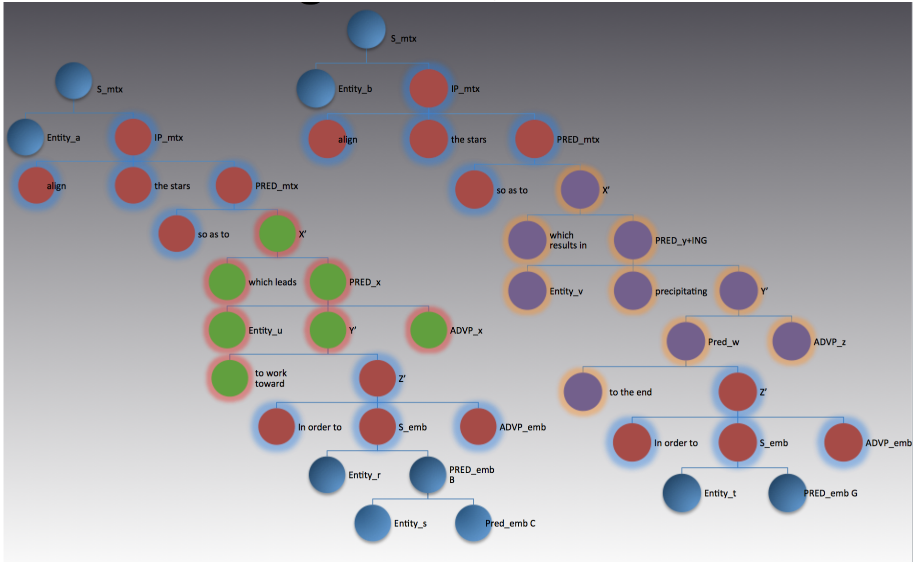
Figure 8.6:



substructures $\langle T_s, T_t \rangle$ that could undergo some homogenizing operation at that generation; the regions where the node-complexes are represented in powder blue are outside T_s, T_t, thus are not relevant for consideration for the homogenizing operation.

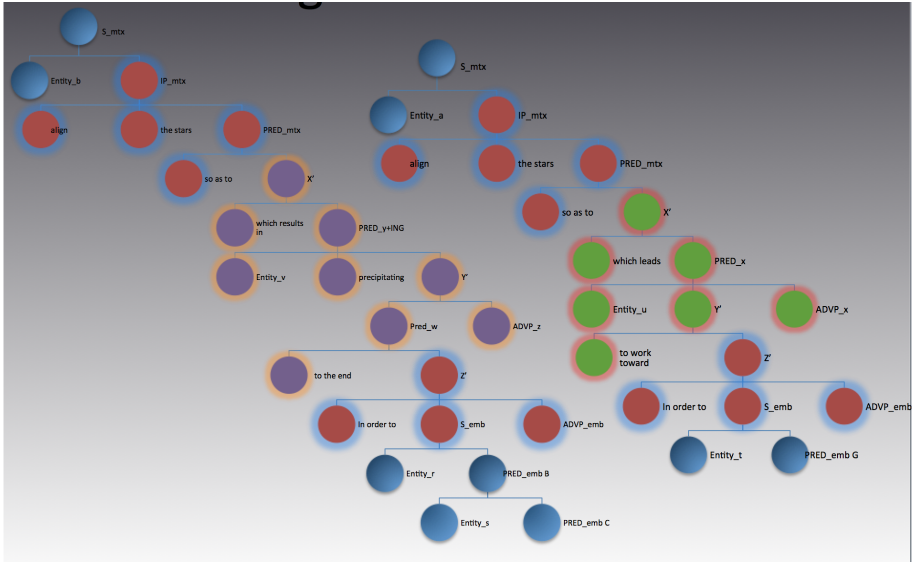
We can observe the the graphical illustration of the $\Diamond-$type homogenizing operation between Figure 8.5 and Figure 8.6. In this pair, there is a single red region of structure that is shared between T_s, T_t. The two regions colored in orange and green within T_s, and the two regions colored in yellow and purple within T_t are the non-shared regions that can undergo $\Diamond-$type *"recombination"*. Given these, we can observe the results of a $\Diamond-$type homogenizing operation on these two sub-structures, which is illustrated in Figure 8.6. Here T'_s preserves the original T_s structure, except for the yellow region, and has a non-shared green region from T_t grafted onto it; and T'_t conversely preserves the original T_t structure, except for the green region, while has a non-shared yellow region grafted on. in this instance, since T_s, T_t each has only two non-shared contiguous regions, an exchange of the yellow and green regions is equivalent to an exchange of the orange and purple regions. However, if there is a third non-shared contiguous region between $\langle T_s, T_t \rangle$, then both forms of the $\Diamond-$type operation are needed to provide all of the possible outcomes for the homogenizing operator.

Figure 8.7:



The Figures 8.7 and 8.8 illustrate the process for a $\bowtie-$type operation. Here again, the powder blue components of the $\langle T, T' \rangle$ graphs are outside of T_s, T_t, and not relevant for consideration of the homogenizing operation; and the red regions are the shared regions

Figure 8.8:



between T_s and T_t. Where as the $\Diamond-$type operation had one such region under consideration, the $\bowtie$ $-$type operation now has two disconnected regions that are shared between T_s, T_t. The graft region is framed by the two shared regions (as stipulated by $\bowtie$ $-$type operator), green in T_s and purple in T_t. There is only one graft region for each sub-structure, so even if there are additional non-shared regions between T_s and T_t, there will only be a single configuration of T_s' or T_t'.

Note that for each of the pictorial example above, we showed those $\langle T_s, T_t \rangle$ where $\varkappa_{\bowtie}^{T_s \xleftrightarrow{\mathcal{I}} T_t}$ and $\varrho_{\Diamond}^{T_s \xleftrightarrow{\mathcal{I}} T_t}$ are the only regions of these sub-structures present. This is for clarity and due to the size of the trees, not generally true. If this were the general case, then there is reason to undergo the $\bowtie$ $-$type operation for any such pair, since then $\eta_{\bowtie}(T_s, T_t)$ would produce the same substructures as the originals. So there should be regions in $\langle T_s, T_t \rangle$ in addition to $\varkappa_{\bowtie}^{T_s \xleftrightarrow{\mathcal{I}} T_t}$ and $\varrho_{\Diamond}^{T_s \xleftrightarrow{\mathcal{I}} T_t}$ in order for this type of operation to perform meaningful addition to the genetic diversity of the population.

8.3.2.5 Culling operator

The processing of *culling* is a technical way of expressing the removal of individuals from the population because of an underlying event of death or infertility; meaning that the individual genome no longer has relevance for the ensuing generation. Death is an essential component of evolution in nature, and only with some significant death rate in a population, could natural selection have an opportunity to apply its pressure. In a biological system, this process is a mixture of some *directed selection*, which depends on the fitness of an organism in a specific ecological niche, or a set of niches that it migrates to and from; and some *randomized selection* process, that provides a chance for the most fit and least fit individual within the population alike to perish without reproducing during that generation. The directed selection can be modeled through the use of a fitness function that best fits the data-type, in which case it is usually mediate only at transitions between generations (non-Lamarckian). The randomized selection in nature is actually a collection of multiple processes, the most important of which are *genetic drift* and *immigration/emigration*. Migration in our case is not modeled, since there is no comparable change of our set of training samples during the evolutionary process.

In the directed component of the culling process, only a fraction of the population is allowed to reproduce, or persist in the population, in order to affect the genetic makeup of the future gene pool. The directed selection is the primary driver for adaptation of the genomes in a population to occur, when the environmental factors remain static over a number of generations; which is the case in our situation, where the *"ecological niche"* for some substructure T_s is the positive sample pair $\langle T, T' \rangle$. The primary metric of usefulness of any substructure T_s is its complexity measured as $n(T_s)$; the higher the complexity of the structure, the less likely it is to occur in some randomly chosen context-free structure, and those complex sub-structures that conform to more than one positive embedded causal samples is highly likely to be relevant for testing whether any future encountered example would also be an embedded causal.

The ultimate goal of the GA process is to produce the largest and most complex set of sub-structures that are shared among all positive training samples. So for each $T_s^g \in \mathcal{G}^g$, we desire to have the most possible extensions of that sub-structure, given the constraints within $\langle T, T' \rangle$. In other words, this entails the maximization of the number of possible non-homogenizing operations that can be done on T_s^l. So we have a basic formulation of the *fitness* based on two factors, the total capacity of the organism to reproduce given the reproduction rate and the span of reproductive life, termed *fecundity*, and the actual reproduction rate given the population and environmental factors, termed *fertility*. We use the representation $f(T_s) = \frac{fertility}{fecundity} = \frac{\mathfrak{f}^{T_s}}{\mathfrak{e}^{T_s}}$ to denote the important ratio that measures how well each T_s uses its reproductive opportunity given $\langle T, T' \rangle$.

In order for the selection pressure in the genetic algorithm to be more relevant to linguistic data and to the problem at hand, we incorporated some basic data-mining techniques into the fitness function. Another potential factor in the usefulness of a sub-structure T_s is the occurrence and distribution of individual terminal symbols of T_s within the corpus. There are several different data-mining metrics that can be plausibly incorporated into the fitness function, for the initial testing, we incorporated *lift* of the token present in the terminals of each tree, where the *lift* of the token-type is computed for trees labeled embedded causative, against all trees in the training data. This process likewise increases the usefulness of the extracted substructures. Let $\tau(T_s)$ be a function linearizes the available terminals of the tree T_s, the fitness function can be formulated as the following, where X_E is the set of terminal sequences that come from trees in the positive embedded causal samples, and $X_{E\&i}$ are samples that show both traits:

$$\begin{cases} f(T_s) \propto \frac{\mathfrak{f}^{T_s}}{\mathfrak{e}^{T_s}} \cdot \frac{\displaystyle\sum_{x_j \in \tau(T_s)} \mathcal{L}(X_E \Longrightarrow x_j)}{|\tau(T_s)|} \\[2em] \mathcal{L}(X_E \Longrightarrow x_j) = \frac{\mathcal{S}(X_{E\&i})}{\mathcal{S}(X_E) \times \mathcal{S}(X_i)} \mid x_j \in X_i \\[2em] \mathcal{S}(X_i) = \left. \frac{\displaystyle\sum_{x_j \in X_i} \mathcal{N}_j}{\displaystyle\sum_{x_k \in X} \mathcal{N}_k} \right| \mathcal{N}_j \propto n(X_j) \end{cases} \tag{8.21}$$

One widely-employed selection strategy is Boltzmann selection, based on the principles of *simulated annealing*, widely used in implementations of GA (Goldberg, 1992; de la Maza & Tidor, 1992 / 1993). This type of selection is used for the employment of variable selection pressure with time, where the tolerance for sub-optimal adaptation of individuals is high early on in the process, but the tolerance is reduced over generations to ensure convergence to (near-)globally optimal solution efficiently. The Boltzmann protocol has a general form that is similar to the following:

$$\begin{cases} P(x_i) = \dfrac{e^{-F-\frac{f(x_i)}{\mathfrak{T}}}}{\displaystyle\sum_{x_j \in X} e^{\frac{f(x_j)}{\mathfrak{T}}}} \\ \mathfrak{T}^g = \begin{cases} \mathfrak{T}^{g-1}(1.0-\alpha)^{\beta} \ |g \geq 1 \\ \mathfrak{T}^0 \quad |g = 0 \end{cases} \\ \beta = 1.0 + c \cdot \frac{g}{G} \\ 0.0 < \alpha < 1.0 \end{cases} \tag{8.22}$$

Here $\mathfrak{T}^g$ is analogous to temperature variable in other simulated annealing processes. $\mathfrak{T}^g$ decreases with each generation. In earlier generations, the relatively high value of $\mathfrak{T}^g$ acts to slow the convergence process by not always selecting the optimal improvements at each step; but in later generations, it gradually cools down to allow the GA to settle in on a global maximum. This works well in most traditional GA implementations, when the maximum degrees of freedom are allowed to be explored early on among the generations. But in our application of GA for the problem at hand, because of data structure in substructure T_s, as well as how the non-homogenizing operator is defined to work with that phenomenon, the degrees of freedom is not constant, and in fact may increase dramatically over time. The sub-structure T_s^g can only build on the complexity of some T_s^{g-1} of the previous generation, and all of the additional degrees of freedom only likely becomes available when T_s^{g-1} is in the population. So in our case, Boltzmann selection not only does not perform the necessary annealing function (at the cost of time-complexity), but may even be counterproductive to arriving at a global maximum. The selection procedure ultimately employed is a *roulette selection* process, given that the variability of fitness within a single generation (generally with a similar level of complexity) is small.

8.3.2.6 Genetic drift

Genetic drift in a natural system is the change in the composition of genes in a population due to random sampling. This process is not directed by some selection pressure, or any anisotropic influence on the population. The best known variant of genetic drift is that of population bottleneck due to some non-adaptive selection process on the population; by non-adaptive we mean that this process occurs with a set of unusual consequences or within a time-scale where natural selection is unable to make meaningful adaptations on the population, e.g. a catastrophic meteor shower where the set of individuals who perished are essentially not dependent on how survivable they are in the natural ecology. Not all drift occurrence are so dramatic, but all of them share the characteristic that population may lose some alleles in the gene pool with no dependence on the phenotypical fitness of these extinct alleles. Genetic drift is the primary mechanism for organisms' difference in evolution for several population in different locations with similar ecological niches, since drift by itself without consideration of any directed evolutionary mechanism, with no genetic information communicated among isolated populations, tends to drive the populations each in a random direction.

An analogous process occurs in implementations of GA, since some randomness are built into each of culling, homogenizing, and non-homogenizing genetic operators. Thus randomness, and the possibility of random movements of the gene poll in the search space, is built into the core mechanisms of GA itself. Drift in relatively small population generally has the pernicious effect of reducing the genetic diversity of that population by extinguishing some arbitrary subset of allele. Drift in a purely parametric model space tends to obstruct large portions of the search space, and prematurely reduce the genetic diversity within the population so that the algorithm never nears optimal solution(s).

However, our formulation of GA have inherent characteristics that allow the system to resist negative effects of genetic drift. Given the way $\mathcal{G}^0$ was defined, for any maximal shared substructure T_s for the pair T, T', there are at least $n(T_s)$ starting points available that could potentially grow into the full T_s at or after generation $\mathcal{G}^{n(T_s)}$; each of these starting point potentially has multiple paths to arrive at T_s. So the more complex (hence more useful and important for the diffuse prototype) the eventual target sub-structure, the more starting points and paths in the algorithm it can be reached through the algorithm. Thus this property inherently makes it unlikely that the most important members of the diffuse prototype would be eliminated by chance through genetic drift. Due to the nature of our chromosome ξ^{T_s} representing a sub-structure, each $\mathcal{G}^g \Rightarrow \mathcal{G}^{g+1}$ generational transition has a strong tendency to make available new degrees of freedom not considered in the previous generation. And this new variability has a high probability of being taken advantage of right away, given two dependents $T_s^1, and T_s^2$ derived from previous generation substructure T_s. So early on in the run of the algorithm, drift affects relatively few degrees of freedom, the optimal value of each has numerous different paths to arrive at in a later generation.

8.3.3 Prototype consolidation

Once the population of substructures have been extracted, we need score each testing tree sample against this collection of diffuse prototype, thus we need to facilitate this process and reduce time complexity by consolidating the number of identical substructures in the population. Those substructures that occur in the greatest fraction of the $\frac{n^2}{2}$ populations would be the ones that we have the most confidence in, in terms of their ability to identify embedded causal structures. Thus, we need to find any possible identities among the population, this is ultimately achieved through tree isomorphism.

This process itself potentially has high time complexity, potentially $\mathcal{O}(n^{k+4.5})$ where k is the degree limit for vertices in the graph (Bodländer, 1988), although the complexity is already much lower than sub-graph isomorphism with the earlier situation before applying the genetic algorithm . The k limit exists due to our transformation of the trees, guaranteeing that a *Chomsky normal form* exists there, where each vertex has degree of at most 3 when the structure is treated as a free-tree.

Additional pre-filters, such as number of vertices, degree-list of the vertices, label-histogram of the vertices, can be applied in addition, to further reduce complexity to far below the original problem, so that the isomorphism comparisons only need to be finally done on a small subset of the substructures generated through the genetic algorithm. Thus the cost in time-complexity in the consolidation of the diffuse prototype is far lower than that of the running of GA, even though isomorphism tests are performed during this stage.

8.4 Results

We conducted some debugging runs with mock dataset, all proved successful and the system functions in principle. Next is the preliminary run with real-world data. As mentioned in Section 5.1.1.2 biggest issue for the selection of dataset was fraction of data that express frames that are sufficiently complex to have a significant probability of being deeply embedded causal.

The testing datasets used was pre-processed from a portion raw BNC data of 1963314 lines, and the raw novels corpus of 129695 lines. As mentioned, the novels corpus is much more monolithic, while having some level of complex structures throughout. The BNC data, on the other hand, has limited portions that are genres much more complex and conducive (than average) for forming embedded causals, but much of it would be in genres where such structures almost never occur (i.e. poetry). The smaller amounts of data used for training was due to computational complexity issues and hardware limitations. The initial run of the embedded causative structure extraction presented some issues with computational space complexity of the algorithm, which reached over 25GB in virtual address space for even these limited sets. Some of this is unavoidable due to the large search spaces for each stage. These may be improved in the future by further optimization for memory usage, and/or running this on a more capable system.

8.4.1 Testing parameters

The data used in this initial phase of testing also comes from BNC and novels, and are annotated after the completion of the run, and the training data comes from BNC. For the **novels** testing set, 26356 instances of semantic frames were detected, and for the **BNC** testing set, 31807 instances of frames were detected. These instances then act as individual candidates within the respective sets in testing against the members of the diffuse embedded-causal prototype.

The procedure performs a linear ranking of the frame in testing data. Each of these extracted frame is measured in rooted tree similarity to each member of the collection diffuse CF-structure prototypes, where each measure is summed to the total contribution for that candidate frame. This procedure does not determine a specific threshold in the range of scores, and thus does not give a binary decision, but rather produces a embedded causality score for the structure that produced each frame.

For evaluation, as described in Section 5.1.3 given that complex embedded causal structures are likely to be highly sparse and potentially has a longed-tailed distribution, the most sensible method is a sparse *quantile*-based annotation. We annotated three sets of $k = 100+$ (actually about 115 each, because we want to ensure that there are at least 100 determinable samples in each) samples. The annotation of this testing phase is performed by the experimenter. Some very large and deep tree structures can potentially contain more than one causal chain. And the possible labels for each sample is Y (positive, contains some form of embedded causality), N (not causal, or the causal structure does not involve any form of embedding), and U (indeterminable, which usually is because of incomplete sentence, lacking some critical context, or too much ambiguity).

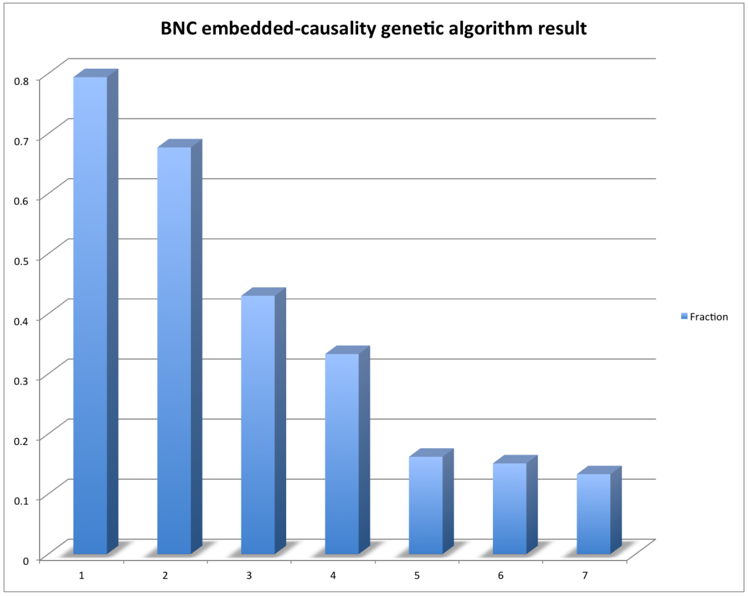
For the **novels** testing set, 26356 instances of semantic frames with some level of embedding were detected, and for the **BNC** testing set, 31807 instances of frames with some level of embedding were detected.

8.4.2 Quantizing ranking

The results from the genetic algorithm are ranked sets of samples; and the bottom ranked part and the parts near the median have very low levels of positives. Thus we know that the result as a fraction of positives in the determined samples must be described by some function that tends to zero near the bottom of each ranking. We explored how quickly the result by annotating the next several quantiles, each with the aforementioned approximately 115 samples to guarantee that each quantile has at least 100 determinable ones.

Since it is very labor intensive, we annotated each until a clear pattern emerges, which happened when there are 6 or 7 quantiles for BNC, and 5 for the novel collection. There are now 805 total annotated samples from the BNC set, and total of 1150 annotated samples from the novels set. The following charts describe quantiles of the BNC and its 7 quantiles:

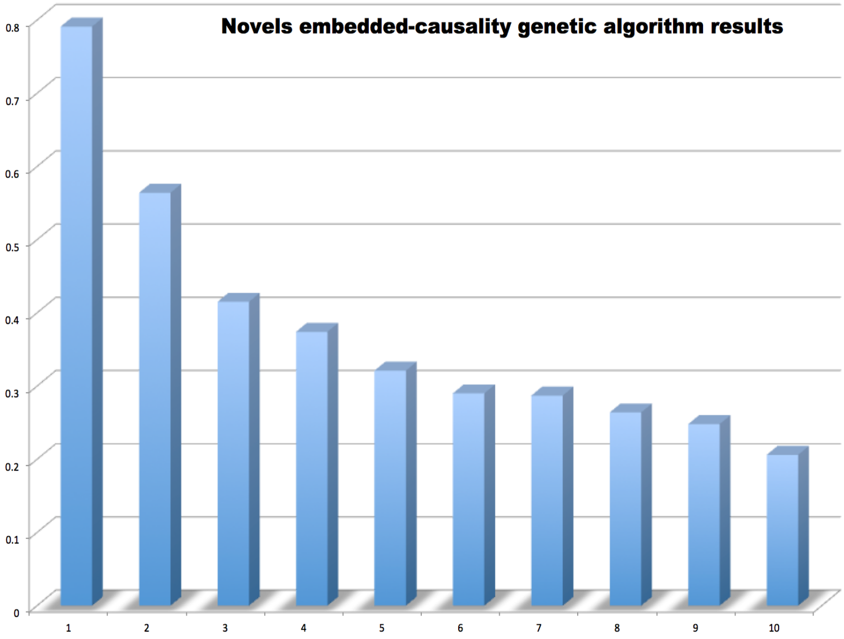
Figure 8.9: BNC results in its top absolute quantiles, each bar is the fraction positive within 115 samples, with at least 100 determinable



We know that the BNC is a mixed dataset coming from multiple genres of sources, some of which include news reporting, formal documentation, or parliamentary proceedings, which have a high emphasis on grammatical correctness and structural complexity. Thus it is a priori likely that the BNC dataset should contain a higher fraction of rank-able results that would be deemed causal and involve some type of syntactic embedding of multiple clauses. This is indeed the case here, the highest several hundred ranked samples have relatively high precision of being embedded causal constructions, dropping off after approximately 15% of the top ranked samples.

The Novels data tells a somewhat different story, as show below with its 10 quantiles: This set belongs to a relatively monolithic genre of modern English language novels, a genre with more potential for character development and plot-lines that cater to related sequences of events. The precision is nearly as high as the BNC data at the top quantile, but drops off relatively quickly, after merely 1% of the top ranked samples.

Figure 8.10: Novels results in its top 10 absolute quantiles, each bar is the fraction positive within 115 samples, with at least 100 determinable



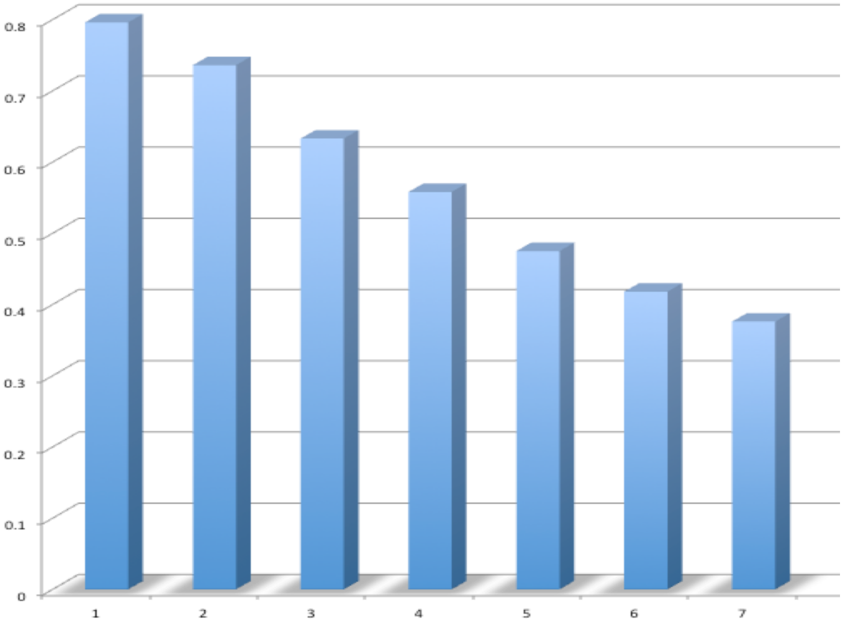
8.4.2.1 Kappa score:

Since the annotation task is highly time intensive, and requires some knowledge of linguistic categories, as well as some training specific to recognizing complex causalities, the second annotator was only able to annotate approximately 100 samples (102 exactly), which were the top ranked 102 samples in our novels corpus output. The second annotator was not a linguist, but a computer scientist with some knowledge of formal language theory, and she was trained over a two hour period on the requirements of the task, as well as on the intricacies of the classifications of causal structures. The guidelines in Appendix I was the main document used in her training. The set of top 102 samples labeled by the second annotator has a precision of 87.3% for the binary classification, which is slightly above the precision of the first annotator for the top quantile. Looking at the agreement between the annotators in the binary classification case, the kappa score (Cohen, 1960) is 0.40, which is acceptable for such a complex cognitive task.

8.4.2.2 Cumulative quantiles:

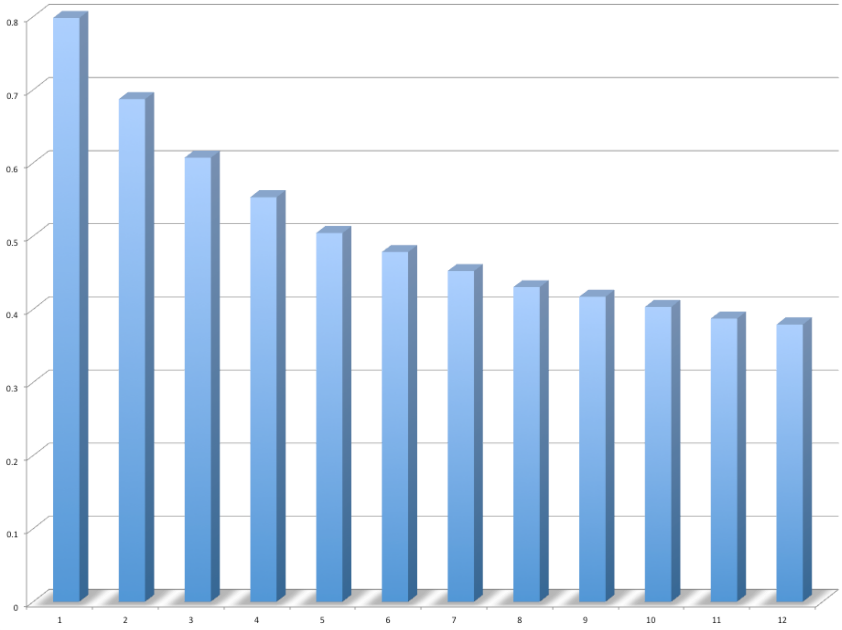
It is also sometimes useful to see the binary discrimination power of an algorithm with respect to a task. So we also show the cumulative quantiles, such that each of the data-points is the precision of the subset of data from the 1^{st} (top) to the k^{th}, where k^{th} is the current quantile in question. This shows the predictable power of each potential boundary, if we turn some division between quantiles k and $k+1$ into the division in a binary classification.

Figure 8.11: **Cumulative** results from BNC, where the k^{th} bar represents cumulative precision from 1^{st} through k^{th} quantile



The BNC set, in terms of cumulative precision is 50% or above until down-to the 7th quantile; while the novels set is above 50% for all cumulative quantiles. So if the task was recast as a binary classification, it would be reasonable to pick a division point between two of the top quantiles for these genres of data-sets.

Figure 8.12: **Cumulative** results from novels, where the k^{th} bar represents cumulative precision from 1^{st} through k^{th} quantile

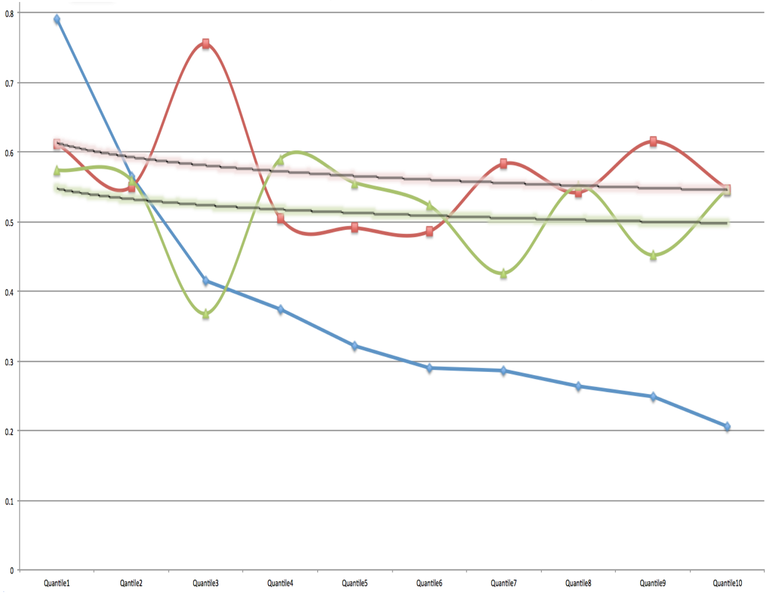


8.4.3 Comparison with baselines

We compared the results of our system to baselines, a textual entailment system as well as an n-gram model; since annotation is highly labor intensive, the annotated data are from the top 10 quantiles of our ranking. Thus these samples are already pre-selected by our system to be relatively likely to be causal; so we mainly test to see is correlation with our system exists, and whether they produce the same gradient of precisions that rank from highest quantile to the lowest among these 1150 samples. For each, we expect some positive correlation with ours; but our system, being more specifically designed for complex causality, should outperform each.

We are unaware of any comparable system for complex causality, so textual-entailment (TE) is the most similar to our task. Thus, we used the TE system VENSES (Delmonte et. al. 2007/2009). This test is not appropriate for the original purpose of VENSES, but is done with our data and annotation to see any correlation to our results, a comparison of the closest system. For any given sample of testing set, we determine whether any pair of the multiple clauses, is identified as entailed by VENSES. We compared the results against our gold standard (for embedded causality). The samples are the top 10 quantiles of the novels data-set (set with the most annotated samples), ranked according to our algorithm. Figure 8.13 contain the TE fraction of each quantile according to VENSES (red), whether VENSES judgment on TE is consistent with our human annotation on causality (green), and our system's output (blue).

Figure 8.13: fractions of TEs according to VENSES, fraction of VENSES Y/N output same as human judgment on causality, and our system; for each of the top 10 quantiles for *novels*. The black lines with shading are the corresponding trend lines



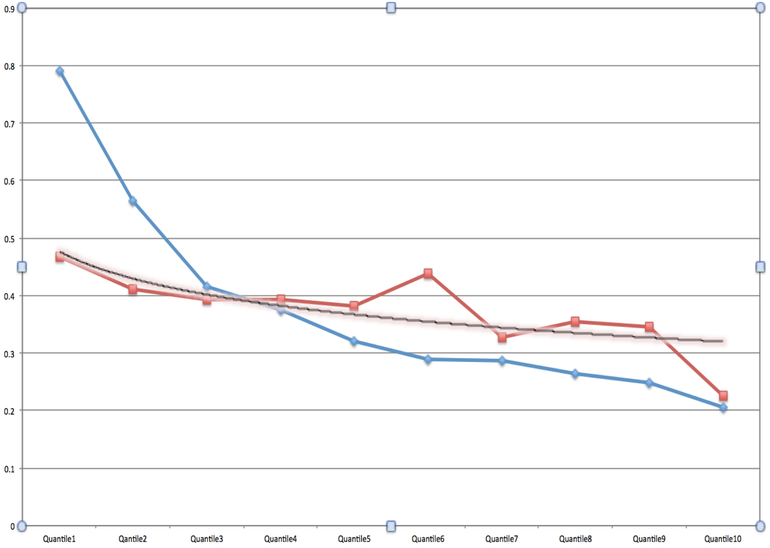
TE results labels contained many false negatives, since it is not designed for causality. This also serves as a baseline for our system, given TE is the closest system available for testing, where our system over-performed significantly given the task of complex causality. There seems to be a spike of those that are ranked by our system in the 3^{rd} quantile, assessed as TE according to VENSES but are not annotated as causal (so in a sense both systems are wrong with respect to human annotation of causality). After examining this subset of samples from the 3^{rd} quantile, we observe many samples with adjacent pairs of clauses where they are speaking on the same topic (which could be construed as TE), but probably not causal in a deep-semantic sense. Some examples include 1) where Mr. Lawrence and Meg are both described as bashful and prim, which together made Jo's contact with him difficult, which definitely falls within the same topic and conversing about one usually entails speaking of the other, but is not causal; 2) describing Collins and Charlotte Lucas were fortunate to find one another; and 3) where the two instances of their attempt to explain the real-estate property entail in the will definitely have logical connection, but are not directly causal.

1. - he looks as if he'd like to know us but he's bashful
   - and Meg is so prim she won't let me speak to him (Mr. Lawrence) when we pass
2. - and I am sure she could not have bestowed her kindness on a more grateful object
   - Mr Collins appears to be very fortunate in his choice of a wife
3. - Jane and Elizabeth tried to explain to her the nature of an entail (of the real-estate)
   - they had often attempted to do it (to explain) before

So while many of these can be construed as TE, the causal relation between them is not clear; these would likely be regarded by some individuals as causal, while others would not see the causal connection. The determination of causality here depends on whether other implied event are perceived; such as in 3) above, where some may perceive the Jane and Elizabeth would likely be frustrated by having to repeatedly explain a real-estate entail. These would most likely be considered *latent causal chains* if they are considered at all causal.

Causal chains are highly sequential structures, so an n-gram model is a reasonable method for comparison. We also produced a standard n-gram model with smoothing and back-off, trained on the same training data as our system. Each sample of multiple clauses/frames is presented a a single sequence of terminal tokens. We determined that a 3-gram model is the optimum to obtain good specificity and avoid over-training. Thus, we tested it against each of the annotated testing samples, and produced a ranked score using the harmonic mean of probability of each token in the sequence according to the 3-gram model. Given that the testing samples are preselected by our system to be top-10 quantile, the n-gram model provides a re-ranking of these. We examined this re-ranking to see whether we get the same differentiation in precision in the new 10-quantiles of the same size after re-ranking (Figure 8.14). Thus the results of our system are also weakly

Figure 8.14: precision in re-ranked quantiles according to n-gram, with trend-line, and original ranked quantiles from our system



correlated with n-gram re-ranking; but our system provides much better Y/N separation of the gold-standard in the trajectory over the top quantiles, and provides a more consistent and monotonic trend.

8.4.4 Per class annotation

During the annotation process, we use a classification system drawn from classical and behavior psychology sources, and further developed by us, to prompt the annotator on what would be considered causal, in order to preserve some uniformity in the semantics (introduced in Section 5.1.2 and detailed in Appendix 9.4.1). Although this was only used as a part of the guideline, we also annotated a small portion of the data, the top quantile of the novels output, to see if any interesting patterns developed with respect to these classes. We labeled the top 150 samples of the novels set, for the presence/absence of each of these 7 classes. Since long causal chains may contain multiple relations of different semantic types in one sequence, a sample may have multiple labels. The number and % of the top 150 ranked samples are ❶ *efficient:* 17, 11.3%; ❷ *necessity:* 36, 24.0%; ❸ *formal:* 42, 28.0%; ❹ *final:* 40, 26.7%; ❺ *inducement:* 44, 29.3%; ❻ *material:* 17, 11.3%; ❼ *latent:* 10, 6.7%;

which has a wide distribution among the 7, and has no particular dominant class. It is unsurprising that *latent causal chain* is contained in the least number of samples, since it is also the most difficult for people to detect.

8.5 Brief Discussion and examples

We discuss some preliminary linguistic observations about the output here, with more detailed linguistic and cognitive analyses of both *adjoined causal* and *embedded causal* results in Chapter 9. The positive samples reflect a variety of different types of embedded causality, with different size in terms of length of the *causal chain.* Here are a pair of of relatively simple positive samples, first from **novels** and second from **BNC**:

1. *there had risen a beautiful, strong friendship, that it blessed them both out of the grave of a boyish passion*
2. .... *his country strives to acquire the far point they need in Poland that it ensures qualification for the world cup next summer.*

e.g. in (1), we can see the direct causal relationship expressed here, where the causality is: $PRED_{rise}(ARG_0 : friendship) \xrightarrow{caus} PRED_{bless}(ARG_0 : friendship,\ ARG_1 : them\{COREF : ....\},\ ARG_{2,to} : grave)$. Next are a pair of moderate complexity, with a non-trivial causal chain (a chain that takes a human significant amount of effort to understand), first from **novels** and second from **BNC**:

a *she decided to make her will, like Aunt March had done, so that if she did fall ill, her possessions might be justly and generously divided ....*

b *forward plan and good communication are the two foundation stones which must be in place to guarantee that people ill with HIV have the good choices and continue to enjoy the very good quality of life wherever they choose to be*

e.g. in (b) we can observe the structure:

$$PRED_{be}(ARG_0 : stones) \xrightarrow{caus} PRED_{be}(ARG_0 : stone, ARG_M in\_place)$$

$$PRED_{have}(ARG_0 : people, ARG_1 : choice) \xrightarrow{caus}$$

$$PRED_{enjoy}(ARG_0 : people, ARG_1 : life)$$

Then are a pair of embedded causal structures that are relatively complex in nature, with longer causal chains, first from **novels** and second from **BNC**:

i *...., I was grateful that the emperor's daughter was so absorbed with her image in the mirror that she didn't notice me slipping my mother's serpent bracelet, wrapped carefully in an old bloody dress under my mattress, I buried it ....*

ii *it can legislate arbitrarily that it deprives a group of citizens of their basic rights or freedom to enlarge its own power at the expense of the local government, to weaken the ability of the media to inform the public or to sap judicial independence.*

e.g. in (ii), it contains multiple predications involved in a series of causal relations, centered on the predicates *legislate*, *deprive*, *enlarge*, *weaken*, *inform*. Here there are two chains that are explicit according to the syntacto-semantic structure, that of $legislate(\cdot) \xrightarrow{caus} deprive(\cdot)$ and $weaken(\cdot) \xrightarrow{caus} NEG(inform(\cdot))$. All of them are linked causally through a set of relations, whether it is a linear chain, or if there are some parallel relations, it is not clear. The syntacto-semantic structure alone does not inform us of the entire causal structure.

Although one of the testing sets comes from the same source as (but does not overlap) the training set, there appears to have little difference; this is potentially due to the fact that BNC itself is an eclectic source with various genera of writings, so it does not offer an advantage to have training and testing sets from that source. The fact that the **novels** corpus contains mostly simple conversational or straightforward narration, and that the **BNC** corpus contains some much more complex structures, can play a role in this. The highly complex and embedded structures or run-on sentence and interfere with the preceding stages of parsing, transformation, and frame extraction, so that the data passed to the embedded-causal module would be much less clean in that case.

8.6 Embedded causal samples

Here is a sample of embedded causal structures found within the first quantile of the ranking by the genetic algorithm informed by diffuse prototyping, from the BNC testing dataset. Since not all causalities are easily assessed without context, some of required an careful reading of the surrounding context, and some required additional real-world information to decide, so the included ones are top-quantiles ones whose causalities are easier to observe in non-contextual forms:

Both the overall surface sequence is shown, as well as the individual links in the causal chain shown as (a, b, c, ...) The basic color-coded labels for frames within embedded structure is as follows, each color denotes the position of the semantic frame in terms of it as a link in the causal chain; here is a selection of causal chains with various types:

- eurotunnel is already in default of its credit agreement with the bank syndicate, [that it] is seeking an extra xx billions on top of the xx billions raise so far .

  eurotunnel is already in default of its credit agreement with the bank syndicate $\xrightarrow{efficient}$ *it is seeking an extra xx billions on top of the xx billions raised so far*

- before the housewives could rest several people called and there was a scramble to get ready to see them (receive them with hospitality)

  several people called [the housewives to visit] $\xrightarrow{efficient}$ *there was a scramble to get ready* $\xrightarrow{purpose}$ *to see them* (here meaning receiving the guests)

- she tries to find high-born women to bear him a son that she can take in as her own

  she tries to find high-born women $\xrightarrow{enables}$ *to bear him a son* $\xrightarrow{enables}$ *she can take in as her own*

- by late afternoon, I (Cleopatra Selene II) joined the rest of the women of the household Lady Octavia took it upon herself to [Lady Octavia] teach me (Cleopatra Selene II) to spin whorl

  I joined the rest of the women of the household $\xrightarrow{constitute}$ *Lady Octavia took it upon herself* $\xrightarrow{purpose}$ *teach me* $\xrightarrow{purpose}$ *spin wool*

- I (Cleopatra Selene II) was a Ptolemy princess (meaning descended from Hellenic-pharonic blood-line), a queen in exile who must bide her time until she could think of some plot, some plan to [some plot/plan] return her to her throne

I was a Ptolemy princess $\xrightarrow{constitute}$ *[I was] a queen in exile* $\xrightarrow{implication}$ *who must bide her time* $\xrightarrow{enables}$ *she could think of some plot, some plan* $\xrightarrow{purpose}$ *return her to her throne*

- one of the guards searched Euphronius he actually put his unclean hands on our wizard's hold person I (Cleopatra Selene II) watched, aghast, trying to ignore the curious motion within the basket an echo of fear that snaked around my heart then the ill-mannered Roman guard approached me and I held my basket out to him hoping he'd reach inside (Counterfactual) hoping that whatever evil spirit lurked there would fly out strike him dead

  one of the guards searched Euphronius $\xrightarrow{efficient}$ *I watched aghast trying to ignore the curious motion within the basket* $\xrightarrow{outcome}$ *the ill-mannered Roman guard approached me* $\xrightarrow{induces}$ *I held my basket out to him* $\xrightarrow{purpose}$ *he'd reach inside* $\xrightarrow{efficient}$ *whatever evil spirit lurked there would fly out* $\xrightarrow{efficient}$ *strike him dead*

Chapter 9

Linguistic Properties of Observed Results

Here we will look at some of the common linguistic properties of causative constructions, introduced in Chapter 6, that we can readily observe in the extracted samples among adjunctive and embedded causals. These properties are selected for the reason that a better understanding in these areas would allow us to perform more accurate extractions in the future.

9.1 Relative tendencies of adjunctive versus embedded type extractions

The *adjunctive* and *embedded* approaches are substantially different in terms of their target causal structures, where one treats frames in the corpus as a single linearly ordered set (with some caveats as discussed in 7.3.4), the other treats the frames with accordance to their relative relations at the syntax-semantic interface representation. Although there is some potential for these to intersect, such as the following example:

- **adjunctive:**

   1. forward planning and good communication are the two foundation stones that must be in place to guarantee that people ill with HIV have good choices
   2. [so that] they enjoy the very good quality of life wherever they choose to be

- **embedded:** forward planing and good communication are the two foundation stones [that] must be in place to guarantee that people ill with HIV have the good choices and continue to enjoy the very good quality of life wherever they choose to be .
   1. forward planning and good communication are the two foundation stones [that] must be in place
   2. to guarantee that people ill with HIV have good choices
   3. [so that they] continue to enjoy the very good quality of life wherever they choose to be

Where both methods located approximately the same causal relation in text, with embedded method providing some more detail for the chain. That not withstanding, the vast majority of the examples extracted by these methods are distinct from one another, and we will examine the major difference between the extractions in the following sections.

9.1.1 Sharing of arguments

There are arguments between different links in the chain that can be identified in extracted samples by both methods, either those that can be machine identified, or those that would take detailed and robust detection of co-reference (highly reliable automatic co-reference resolution for all types of data does not yet exist). We observe that *adjunctive* causal forms $\langle e_i, e_{i+1} \rangle$ generally have a strong tendency where the external arguments of both e_i, e_{i+1} are the same entity. On the other hand, the *embedded* causal forms $e_1 \xrightarrow{caus} e_2$ do not have this tendency, but rather often occurs that the external argument for e_2 is identified with one internal argument of e_1, but this tendency is not as dominant as the tendency among *adjunctive* forms. The contrast can be exemplified by:

1. **adjunctive:**
   (a) time had appeased her (Aunt March) wrath
   (b) [time] made her repent her vows (March stated that she would not give money to the couple if Meg married Brooke)

2. **adjunctive:**
    - (a) an acetlink person, possibly an existing volunteer, would keep their church informed about our work
    - (b) [that acetlink person] would encourage people to consider becoming volunteers helping with fundraising ideas
3. **adjunctive:**
    - (a) Jo took Beth down to the quiet place, where she could live much in the open air
    - (b) [Jo] let the fresh see breeze blow a little color into her pale cheeks
4. **embedded:**
    - (a) the next stage of our preparation will be a leaflet
    - (b) [the leaflet] to be used during the election campaign
5. **embedded:**
    - (a) the success of the initial pilot program has been recognized by the ministry of health and the institute of health education
    - (b) [the success ...] led to maurice, kate, and ana meeting with government officials and representatives from unicef ....
6. **embedded:**
    - (a) the practical development of its work as taking place mainly at local levels, [it] encourages the formation of autonomous local groups
    - (b) which (the practical development) can respond to need in [their] own areas of the IRS

For the examples (1-3), the external arguments of *'time'*, *'an acetlink person'*, and *'Jo'*, all reside in the external position at LF, sometimes playing the same $\theta-$role in both (as *force* in (1)), while other times having different roles (as *expressor* and *trigger* in (3)). While for examples (4-6), there is an argument that is shared between the two causal chain links, such that $Entity_x \rightarrow Entity_y \rightarrow Entity_z$, for instance, *'a leaflet'* would be $Entity_y$ in (4). Note that in (5), the argument *'the success of the initial pilot program'* is actually an internal argument at syntax-semantics interface, but due to passivization appears to be in the external position at the surface. The ability of embedded causals to form a chain such that the internal argument of one link becomes the external for the next, requires that any non-final link in the change must be at least dyadic in terms of the essential argument list. This difference is in part inter-related to the fact that there are far more parallel and coordinate structures that are extracted by the adjunctive method (as expected); but it doesn't tell us whether the tendency for coordinate structure is caused by adjunctive causal structure at the semantic level, or vice versa. This could be an avenue for further investigation.

In addition, there is a substantial amount of embedded causal structures where there is no shared argument between the neighboring links of the causal chain. Or similarly, only some non-essential argument (*temporal*, *manner*, etc) from the first link in the chain would be shared with the second causal link. We can see that in the following: the more stubbornly East Germany's old men resist the changes, greater becomes the danger that the whole edifice they have constructed will collapse around them, threatening stability in europe and raising the question of reunification for which no one is prepared

a the more stubbornly East Germany's old men resist the changes

b greater become the danger that

c the whole edifice they have constructed will collapse around them

d i threatening stability in Europe

   ii raising the question of re-unification (with B.R.D.)

Here we see that (a) does not share any argument with (b), which is a monadic frame about some consequence of resistance to change. The relationship between (b) and (c) is similar, with the latter addressing the consequence of *'danger'* in metaphorical terms. The relationship between (c) and (di) / (dii) is slightly more complex, where the frames in (4) are derived from a gerundic form, and have an external argument place that is should be filled by not the immediately preceding entity *'them'*, but rather the entire event described by frame (c). This additional level of complexity, added by the lack of any apparent argument sharing, makes any further automatic analysis of the causal chain more problematic, meaning no apparent connection exists between neighboring links in embedded causality outside of the indication by the embedded-causal structure itself. So in these cases, no addition of scoring help using argument similarity, even with augmentation of nominal similarity measure, would improve the results of embedded causal extraction. So in these cases, only the overall structure of the embedded causalities as represented in language, as in our algorithm, would be able to detect the causal chains (with possible exception of wide discourse-context or knowledge-base methods, depending on the type of dataset).

9.1.2 Temporal sequence and separation

Although exact times of occurrence are rarely indicated in the samples, a careful human consideration of the sample could usually gives some perception of the likely time-separation among real-world events represented by links. From observation of the top examples extracted, through analysis, the *adjunctive* method seems have a tendency to extract pairs of events that are relatively close in time, whereas *embedded* method seems to be able to extract chains that are widely separated in time. Here are some typical examples from extracted adjunctive causality.

1. (a) Jo took Beth down to the quiet place, where she could live much in the open air
   (b) [Jo] let the fresh see breeze blow a little color into her pale cheeks
2. (a) discordant din from those in the car had been audible for some time
   (b) [discordant din added to the violent confusion of the scene]
3. (a) the man peered doubtfully into the basket, plunged in his hand
   (b) [the man] drew one up (a dog up from the basket)
4. (a) consumers reject inferior products (while discussing economic theory)
   (b) [consumers] make them unsellable

Some examples of the adjunctive set, such as (1) or (3) above, where the second event is an action that necessarily follows the first with little to no intervening temporal separation, such as the man plunging his hand into the basket to draw up a dog. These types seem to have a strong tendency for e_2 to be a purpose of e_1, and that the actor of e_1 having volition. Other examples, such as (2) and (4) show e_1 to be some state of the *theme*/*phenomenon*/subject matter of e_1, or some persistent action on the part of the actor of e_1, where e_2 is a natural consequence concomitant with e_1's occurrence. This type has a tendency to have e_1 as the occurrence of natural phenomenon or some generalization of human behavior, rather than the action of some specific individual. The samples from the embedded procedure are like the following:

1. (a) eurotunnel is already in default of its credit agreement with the banks,
    (b) [it] has in effect been given until the end of the year to settle its differences with the contractors
    (c) to permit a viable financing strategy to be put in place
2. (a) the next stage of our election preparation will be a leaflet
    (b) to be used during the election campaign itself
    (c) [t] entitling [you to] your vote count
3. (a) the cutting of equivalent to a p reduction in the basic rate of income tax
    (b) to be replenishing some of the consumer depleted ammunition
4. (a) smurfit is looking to Europe as its main engine for growth in the months to come
    (b) it must decide how
    (c) to spend the xx billion cash raised from the recent financial restructuring with morgan stanley

For all of these instances, there is a perceived time lapse within the duration of the set of events represented by the causal chain. The determination requires a significant amount of real world knowledge, about things such as *credit in financing*, *the length and stages of the electoral process*, *the taxation system*, and *the process that leads from capital expenditure to economic growth*. As such, (1b) gives an explicit piece of temporal information of *'until the end of the year'* to support a temporal separation in the real-world, while the rest requires knowledge and a system of deep pragmatics to arrive at that conclusion.

The difference in likely temporal separation for these two extraction methods likely relates how causal chains are manifested in linguistic constructions, as discussed in Section 6.2.5. This is readily explained, if we presume that adjunctive and embedded causality have real psycholinguistic preceptive impact on listeners. As discussed there, there is a strong tendency for embedded causal structures to represent temporal separation by providing more complex structure between the matrix clause containing the original *cause* and the final embedded frame of the chain. One may view this as embedded forms having more inter-frame "structural resources", more elements and structures (demonstrated by the array of sub-structures extracted to form the diffuse prototype) to provide cues of the length of the chains, and indirectly provide the hints for temporal separation. Adjunctive causal structures, being perceived linearly, lack an equivalent mechanism to do so, although it could (as could adjunctive causality) still use lexical semantics with associated real-world knowledge to do so.

9.1.3 Expression of purpose

Extracted causalities can also be perceptually assessed to see whether the relation is incidental or with pre-purpose in e_1. The causal structures from the adjunctive procedure are observed to contain more incidentally causal, whereas those from the embedded procedure contain more that are purposeful, this is partially related to the entity property discussed in Section 9.3.2 in terms of the role of the external argument entity in e_1. The following are samples from embedded causality that demonstrate this property:

1. (a) it is asking readers to say if they have been misrepresented in a program
   (b) they expose the tricks and deceptions used by tv
   (c) to deliberately mislead people
2. (a) these will all be present to interested donors in a common summarized format
   (b) [which] will provide them with the information they are looking for
   (c) to make their decisions
3. (a) the cutting of equivalent to a p reduction in the basic rate of income tax
   (b) to be replenishing some of the consumer depleted ammunition

All of the above are perceived to be purposeful in terms of the e_1 taking place to bring about the rest of the chain, each uses different means to convey that purpose. In (1) it is done primarily through the action of *'asking'* the readers to perform the subsequent task, which predisposes us to perceive that original intention is there in e_1. For (2), the second frame $\phi_{provide}(X, donor, information)$ strongly biases us to perceive that the information is related to the decision process stated by the final frame, which then allows us to conclude that the e_1 was purposeful. (3) is seen as purposeful, but it requires some form of knowledge of how tax rate and economic growth are related, as well as the correct interpretation of the metaphor *'ammunition'* in this context, in addition to the structure information. The situation among the adjunctive causal set is different, typified by the following:

1. (a) this type of display is fool-proof
   (b) [this type of display] takes only minutes to arrange (blooms on souffle dish)
2. (a) i became close to friend of a friend
   (b) [i] gradually fell in love with her
3. (a) [Laurie] to lose heart at the first failure
   (b) [Laurie] to shut himself up in moody indifference
4. (a) the hundreds (Gatsby's acquaintances) who had accepted his hospitality
   (b) [Gatsby's acquaintances] so became authorities on his past

These show the prevalent types of examples from the adjunctive set, and are generally perceived not to have any purpose in e_1 for the occurrence of e_2. Many examples occur with the first argument of the e_1 frame being *theme*, *experiencer*, *force*, *event*, *trigger*, and other $\theta-$roles that are inherently inanimate, or are cast with $[-m]$, so that the possibility of a purpose is precluded, such as (1) (*theme*) and (3) (*experiencer*) above. There are other instances where more complex pragmatics dictate that there is no purpose in the chain, such as in (2) where a combination of the knowledge of how people normally *'fall in love'* along with the implicature conditions, where e_1 failed to provide any information contrary to the normal cause of *'falling in love'*.

9.2 Causal chains structure

Complex chains of causation often present interesting internal structure to express complex thought processes of the speaker. The structure of each causal chain ultimately contributes to the social network representation in its topology, whether the chain is simple linear in structure, or has more complex branching structures. As previously mentioned, the embedded causal detection is only with regard to whether some complex embedded structure contains a set of causal relation or not. When the internal logical structure is linear, it is relatively easy to infer the individual pair-wise causal relations as in a ordered linear sequence; but when that is not the case, more complex analysis would be required to come to the correct individual causal relations contained within the larger embedded causal structure.

9.2.1 Linearly arranged chain causality

Most of the longer (length > 2) causal chains extracted with embedded module contain a single chain that can be linearly arranged. Many samples, especially from embedded causals, do not follow a standard structure of $entity_1 \xrightarrow{\phi_1} entity_2 \xrightarrow{\phi_2} entity_3 \xrightarrow{\phi_3} entity_4$ or $cause_1 \xrightarrow{\phi_1} (cause_2 \xrightarrow{\phi_2} (cause_3 \xrightarrow{\phi_3} effect))$. Rather, some exhibit frame-structural patterns that are unexpected and non-uniform, and yet still produces a semantically relevant causal chain through all the involved frames, such as the following:

a the more stubbornly East Germany's old men resist the changes

b greater become the danger that

c the whole edifice they have constructed will collapse around them

d (a) threatening stability in Europe

  (b) raising the question of re-unification (with B.R.D.)

Here, we represent an abbreviated version of frame structures, using $\langle\rangle$ to represent coordination, including the most important attribute(s) of each argument and predication as subscripts, and covert forms as $[\cdot]$. We can see that the most likely frame-based chain structure of this form comes to:

$$men^1_{old} \xrightarrow{resist} changes^2(danger^3 \xrightarrow{BECOME_{greater}}$$

$$(edifice^4_{constructed} \xrightarrow{collapse})(\langle[collapse^5_{edifice}]$$

$$\xrightarrow{threaten} stability^{6a}_{europe} \wedge [collapse^5_{edifice}] \xrightarrow{raise} question^{6b}_{reunification}\rangle\ )$$

We can see that first, there are several different adicity structures for the frames involved, including monadic (*BECOME* and *collapse*) and dyadic (*resist* and *raise*). The entire complex of the final three frames acts as an adjunct of *old_men* $\xrightarrow{resist}$ *change*, which is similar to the average adjunctive causal structure, the third frame likewise is an adjunct to the second of *danger* $\xrightarrow{BECOME_{greater}}$, which contains the next frame $\xrightarrow{collapse}$ as a *goal.* The third frame $\xrightarrow{collapse}$ at the same time behaves as the *event* and *expressor* $\theta-$roles of the final coordinated frame with two predicates *'threaten'* and *'raise'.* The final frame $((\xrightarrow{collapse}) \xrightarrow{collpase}) \xrightarrow{threaten \wedge raise}$ .... . The final frame at the same time represents a conjunction of two events, thus is a composite of two frames that shares the same external argument, which produces two purported final outcomes of the series of events. Thus there is a convoluted structure that from a frame-theoretic point of view is in no way linear, but it ultimately results in a semantic interpretation of a single sequence of events. Similar phenomena can be observed in the following:

a the success of the initial pilot program has been recognized by the ministry of health and the institute of health education

b led to maurice, kate, and ana meeting with government officials and representattives from unicef

c to establish a long term training program in five romanian regions

d [the long term training programs] starting in November of this year

Here the most likely frame-based chain structure comes to

$$(\langle ministry^{1a}_{health} \wedge institute^{1b}_{HE} \rangle \xrightarrow{recognize} success^{2}_{program})$$

$$\xrightarrow{lead\_to} (\langle maurice^{3a}, kate^{3b}, anna^{3c} \rangle \xrightarrow{meet} official^{4}_{government}$$

$$([official^{4}_{government}] \xrightarrow{establish} program^{5}_{training}$$

$$([program^{5}_{training}] \xrightarrow{start})))$$

The chain's initial entity is a compound entity, being a conjunction of $ministry \wedge institute$, is the experiencer in a perceptual event that precipitate the following chain. The initial $\xrightarrow{recognize}$ event is in turn the external argument of the *manner-of-causation*-type predicate $\xrightarrow{lead\_to}$; and the entire initial $\xrightarrow{recognize}$ event is in the *cause* role. The remaining chain is composed of three frames, the first of which $\xrightarrow{meet}$ has the second $\xrightarrow{establish}$ as a *purpose* adjunct, which has a covert *agent* argument. The *xlongrightarrowestablish* frame itself has an adjunct frame of $\xrightarrow{start}$, which has a covert *phenomenon* argument. Some of the linearly arranged chains can be of considerable length, but seems to have a limit of four or five (with one or two exceptions) separate frames in the chain, depending on the corpus, such as the following:

[Amy] feeling that the neighbors were interested in her movements she wished to efface the memory of yesterday's failure by a grand success today so she ordered the 'cherry bounce' (a strong drink made with whiskey, cherries, and lemon juice) and drove away in state (to achieve a state of mind) to [Amy] meet and escort her guests to the banquet

1. [Amy] feeling that the neighbors were interested in her movements
2. she wished to efface the memory of yesterday's failure by a grand success today
3. she ordered the 'cherry bounce'
4. drove away in state
5. [Amy] meet and escort her guests to the banquet

The lengths of chains and its limits can be an artifact of the type and size of the dataset. But alternatively, it may well be due to some cognitive processing capability that speakers use to logically analyze long chains, beyond which the causal structure becomes difficult for the average user to see. This is an area that can potentially use larger dataset and more psycholinguistic research.

9.2.2 Out of order chains

As stated previously, there are some of the exceptions to the surface sequential order corresponding to the logical sequence of causation. As in the following:

Brandon Ormsby's transfer from Leeds United to Cardiff City was called off yesterday, after the defender decided he does not want to play in the third division

1. the defender (Ormsby) decided

2. he does not want to play in the third division
3. Brandon Ormsby's transfer from Leeds United to Cardiff City was called off yesterday

Here, there is a sequence from a frame perspective such that:

$$Entity_x \xrightarrow{call\_off} transfer_{Brandon\_Ormsby}$$

$$(Brandon\_Ormsby \xrightarrow{decided} (Brandon\_Ormsby \xrightarrow{\neg want}$$

$$(Brandon\_Ormsby \xrightarrow{play\_in} division_{third}\ )\ )\ )$$

The sequentially first frame occurs as an adjunct to the remainder of the chain. While the remainder (the non-initial links) in the chain occur in logical order, the initial adjunct frame is actually the logical final outcome of the causal chain (the embedded frame $\xrightarrow{play}$ is not a really a part of the chain, as it does not express an event that occurs in the immediate temporal sequence, but the desire to play is). Any time when we have adjunct structures in embedded causals, there is a significant chance that the adjunct's linear sequential position may not correspond to the causal semantics at a deep level. There are often, but not always clues that they adjunctions are out of order with respect to the deep semantic structure, such as the temporal adverbial *'after'* which makes the causal chain in the exact direction of the surface structure highly improbable. These misalignments between two levels of semantic structures awaits better temporal inference methods to be completely resolved, as temporal precedence is the final arbiter in the directionality of a causal structure. Something similar can be observed in the example below, where the preposition *'for'*, as an indication of causal link, allows for the rearrangement of ordering of the logically final frame to be first in surface sequence:

Beth smiled and felt comforted for the tiny thing (a little grey coated sandbird that looked at her with friendly eyes) seemed to offer its small friendship and [the little grey sandbird seemed to] remind her [Beth] that a pleasant world was still to be enjoyed

1. the tiny thing (a sandbird) seemed to offer its small friendship
2. [the sandbird] remind her that a pleasant world was still to be enjoyed

3. Beth smiled and felt comforted

The *'for'* in this case puts the remainder of the structure (frames 2 and 3) into a *reason* role with respect to the first frame of *'Beth smiled and felt comforted'*, which allows it to express the causal relation in the opposite direction of the surface order.

As we observe the chain of causation, especially in deeply embedded causal structures, expresses the connections between chain links through a large variety of ways of differing complexities, at the frame-theoretic representation. The $\theta-$roles of each frame can also be filled with a variety of lexico-syntactic structures, ranging from standard nominals to entire complex frames which may itself contain internal structures. However, in a majority of the cases, the semantics of the chain of causation can be generally analyzed in a linear sequence from the initial to the final frame, even if there are at times confluences and branching of this linear structure. So we see that there are at least two separate levels of structures in these complex causal constructions, one at the frame-theoretic level, while another at a level closer to the real-world referent as the causal chain itself.

9.2.3 Complex topology logical structures

There are certain embedded causal structures that go beyond representing linear chains of causality, which is a small portion of the extracted causals. (So tendencies seen here are from observation on very small sample set, and should only preliminary, and await larger datasets to confirm.) These involve a chain-like structure in which some link in the chain consisting of multiple events (frames), and thus would be better described as a *causal lattice* rather than chain. The most typical position for this bifurcated form of causal structure occurs at or near then end of a causal sequence, this occurs with or without the surface sequence corresponding to the order in logical causal structure, such as the following:

1. prone upon the floor lay Mr. March with his repeatable legs in the air likewise prone was Demi [Demi] trying to imitate the attitude with his own short, scarlet-stockinged legs, both grovelers so seriously absorbed that they were unconscious of spectators till Bhaer laughed his sonorous laugh, and Jo cried out, with a scandalized face ....
   (a) lay Mr. March with his respectable legs in the air

   (b) [Demi] trying to imitate the attitude with his own short, scarlet-stockinged legs

   (c) both grovelers so seriously absorbed that they were unconscious of spectators

   (d) i. Baher laughed his sonorous laugh

      ii. Jo cried out

2. so it was uncomfortable to see the stoicism with which Agrippa and Marcella now approached the altar Marcella held her lips tight [Marcella's lips are] sewn up, fastened down I (Cleopatra Selene II) wonder whether she had ever breathed a day in her life or whether she'd always been that girl arguing with me while we decorated boughs for the Saturnalia

   (a) Agrippa and Marcella now approached the altar

   (b) Marcella held her lips tight

   (c) [Marcella's lips are] sewn up, fastened down

   (d) i. I (Cleopatra Selene II) wonder whether she had ever breathed a day in her life

      ii. I wonder whether she had always been that girl arguing with me ....

   (e) it was uncomfortable to see the stoicism with which ....

The first example above describes a chain of events that ends in a conjoined event that is composed of two separate actions by two entities, representing two distinct reactions of the same phenomenon (Mr. March & Demi on the floor) of two independent observers. The second example describes a causal lattice that has a bifurcation into a disjoint pair of situations near the final link, where the situation described by the pair triggers the final link in the causal structure. (The verb *'wonder'* and the WH-elem *'whether'* seem to be distributed between the two frames of this pair, and does not appear to mean only one of the pair can be true, but rather they are independent.) These allow the surface structure to express more complex causal relationship among events and states of entities, but does introduce additional complexity in the analysis of a causal lattice. Like the out-of-order causal chains, this subtype prevents a simple analysis of the frames expressed in the surface form of a embedded causal structure to be always analyzed as a linear causal chain, and necessitates a more complex mechanism for ordering the causal relation in the lattice. There are very few examples of structurally more complex lattices, such as expressed below:

I (Mrs. March) had all my girls to comfort me at home and his (Mr. Lawrence's) son was waiting, miles away, to say goodbye to him, perhaps I felt so rich, [I felt] so happy thinking of my blessings that I made him a bundle, [I] gave him some money, and [I] thanked him heartily for the lesson he had taught me

1. (a) I (Mrs. March) had all my girls to comfort me at home
   (b) i. his (Mr. Lawrence's) son was waiting
       ii. [Mr. Lawrence's son was] miles away
       iii. [Mr. Lawrence's son] say goodbye to him (hypothetical)
2. (a) I (Mrs. March) felt so rich
   (b) [Mrs. March felt] so happy thinking of my blessings
3. (a) I (Mrs. March) made a bundle
   (b) I (Mrs. March) gave him (Mr. Lawrence) some money
   (c) I (Mrs. March) thanked him (Mr. Lawrence) heartily for the lesson he had taught me

Here, at each link, there are multiple events that can be interpreted as causally influencing any of the following events. In one instance the *'Mrs. March felt so rich'*, *'Mrs. March felt so happy ...'* pair itself could be interpreted to contain a causal relation, in addition to the causal relations with events up and down stream in the lattice. Even more curious forms exist, such as one of the longest chains detected in the novels corpus, below:

one of the guards searched Euphronius he actually put his unclean hands on our wizard's hold person I (Cleopatra Selene II) watched, aghast, trying to ignore the curious motion within the basket an echo of fear that snaked around my heart then the ill-mannered Roman guard approached me and I (Cleopatra Selene II) held my basket out to him [Cleopatra Selene II] hoping he'd reach inside (Counterfactual causal) hoping that whatever evil spirit lurked there would fly out [evil spirit] strike him dead

1. one of the guards searched Euphronius
2. (a) I (Cleopatra Selene II) watched
   (b) [Cleopatra Selene II was] aghast
   (c) [Cleopatra Selene II was] trying to ignore the curious motion within the basket
3. the ill-mannered Roman guard approached me

4. I (Cleopatra Selene II) held my basket out to him
5. he'd reach inside (Counterfactual causal)
6. whatever evil spirit lurked there would fly out
7. [evil spirit] strike him dead

For most of the length, the causal lattice here appears to be a straightforward chain, such as in the latter part, where the Cleopatra Selene wished that the soldier would reach inside, and the creature in the basket flies out, so that it could kill the soldier that would be searching her at the check point. It is not clear whether her standing there watching with an *'aghast'* demeanor attracted the attention of the guard to come over to examine her. There is also the potential of causality within that parallel portion between the event of her *'watching'* and her *'being aghast'*, her observation causes angst of what could happen to herself. That aside, the part of the conjoined event *'she trying to ignore the curious motion within the basket'* lets us know that she detects the presence of the creature inside, and this is a necessary precondition for the remainder of the causal chain, which presupposes that she knows the creature is there and senses that it might be dangerous to those who approach the basket. That fact, along with the event of the Roman guard approaching her precipitated her thoughts on potentially killing the guard with the creature in the basket. If there is a significant number of such structures, it would post quite a challenge to correctly identifying all of the individual pairwise causal relations expressed by the surface form.

9.3 Argument characteristics

Certain tendencies in the traits of arguments and the corresponding entities can be observed in the positive samples extracted by *adjunctive* and *embedded* causalities. These are intimately related to the arguments' roles in the respective frames, and in turn related to their functions in the causal chain, and in part determines the class of the causal relation.

9.3.1 Causal $\theta-$roles

The entities involved in the causal constructions exhibit a variety of $\theta-$roles within their respective frames. There is an obvious difference in the likelihood of certain types of roles within the causal pair/chain, some differentiation also exists between the roles present in adjunctive and embedded constructions. As discussed earlier, there are certain $\theta-$roles that have a tendency to appear in the non-final frame's external position, such as *cause*, *force*, *origin*, *expressor*, *source*; certain $\theta-$roles often appear in an intermediate position, the final frame's external position, such as *trigger*, *constructum*, *destructum*, *theme*; some roles that could appear in both the initial and intermediary position, such as *agent*, *experiencer*, *trigger*, *benefactor*; and others that most likely appear as an internal argument of the final frame, such as *patient*, *disposition*, *phenomenon*, *benefactive*, *goal*, *recipient*, etc.

These are just general trends, there is a great deal of variability for each of these tendencies depending on genre. The more interesting case are where a $\theta-$role in some frame is occupied by a frame that is not syntactically subordinate, such as the following:

1. (a) advance in the field of cosmetics means that, today, superfluous ingredients and allergens can be identified and substituted
   (b) this is great news indeed for anyone with sensitive skin
2. (a) exchange rate [is] held up artificially high by interest rate
   (b) [it] prejudices exports and encourages imports
3. (a) you can be more extravagant
   (b) you can go for more dramatic, impressive displays for special occasions
4. (a) when the nazis took power, he went underground
   (b) [he] worked in the communist resistance, was arrested

For (1) and (2), the first of two frames (in case of (1), the initial frame is itself complex, with an embedded frame $\xrightarrow{identify}$ as a *subject matter* argument within the $\xrightarrow{mean}$) becomes a $\theta-$role in the second in the sequence. For (1), the $\xrightarrow{mean}$ frame is a piece of information that becomes the *theme* in the predicate copula frame, mediated through the demonstrative co-relative *'this'*. For (2), the $\xrightarrow{hold\_up}$ frame is a event in the financial world that behaves as a *phenomenon* role in the $\xrightarrow{prejudice}$ frame, mediated by the coindexed pronoun. For (3) and (4), the second of the two frames becomes a $\theta-$role in the first frame in the sequence (example (4), the initial frame $he \xrightarrow{go} underground$ has its own adjunct frame $nazis \xrightarrow{take} power$). In (3), the $you \xrightarrow{go} for_{display}$ in some manner is semantically identified with being *'extravagant'*, and is an *attribute* argument for the copula frame; and in (4), the $\xrightarrow{work}$ is the definition for being *'underground'*, and could act as a *manner* argument for the first frame and yield the same meaning, as in *'when the nazis took power, he went to work in the communist resistance'*. We can see that in causal sequences, individual frames can behave with much versatility, and fill a variety of $\theta-$roles in other frames. Even when the structures at the syntacto-semantic interface is non-hierarchical, a hierarchy among these frames is often built through some indexical mechanism that allows the speakers to perceive the deep causal semantics.

9.3.2 Entity volition and intention

We observe from different types of surface forms expressing causal relations, there is a difference in the efficacy of each in presenting some level of volition and intention in effecting the causality. The ability for embedded causal structures, through the manipulation of the length of the expressed causal chain in the number of frames occupied, to convey the likelihood of volition is already aforementioned. In most examples of the adjunctive causal types, we observe that either there is no entity capable of volition in non-final frames of the causal chain, or that such individuals had no intention in bringing about the final outcome.

There are a small number of examples of adjunctive causal structures that do convey the causal relation being intentionally put into effect by a clearly stated volitional entity. These are the exception rather than the norm, such as the following:

1. (a) **my husband** left me
   (b) [he] went back to his ex-wife
2. (a) **the government** has shied away from forcing unions to discipline members
   (b) [it] has put proposals to curb strikes in essential services on the back burner

Here the bolded argument are the entities that likely carry volition in their initial action to start the chain of causation. But many examples of the adjunctive causals that have been extracted are those where no specific entity with intention is present in some earlier event that brought about the occurrence of the latter. The adjunctive causals have a tendency to express *material cause* or *efficient cause* as mentioned in the introduction. Some examples are in the following:

1. (a) advance in the field of cosmetics means that, today, superfluous ingredients and allergens can be identified and substituted
   (b) this is great news indeed for anyone with sensitive skin
2. (a) granting of relief to an applicant who had delay would cause substantial prejudice or hardship to some person
   (b) it would be detrimental to good administration
3. (a) yesterday barklays bank announced an increase in its base rate
   (b) it is quickly followed by the other clearing banks
4. (a) he is trapped in a body [that] is severely disabled
   (b) [he] will spend the rest of his life in a wheel chair
5. (a) it is very low in mineral, it suits a wide range of people
   (b) it is ideal for families with young children (this refers to spring water from Britain)
6. (a) the wicker screen is ideal for dividing your lounge
   (b) [it] conceals unwanted clutter
7. (a) your boyfriend is finding it hard to come to terms with the prospect of fatherhood
   (b) [he] is taking his resentment out on you
8. (a) your son resents his step-brother for taking up your time
   (b) [he] is worried that you do not love him any more

Some of the examples above, such as (1), (4), (5), (6), do not have any expressed *cause* or *agent* that is capable of any intention, given that no animate argument exists for the first event e_i. Thus these causal relations cannot be construed as having volition between e_i and e_{i+1}. Examples such as (3), (7), (8) have an entity capable of intentionality in e_i, but all of them clearly have no intention of precipitating e_{i+1} through their participation in e_i. (3) is a likely case of unintended consequences, while (7) and (8) have to do with emotional states of the entities, and the individuals involved typically do not have volitional control over such properties of these entities. (2) has an implied entity that performs the action of *'grant'* in e_i, which could well be an animate being, but the volition is precluded for e_{i+1} unless the unexpressed entity is favor of *'detriment to good administration'*.

In all of these cases, we see that event though there is a temporal sequence and adjacency in each pair, any participant in e_1 is not immediately aware of the occurrence of e_2. It may be due to the adjoined and paralleled nature of these pairs, where any logical notion resembling hierarchy is unlikely to be expressed by such. There are other samples where the volitional nature of any entity in e_i is possible, but there is no indication to that effect, such as in the following:

1. (a) the design is open ended
   (b) it could be adapted to separate passengers from their cars
2. (a) forward planning and good communication are the two foundation stones that must be in place to guarantee that people ill with HIV have good choices
   (b) [so that] they enjoy the very good quality of life wherever they choose to be
3. (a) the vehicles will be given to the national association of boys' club
   (b) [they] will tour rundown inner-city areas

For the embedded causal structures, it is easier to convey volition of an entity in the non-final frame in the chain. These have a tendency to express *formal cause* or *teleology* as mentioned in the introduction. We can look some examples with a range of different lengths of chains of causation, as in the following:

1. (a) **the management team** led by peter jansen originally put up xx percent of the m-equity elements of the buyout
   (b) [which] have seen its investment increase x fold at the mb group offer price
2. (a) **shelford** is selling to rugby club in england and wales by marketing agents
   (b) he is available for almost any kind of promotional activity
3. (a) **akia maria** recorded that his fledgling company was enabled (by akia maria)

   - (b) to buy a license from western electric
   - (c) to develop transistor technology

4. (a) **we** have a very full agenda for our scheduled meeting on October xx
   - (b) it is decided that
   - (c) we meet on the xxth with cricket as the sole topic of discussion
5. (a) **it** (the legislature) can legislate arbitrarily
   - (b) to deprive groups of citizens of their basic rights or freedom
   - (c) to enlarge its own power at the expense of local governments
   - (d) i. to weaken the ability of the media to inform the public
     - ii. to sap judicial independence
6. (a) the success of the initial pilot program has been recognized by **the ministry of health** and **the institute of health education**
   - (b) led to maurice, kate, and ana meeting with government officials and representattives from unicef
   - (c) to establish a long term training program in five romanian regions
   - (d) [training programs] starting in November of this year

Each of the likely volitional entity is bolded in the examples. All of the above structures, to some degree, conveyed volition on the part of one of the entities in a non-final frame of the sequence. Some substructures such as *'... in order to ...'* automatically indicates the likelihood of volition being involved, because of the explicit *purpose* role of the embedded frame. None of the above examples has such explicitness built in, but many of the SC-structures (... *to* $[_{VP}INF$ ....]), as well as the presence of certain lexical verbs like *'lead'*, *'decide'*, *'enable'*, also seem to aid the perception that these activities and resulting causal chains are intentional.

9.4 Semantic characterization

By *semantic characterization* of causality, we mean how individual language users view causal expressions, and how likely they are to judge a complex expression including two or more events as having a causal relation in between. This task of characterization in a large measure requires a classification scheme that sharpens what individual mean by *'causality'*, each category contributing a unique property that would be identified as causal; as such, we do not need a classification system where all sub-classes are mutually disjoint, so here they may overlap or refer to potentially orthogonal properties. Here, the paramount concern is the individuals' perception of relation between two real(or hypothetical)-world events (and thus how they are likely to semantically annotate it), which is orthogonal to the definition and classification of causative constructions in Section 6.1, which are classes based on the expressions' morphosyntactic structures, and should provide an intuitive way of identifying that semantic type.

The semantic classification we adopt here also have a direct impact on the types of causal expressions we are able to extract for *embedded causality* (Section 8.3), since the training data would originally need to be annotated with a specific definition in mind that includes certain semantic classes of relations while excluding others in a semantic definition of causality. Moreover, among the users of a language, there is a great variety of personal interpretations of what the semantic definition of causality might be, and tend to have a large degree of disagreement. So it is important to both establish a consistent semantic definition, and ensure annotators adhere to that as much as possible; thus for the definition and classification that adopted for this study, the foremost consideration is how such a system can guide annotators and other individuals in identifying the types of causalities relevant to the study, providing them hints and sign-posts to that end.

9.4.1 Classification in corpus studies

Causal expressions have different possible categorization schemes for works in computational semantics, and as a concept is not well agreed on for formal logic, psychology, or other branches of cognitive sciences. Currently, Penn Discourse Treebank (PDTB) is the foremost resource for empirical studies discourse structures of language (Webber & Joshi, 1998, Miltsakani et. al. 2004, Joshi et. al. 2006, Prasad et. al. 2007a, Prasad et. el. 2014), and contains a variety of relational types that can be studied in detail in their structural and statistical properties. This resource has been used in experiments in the automatic extraction of certain classes of discourse relations (such as: Prasad et. al. 2004, Lin et. al. 2009, Pitler et. al. 2009), including many classes that are considered causal. The PDTB is focused on explicit and implicit *discourse connective*s between two arguments, where arguments are defined as *'abstract objects'* that often can be correlated to eventualities in semantics.

9.4.1.1 PDTB annotations

There is a well defined annotation classification system for the PDTB (Prasad et. al. 2007b), where the concept of *causality* is sub-categorized along several dimensions of variability. The classification of the causal senses fall under the broad category of *CONTINGENCY* in PDTB, where the surface structure is ... arg_1 $CONNECTIVE$ arg_2 , with the *arg*s are the surface forms representing the events, states, properties, or entities under consideration; the subtree of which exhibits several modes of differentiations among subtypes. (Prasad et. al. 2007b) The lexical *connective*s between a pair of clauses plays a prominent role in determining the identity and classification of the causal relation represented.

First, the different types of connective-mediated relations with causal semantics are divided into *CAUSE*, and *CONDITION*, depends on whether the truth condition for each argument can be individually determined or not. There is also a differentiation between semantic and pragmatic cause in expression; the former is a causal relation between the world-extensions represented by the surface forms of the constituent arguments; while in the latter, one or more of the arguments does not directly represent the events / entities involved in the causal relation, but rather some implicature or pragmatic consequence of the surface argument. *CAUSE* is subdivided into *REASON*, where the direction of causality is $arg_2 \xrightarrow{caus} arg_1$, and *RESULT*, where the direction of causality is $arg_1 \xrightarrow{caus} arg_2$. *CONDITION*s are also further subdivided into types depending on values of truth conditions of arg_1, arg_2, and the dependencies between the truth conditions of the two args, and any implicit conditional mood involved.

9.4.1.2 distinct PDTB emphases

This classification scheme mostly makes distinctions among causal relations that pertain to morphosyntactic structure, the computation of truth values and pragmatic consequences of the component events. The surface direction of $\langle arg_1, arg_2 \rangle$ with respect to the underlying causal sequence is not relevant in our annotation tasks, since we only seek to find whether the underlying semantic structure is causal. The distinction between real and hypothetical (which conditionals may represent) sequences of event also is not weighty in our consideration, since we only test whether the underlying relation between $\langle e_1, e_2 \rangle$ is plausibly causal, not whether they actually occurred in the real world; much of the datasets we use, in fact, (such as a number of novels) contain long stretches of speech or thought of characters. The temporal distinctions likewise bears minimal relevance for our tasks. So all of the distinctions that this system makes on types of causal structures, with the exception of the pragmatic consideration, only pertain to the relation between surface structures and sequence, or the manner in which the pair resides in possible worlds semantics; they do not get to the core question of this study, whether the underlying semantics of e_1 and e_2 themselves indicate some causal relation is probable between them.

9.4.1.3 Deep semantics

The primary purpose of our classification scheme is provide annotators and our system a relatively well defined notion of causality, through the elucidation of various notions and properties of events that make them causal in human perception. This ultimately refers to how humans perceive the real-world events, as expresses by the whole linguistic constructions in semantic frame or larger structures, rather than how pairs of linguistic expressions relate to each other at some level not considering their full semantic and pragmatic contents. The prominent role of the lexical entries for *discourse connectives* in semantic definitions is particularly not suitable for the types of cognitive semantics of **events** that this study pursues.

While our system is informed to a degree by lexico-semantic and structural information, the algorithms themselves, especially *embedded causal* extraction, is not dependent on the enumeration of any single set of lexical items, or any single pattern or type of morphosyntactic structure; so it is essentially independent of other levels of linguistic representations above real-world extensions in event-level semantics and pragmatics, since that is the level of representation that can be connected to individual's perception of causality among events. Furthermore, in an effort to approach real-world semantics of events closely, this study pursues the type of semantics of relations among events that are informed by contextual information, and textual context of each sample is normally utilized for annotation, and sometimes even extra-corpus information as well. Thus, for the variabilities the the PDTB classification scheme makes distinction on, aside from the pragmatic consideration, while they are all relevant in many computational semantic tasks, they do not help individuals or annotators to sharpen and unify their common definition and conception of causal relations among real world events. Thus we will pursue a classification scheme that is primarily focused on cognitive considerations and real-world semantics.

9.4.2 Considerations of a viable scheme

The issue of assigning a relation between $\langle e_1, e_2 \rangle$ to a large extent depends on the knowledge base and perceptions of individual observers, so there is always a good chance for significant disagreement among individuals. One way we can partially mitigate this problem is to redefine causal relations as a collection polyphyletic semantic types that share the common statistical characteristics described in Section 7.2.1, but each class having a set of criteria for definition that is potentially independent of another. In order to design a classification scheme well suited for our task and relevant to the annotation task, it is important to take common concepts of individual language users into account, with cognitive concepts employed practically. Given the perceptual nature of judging linguistic causality, the classes and their associated properties also refer to the logical system as perceived by the individual, not an external system that is independent of that individual. So the system that we use for enforcing a relatively uniform notion of causality while annotating samples, as well as for the post analysis of the testing output, is one that aims at some common perceptual view of individuals of elements of a formal logical system of causal classification.

9.4.2.1 Classical scheme

The Aristotelian notion of causality contained four discrete types that are conflated into a single cognitive category. (Ἀριστοτέλης 1957, Ἀριστοτέλης 1994, Sachs 2001) This is a very well researched framework that persisted for the entire duration of human study of causality, and is highly intuitive for non-experts to grasp in relation to every day experiences. The fundamental conception of this categorization has root in the hylomorphic view of entities (Irwin, 1988) in the world able to be described through orthogonal compositions of *matter* and *form* (Caston 2006).

❶ There is one sense of causality in this classical framework that pertains to the substance that an object / entity is composed of, the *material cause*; this refers to change of state by some constituent element of the object / entity, such as

- *'the rust makes the steel column crumble'*
- *'the logs provide the fuel for the fire'*
- *'the water in the solution dissolves the crystalline salt'*

❷ A second sense in this framework pertains to the configuration of different components of the object / entity, or its *'form'*; this is *'formal cause'*, where the change of state or movement comes from the functional design or configuration of the subcomponents of an object, and is connected to the idea that information about the configuration can be one of its causes, such as

- *'formulae in newtonian physics allow us to send spacecrafts into earth's orbit'*
- *'the blueprint provided the design for the building'*
- *'the digestive system of the ruminants afford them a large advantage in surviving in grassland ecosystems over other herbievores'*

❸ Another sense in this framework pertains to the action or participation of an external party in bringing about the change or movement, the *effective cause*; this refers to the agency or instrumentality aspect of the causality, and is the most transparent type of causality in linguistic expressions, such as

- *'John broken the window while playing a game'*
- *'Mary sent the samples to be analyzed'*

- *'the rolling thunder frightened the children so that they ran back into the house'*

❹ The final of these senses pertains to goal or purpose of the change of state or movement, the *final cause* or the *teleological cause* of the system; this does not necessarily imply mental deliberation of the causer, but frequently involves some sentience and volition of the being that initiated the event, such as in the following, where the content in the paratheses make the teleology unambiguous, but may not always be explicit:

- *'the gentleman purchased a solitaire engagement ring (in order to make the proposal)'*
- *'the girl aced all her classes (so that she can gain entry into the best university)'*
- or in cases without sentience: *'the bacterial colony multiplied and grew in size (such that it consumed all of the nutrients on the agar plate)'*

9.4.2.2 Behavioral psychology's take on causality

There are also modern behavioral psychology adaptations of these basic principles for understanding human cognition on causality (Rachlin 1992, Hogan 1994, Killeen 2001, Killeen & Nash 2003, Alvarez 2009). These adaptations are interested utilizing the classical ideas for a consistent framework for describing causality in operant behavior and conditioning derived from B.F. Skinner's work (1938, 1953) for the analysis of behavior dependent data. One type of adaptation makes analogies between the four classical causes and elements of experiments in changes in neurology and behavior modification. Here, the material cause can be analogized with operant response, which is the prior operant behavior and neurological configuration; where as the final cause would correspond to some reinforcement, either the addition of a rewarding stimulus, or the removal of some aversive stimulus, thus providing a goal for the respondent; The efficient cause is likened to discrimination performed in the process, involving some intelligent agency in devising discriminative learning, tying specific operant response to corresponding reinforcement to produce the desired change in behavior; while the formal cause would be some model representation of this process, the internal structure within human psyche or the nervous system that reflects this learned association (Killeen 2004, Alvarez 2009).

Alternatively, behavioral psychology can adapt the four causes to be directed at the ontological considerations of an individual in relation to real-world phenomenon rather than pure experimental considerations (Alvarez & Montes 2006, Alvarez & Sass 2008, Alvarez 2009). Here the material cause can be viewed as the entire organism as a collection of substances, subcomponents (organs and systems), and internal processes, considered the raw material from which behavior can take shape and be changed by external stimuli; the formal cause corresponds to the model, a prior representation of the individual's behavioral patterns, which can be innate, phase sensitively imprinted, or acquired through cultural context; the efficient cause is then the agentive reason for a behavioral pattern, which can be attributed to some external being's explicit influence, but Alvarez (2009) also attributes to the individual's internal desire and impetus; the final cause then, is related to the individual's purpose and intent, the individuals mental state regarding the behavior within a temporal, if not predictive, structure.

The baseline form of causal relation in behavioral psychology is $E \rightarrow C$, similar to our event based representation. Causality couched in this behavioral framework at a fundamental level refers to the relations between the probability of the cause $P(C)$ and that of the effect $P(E)$ in the sequence of events, this is true of nearly all relations among *stimulus*, *response*, and *reinforcement* in behavioral sciences. This is more specifically $P(E|C) > P(E|\neg C) \wedge P(\neg E|C) < P(\neg E|\neg C)$, which is also the view of causal relation of this study and in behavioral psychology; but unless specifically warranted in a particular definition, we will omit this detail for simplicity, assuming the relation is underlyingly probabilistic.

9.4.2.3 Cognitive behavior of causal perception

A uniquely relevant area in behavioral psychology in relation to perception of causality is the study of hypnosis (Killeen 2003; Alvarez 2009), there an application of categories similar to the classical four causes are utilized for explanation of hypnotic phenomena. The act of hypnosis primarily relies on an alternation of the state of mind of the experimental subject, such that his/her perception of causality among event around him/herself is different than in the objective world. A hypnotism can be described as an inductor of the phenomenon alters the subject's state of consciousness such that he/she will disassociate the normal cause from some observed effect, and reassociate it with some imagined cause. An example would be when the subject moves his/her arm with own locomotion, but is entranced by the inductor to believe that the effort on his/her own part lifting the arm is not effecting the motion, but rather an imaginary balloon tied to his/her hand is, due to this altered state of consciousness called *hypnotic state* (Weitzenhoffer 1978, Killeen 2003). This decoupling of real cause and effect and mentally reconstructing the mental representation of the $C \rightarrow E$ relation necessitates an in depth analysis of the perception of causality.

The class *efficient cause* here are conceived as unitary events or phenomena that rises above some base-line context, which are generally regarded as facilitators and inhibitors of from response from the subject. In this type of Cognitive experiments, given the subject being in hypnotic state (HS), these are suggestions and stimuli that induces some response, whether sensory, motor, or verbal from the subject. Specifically for HS, external stimuli, particularly those introduced by the inductor, may be perceived and trigger responses from the subject without his/her awareness (Bargh & Chartrand 1999; Killeen 2003; Alvarez 2009). Thus the perception of causality may be subconscious and separate from an individual's conscious attribution of a cause to an event. The class *material cause* in this conception, refers to a collection of subcomponents of a larger event. But unlike the classical notion, it does not primarily refer to actual material making up an entity, but rather neuropsychological processes that underly the subject's response, the underpinning mechanisms of the CNS; such as the neurological in the *anterior cingulate cortex* which are associated with imagination and hallucination (Bush et. al. 2000; Killeen 2003), or electro-cortical activity in theta-frequency that are associated with hypnogogia (transition period in and out of sleep) and meditation (de Pascalis et. al. 1998; Graffin et. al. 1995; Ray 1997; Killeen 2003). Thus *material cause* in behavioral psychological sense relates to detailed sub-processes within events, which is a useful concept that is to be used in our definition, although not restricted to neuro- or psychological sub-processes.

The class *formal cause* here is described as a configuration of the mental state of the subject and the stimuli posed by the inductor of the HS. This especially refers to the how the experimenter induces an alignment the subject's perception of events with that purposed in the experiment (Killeen 2003; Alvarez 2009). This often refers to the framework of *signal detectability theory* (Green & Swets 1966; Killeen 2003), which regards how an individual resolves ambiguity in the stimuli to arrive at a decision on response. The individual seeks to maximize the payoff by adjusting their internal representation of an event based on a history of reinforcement; this acts as a model of the event for the individual developed in interaction with stimuli associated with the event, thus can be seen as strategy to develop a response geared toward payoff. Our study uses a similar concept of *formal cause* being a model of the event in the effect of a causal relation, a configuration that can be represented as internal structure of the events (among its subprocesses), described in detail later. For the class *final cause*, the behavioral psychological conception is largely based on motivation of the individual, or the functional end if there is no volitional being involved in the process (e.g. biological evolution). This is very similar to the classical concept, where individuals formulate responses based on the reward or punishment in the ultimate consequences of the event (Damasio 1994; Killeen 2003).

The *cause of necessity* is not represented as a separate category in behavioral psychology, but rather an orthogonal property to classification of behavioral semantics; this is also the view of this study (but for convenience of explanation to annotators we have placed necessary cause as a category). Our unmarked conception of causal relation can be viewed as cause of *sufficiency* can be represented as $C_s \cdot C \rightarrow E$, where C would include a set of default assumptions about the world context where C_s is placed; and with the addition of this one additional cause C_s, there is sufficient factors in the environment for E to come about. The complementary situation is then $\neg C_n \cdot C \rightarrow \neg E$; where some necessary cause C_n being missing from the default assumed world context and prevent effect from taking place ($\neg E$). Very few causal relations in the real world are verifiably entirely of sufficiency or necessity; such as the coming of winter is neither sufficient nor necessary for the occurrence of snow; but there tends to be a strong perceptive tendency of people viewing a particular causal relation as of sufficiency or necessity. In terms of *cause of intermediate volition*, although not explicitly represented as a subtype in this area of behavioral psychology, the hypnosis process itself is a model that necessarily involves an intermediate agent/experiencer. The subject in the experiment always is acting in a position of an intermediary with his/her own volition, although one that is continuously manipulated by the experimenter; thus the notion of *inducement* is incorporated into the overall experimental design itself, even if not in the HS processes that occur within the experiment.

9.4.3 A useful classification scheme

So starting with the classical and behavioral psychological frameworks of causality classification, we and make further adaptations to these to come up with a scheme that best fits the purposes of this study. The conception of causality in this study is focused on causal relations between events and states, in contrast with the largely entity / individual centered systems discussed earlier, with some concept of probability integrated; there is also an increased level of potential complexity for especially the embedded causality type, given that structurally it can contain recursively embedded sequence of many events, and involve more than two entities; these issues must be taken into account. For some of these classes, we may also need an intuitive name and description intelligible and helpful to the annotators and other not trained in this area.

9.4.3.1 Material cause

Here, substance no longer is relevant for events or states, but an analogous concept of subcomponents of events applies. Events and processes frequently composed of temporally or spatially distinct sub-events/processes. The component event processes each may have its own time duration within the duration of the encompassing event, may have a more specific spatial location within the overall spatial expanse where the encompassing event takes place, and each may have its own discernible causes and effects structured in some POSET form. For instance, the construction of a hydroelectric dam involves the sub-events of blocking the river flow, building a coffer-dam, installing generators, building a ship-lock, etc; the construction of the dam requires the installation of generator, and the installation in part completes the dam construction. Similarly, the craft of sewing requires of the sub-events of threading a needle, cutting the thread, piercing the fabric, running the stitches, etc.

For intelligibility we will term this class *constitution*, it always refers to at least the protasis A as a complex formulae of logical statements. It is logically expressed as $\Gamma_\iota, \Gamma_\kappa \subseteq_U \Gamma$, $\Gamma_\kappa \subseteq \Gamma_\iota \mid A = \wedge_{A_i \in \Gamma_\iota}$, $B = \wedge_{B_j \in \Gamma_\kappa}$. It can be viewed as a special type within *implication*, if we treat each A_i and each B_j individually. Here it means that both A and B are conjunctive events that takes some subset Γ_ι, Γ_κ, from among the set of environmental logical formulae Γ; and one of these is a subset of another, meaning that one of A, B is part and parcel of another; this presupposes that some events perceived by humans are conjunctive in nature and are naturally composed of sub-events that are required to make up the conjunctive event. It is likely perceptually distinguished from other *implication*s in linguistic data. This class requires some type of relation between $A = \wedge_{A_i \in \Gamma_\iota}$ and $B = \wedge_{B_j in \Gamma_\iota}$ to have some relation that is analogous to the *meronymy* relation among entities. Such relations could be: playing hockey is to handling a stick; shopping in the mall is to paying for an outfit; becoming a martial artist is to obtaining a particular belt; piloting a plane is to handling the control stick, watching the navigation instruments, and heeding ATC instructions; etc. These are basically component events or predications that are a part of a longer process or description. This can also be viewed that $\exists B_j \in \Gamma_\kappa$ that is a *generalization* of some $A_i \in \Gamma_\iota$, such as being a *"master blacksmith"* (a.k.a. being proficient at all the tasked involved in iron-work) is a generalization of being *"proficient at controlling the smelting furnace"*.

9.4.3.2 Formal cause

Similarly, the formal cause in our scheme is a functional configuration / pattern of the its constituent events, analogous to the arrangement of subsystems, organs, or other components in the behavioral psychological sense of formal cause. The effect of the cause here is the logical consequence that we can draw from a static collection and configuration of sub-events. Since this type is restricted to one point in time, the effect is mostly not another dynamic event, but a state or condition that arises from the combination of causes that can be instantaneously verified. We use the term *implication* for intelligibility, which also has several different meanings depending on the domain of knowledge, we mean here the *material conditional*, and is combinatorially defined within U. For instance, the coffee is finished brewing and it has been poured into a drinking vessel implies that it becomes potable (drink-able); or when the route is planned, the transportation is prepared, the hotels have been booked, and a house sitter is found implies that all of the conditions for the trip are met, so that the family are ready for the vacation.

It is formally expressed as $\nexists \Gamma_\iota \subseteq_U \Gamma$, $A \in \Gamma_\iota \wedge B \notin \Gamma_\iota$, where $\subseteq_U$ again is constrained by U. This is different from *entailment* in a subtle but important way, given that the logical system that people employ to judge semantic causality is not necessarily monotonic, and their belief system not necessarily static. that *entailment* allows for a sequential change within Γ that ultimately results in B, but *implication* here speaks of a purely combinatorial property of the set of formulae, and does not necessarily include a notion of semantic consequence with time.

9.4.3.3 Efficient cause

This class in our case is the same as the psychological adaptation of the classical notion of efficient cause, it is the dynamic impetus for the event, usually in the form of some volitional being or key instrument to bring about e_2. Unlike *formal cause*, there is dynamism involved in these relationships, such that the e_1 brings about some change in state of one or more entities/objects such that e_2 occurs. It is also the most familiar and apparent semantic type of causality in terms of their linguistic expressions, since there is always clear *causer* or *instrument* indicated or implied in the language. For instance John throwing a baseball may lead to the event that the window breaks; on the other hand, a meteor streaks through the sky may also lead to the same consequence.

Since this is the perceptually most obvious class of causality, we may term this *entailment*. We mean here *semantic entailment*, which is also a comparatively well defined model in terms of formal logic. It is formally expressed as $\forall \Gamma_\iota \subseteq_U \Gamma, A \in \Gamma_\iota\ [\Gamma_\iota \models_U B]$, U is the closed system being modeled, Γ is a set of logical formulae congruent with the conditions in U, $\subseteq_U$ here is not simply a subset notation, but only those subsets that are permissible under conditions within U. and $\langle A, B\rangle$ is the pair of candidate *protasis* and *apodosis*. In this ideal formulation, this is saying that no matter what the remaining conditions in the real world is, the presence of A will somehow lead to the presence of B, as long as the set of underlying assumptions is constant.

9.4.3.4 Final cause

This class is also very close to the classical and behavioral psychological notion of final or teleological cause. In our case, we will simplify the definition by limiting so that the *causer* (the primary entity of e_1) must have sentience and some fore-knowledge of the likelihood of the effect from cause. This can be intuitively explained to the non-experts as causal *'purpose'*, and associated with connective structures such as *'in order to'*, *'in a way such that'*. For instance, John sets a mouse-trap and the mouse is trapped is causally related in this fashion; or Mary shoots Patrick and Patrick dies from a gunshot wound is likewise.

This is a more esoteric case than the remainder of the four traditional types, since its defining distinction is a mental state, the *volition* of the primary entity involved in the causal chain in order to be classified as such, and intentionality is not a logically well defined concept. Since it is primarily based on volition, it does not require that B already occurred, but only that the intention of the *causer* is there to set events into motion to bring it about. This type of causality is also very often expressed using embedded structures, especially those a number of prominent patterns like *"for the purpose of"*, *"so as to"*, *"in order to"*, etc.

9.4.3.5 Intermediate volition

The classical notions of causal semantics takes one intelligent and volitional being in the causal sequence into account, which is reflected in the *efficient cause* type. The causal structures in our study, especially embedded causality, is potentially sufficiently complex so that multiple sentient beings may take part in the causal sequence. In these complex situations, there is a possibility that the primary entity in e_1 has no volition related to the causality (thus is not a *final cause* as defined above), but some intermediate agent, whose mental state was changed by the e_1, so that he/she effects e_2's occurrence having gone through that mental process; this we will term *inducement*. There is nothing to prevent both the primary and secondary entities to have volition at different points within the causal chain for the final effect; in that case both *final cause* and *inducement* are appropriate. This type is associated with many of the *manner-of-causation*-type verbs such as *'influence'*, *'enlist'*, *'persuade'*, *'compel'*, etc, where the direct object entities of those content verbs are the individuals (of e_2) whose mental states have been altered by the original action of the primary within e_1, so that event described in e_2 is effected by that intermediary agent; although *inducement*s are not limited to structures containing such clear indications by verbs.

This can be illustrated as the morning sunlight awoke the border guard, he started conducting his morning patrol; or inJohn conspired with Mary on the assassination plan, Mary shot Patrick. Inducement for our purposes is defined as a trigger that brings about the change in someone's disposition which leads to action. It is in that way in symmetry with *purpose* which requires *volition* on the part of entity in the causing event, *inducement* requires some change in mental state of an entity in the caused event to precipitate a new or altered event. The causal relation between $\langle e_1, e_2 \rangle$ induces the primary sentient entity of e_2 to alter his/her mental state or disposition in such a way as to be reflected in the completion of e_2.

9.4.3.6 Non-sufficiency causality

For the four causes, there is normally an underlying assumption of sufficiency of the cause, provided some default (unmarked) conditions in the remaining circumstances, which are assumed to be met; meaning that the occurrence of e_1 likely would be sufficient to bring about the occurrence of e_2. Some causal relations have protasis that provides the necessary preconditions, but no the necessity to bring about the apodosis. This can be evident in instances such as John opened the curtains, Mary saw him from outside; or the beaver built a dam on the river, several geese fish in the pond. In each of these cases, under the normal assumptions of default conditions in the real-world, the former event provides the necessary, but not sufficient condition for the latter to take place; this type we will term *enablement*.

This type has similar consequences as that of the four classical types in the real world, but it has far less stringent set of scenarios in which it is applicable; so when we speak about *enablement*, we mean the set of *enablement*s excluding any actual *entailment*. It is formally expressed as $\exists \Gamma_\iota, \Gamma_\kappa \subseteq_U \Gamma,\ A \in \Gamma_\iota \setminus \Gamma_\kappa\ [\Gamma_\iota \models_U B \wedge \overline{\Gamma_\kappa \models_U B}]$. In every day terms, it essentially means that the presence of A provides a presupposed condition for B, without which B is never likely to be true in the system, so A is said to be a pre-condition for B. By its nature, *enablement* predicates nothing about whether the apodosis event actually occurs, only that now one of the pre-conditions has been fulfilled by e_1. So e_1 provides some necessary but not sufficient condition for e_2 to take place.

9.4.3.7 Latent causal chain

Since the data here largely are complex causal structures each containing multiple complete events, many of these will naturally contain complex causal chains representing event sequences. A certain fraction of these sequences will will leave certain intermediate events within the chain unexpressed. Thus these cases are non-obvious causal structures, and perhaps the most difficult for humans to detect. This can be seen in many instances where the knowledge and expertise of the reader lags behind that of the author, such as the early hominids evolved a gait for efficient energy expenditure during long distance migration, so then these hominids evolved a more complex neurology and higher level of intelligence than their ancestors. This causal logic connecting the two above events would certainly be non-obvious to those not trained in evolutionary biology; the actual chain of causation is: ❶ a gait built for efficient (as opposed to speed or acceleration) prefers less unnecessary movement and musculature → ❷ the resulting bipedal locomotion changes the spatial and kinetic relation between the pelvis and the spine → ❸ this in turn allow for the skull to sit on top rather than forward of the C1 (atlas) of the spinal column → ❹ which allows much less requirement of the neck strength in bone an musculature to support a head of certain weight → ❺ which then allow for a skull of greater internal volume for the same neck strength → ❻ that then provides the necessary space for large area for cerebral cortex → ❼ finally providing the basis for complex neurology to develop. There are also many far less extreme, but nonetheless latent examples: Jane drank five cups of coffee yesterday, so then she is feeling very tired right now. Although these do not constitute a distinct type in the classical sense, since these are actually complex causal structures containing multiple individual causal relations, it is very helpful practical distinction to make for this study, which we will term as the eventual *outcome*.

Outcome is a type that share many similarities with the previous classes, but the description of its apodosis does not necessarily occur in the current system U. It is formally expressed as $\forall \Gamma_\iota \subseteq_U \Gamma, A \in \Gamma_i\ [\Gamma_\iota \models_U C, \{C\} \cup \Gamma_\iota \models_U D,\ \{C, D\}\Gamma_\iota \models_U E,\ \ldots, \Gamma'_\iota \models_U B]$, such that $\Gamma'_\iota = \{C, D, E, ...\} \cup \Gamma_\iota$. This is basically the expression of a longer (more than 2 link) causal chain of condensed into a single pair. Notice that this is far more likely to occur when each of the individual causality involved in the chain (between any two links) is an expressed *entailment* type, but unlikely with any other type, since entailment normally guarantees the occurrence of the next link event given the conditions in U; and there is not a logical reason to express a longer chain with a series of *implication*s, since Γ_ι is the same for each causal relation in the series.

9.4.3.8 PDTB annotation examples

Given the above classification scheme used for this project, we will do a brief comparison with the causal structures in the PDTB's annotation scheme. We will do this by going through the examples in the PDTB's annotation manual (Prasad 2007). Due to the fact that the emphases of the PDTB annotation and that of this study are mostly orthogonal, there is very little correlation between any single category from one scheme to another from the other scheme. All of the examples are drawn from the "CONTINGENCY" overall category, which is the categorization most similar to what we mean by *'causality'*. The first broad category under CONTINGENCY is CAUSE, which are causal relations between a pair of events that are known for fact to have happened in the real-world. Samples normally have some *discourse connective* that serves as the primary cue for detection of the causal relation, which is underlined below; the two clauses that are dominated by the connective are the ‘ARG’s of the identified relation.

1. **PDTB (semantic) “CAUSE” :**

   i **PDTB reason:** *use of dispersants was granted when a test on the third day showed some positive results, officials said*

   √ THIS WOULD LIKELY BE IDENTIFIED AS *efficient cause* DUE TO THE FACT THAT A UNITARY EVENT PROVIDED THE TRIGGER FOR THE SUBSEQUENT EFFECT; IT COULD ALSO BE REASONABLE IDENTIFIED AS *enablement*, ALTHOUGH THERE IS NO EXPLICIT INTERMEDIATE AGENT, HIS/HER PRESENCE CAN BE INFERRED

   ii **PDTB result:** *in addition, its machines are typically easier to operate, so customers require less assistance from software*

   √ THIS WOULD LIKELY BE LABELED AS *formal cause*, AS THE CUSTOMERS REQUIRING OR NO REQUIRING ASSISTANT IS A STEADY STATE OF EVENTS, AND INVOLVES A COMPLEX CONFIGURATION OF PROBABLY MULTIPLE CUSTOMERS’ STATES OF SERVICE NEEDS

2. **PDTB “PRAGMATIC CAUSE” :**

   i **PDTB justification:** *Mrs. Yeargin is lying, because they found students in an advanced class a year earlier who said she gave them similar help*

   √ THESE TYPES OF CONSTRUCTIONS REQUIRE IMPLICATURES TO OBSERVE THE CAUSALITY, SO IT DEPENDS ON THE ABILITY AND EXPERIENCE OF THE ANNOTATOR; IF SUCH CONNECTION BETWEEN ‘MRS. YEARGIN IS LYING’ AND ‘THE KNOWLEDGE THAT MRS. YEARGIN IS LYING’ IS MADE, THEN THIS IS LIKELY IDENTIFIED AS *material cause*, SINCE FINDING OUT THAT A STUDENT ILLICITLY RECEIVED HELP FROM HER IS A SUB-EVENT IN CONSTITUTING THE KNOWLEDGE THAT SHE IS LYING

The second broad PDTB category under CONTINGENCY is CONDITION, which describe causal relations between a pair of events that are not known to have occurred with certainty, and thus has at least one of the events as possible, potential, or occurring in some other world time-line not our own. Given that we only observe some relation as causal if the frame-semantics of the events are causal by individuals, without regard to whether event sequences occurring in the real world or some other possible world, this is not a fundamental variability that we take into account, and similar types according to our scheme should be observed.

A **PDTB (semantic) “CONDITION” :**

i **PDTB hypothetical:** *both side have agreed that the talks will be most successful if negotiators start by focusing on areas that can be most easily changed*

√ this is most likely to be regarded as *cause of necessity (enablement)*, since focusing on areas that can be changed by both side is a prerequisite to agreeing on changes; it could also be reasonably seen as *efficient cause* for the causal relation is base on one change of state triggering another

ii **PDTB hypothetical:** *in addition, Black & Decker had said it would sell two other undisclosed Emhart operations if it received the right price*

√ this is most likely to be annotated as either *cause of necessity (enablement)* since receiving the right offer is a pre-condition for completion of the sale; and like the previous example, it can be equally validly regarded as *efficient cause* due to receiving the price potentially triggers the buy

iii **PDTB general:** *it explains why the number of these wines is expanding so rapidly. But consumers who buy at this level are also more knowledgeable than they were a few years ago. “They won’t buy if the quality is not there,” said Cedric Martin of Martin Wine Cellar in New Orleans*

√ this is likely to be recognized as *formal cause*, because the fact that *‘the quality is not there’* describes the composite states of these many wine products, which is a configuration that prevents the purchase of these products

iv **PDTB factual present:** *“I’ve heard that there is $40 billion taken in nationwide by boiler rooms every year,” Mr. McClelland says. “If that’s true, Orange County has to be at least 10% of that.”*

√ this is also likely described as *formal cause*, as the cause here is a persistent (year over year) condition, that circumscribes a set of events during each time period, which results in another persistent state of boilers within Orange County in relation to that larger picture

v **PDTB factual past:** *“If they had this much trouble with Chicago & North Western, they are going to have an awful time with the rest.”*

√ this is likely to be regarded as *formal cause*, as the protasis describes a complex arrangement of previous events that lead to the lack of rest; it can also be reasonably regarded as *efficient cause* if someone takes the reading that *'trouble with Chicago & North Western'* to be a single unitary occurrence

vi **PDTB unreal present:** *Of course, if the film contained dialogue, Mr. Lane's Artist would be called a homeless person.*

√ this is most likely to be seen as *cause with intermediate volition (inducement)* as there is an implied intermediate agent (the person that called him/her a homeless person); if the reader takes a different reading that this is just a generic description of the action of *"calling"*, then *formal cause*, with the state of the film with regard to dialogs being the causal configuration, would be the most likely type

vii **PDTB unreal present:** *I'm not saying advertising revenue isn't important," she says, "but I couldn't sleep at night" if the magazine bowed to a company because they once took out an ad*

√ this is most likely to be seen as *cause with intermediate volition (inducement)* since an intermediate experiencer here is clearly involved in effecting the second event, which is directly brought about due to a change in the mental state of this entity

viii **PDTB unreal past:** *"if I had come into Friday on margin or with very little cash in the portfolios, I would not do any buying*

√ this is most likely to be regarded as a *formal cause*, as the lack of funds in the portfolios is a configuration involving a confluence of states which results in another state of the individual described (inability to purchase); although some may view this effect as a choice and not a natural outcome (if he/she can borrow money, e.g.), in which case *cause of intermediate volition* would be a better label

B **PDTB "PRAGMATIC CONDITION" :**

i **PDTB relevance:** *if you are thirsty, there's beer in the fridge*

√ as with PRAGMATIC CAUSEs, this requires knowledge of the implicature of telling someone there being beer in the fridge as a permission for consuming the beverage; if that connection is made by the annotator, then this is most likely *formal cause*, as a configuration of the physiological state of the individual causes him/her to drink the beer

ii **PDTB relevance:** *if anyone has difficulty imagining a world in which history went merrily on without us, Mr. Gould sketches several*

√ this also requires some logical extension beyond the explicitly described events; here the difficulty of someone imagining this scenario may explore some ways of describing such, and then may come to know that Gould has such an ability, which lead him/her to ask Gould to make such a description; so this is best considered to be a *latent causal chain* where some of the links are implicit

iii **PDTB implicit assertion:** *in 1966, on route to a re-election rout of Democrat Frank O'Connor, GOP Gov. Nelson Rockefeller of New York appeared in person saying, "if you want to keep the crime rates high, O'Connor is your man."*

√ this also requires some extension of the logic to arrive at a *latent causal chain*; Nelson Rockerfeller is saying here that if O'Connor is elected, he would implement law-enforcement and other policies in such a way, possibly through the laxness of enforcement or economic malaise, as to bring about more crimes in the city.

The number of samples picked as representative by PDTB project is relatively small, but some observations can still be made. As expected overall, categories such as *efficient cause* and *formal cause* are widely distributed among PDTB categories, since these are both relatively easy detectable by humans, and are almost universally regarded as causal. The only surprise is that *cause with intermediary volition* have multiple examples, which usually occurs when the causal chain is sufficiently complex so that multiple sentient entities are involved; this could mean that the PDTB has fair amount of complexity in its corpus, or an artifact of example selection. For the majority of these cases that we have examined, there is little correlation between any specific category in PDTB scheme and some class of our scheme. The lone exception to that is that *latent causal chain*, where complex causal sequence is broken by some links which are not explicitly represented in the linguistic form, but takes additional inference, seems to be PRAGMATIC CONDITION, or perhaps pragmatic categories overall; this is unsurprising since these structures do require some logical inference by humans to detect as causal.

9.4.3.9 Intersections of subclasses

The four classes derived from the classical types are largely mutually exclusive, and have very little if any interaction, but there are some intersection between any one of the original four types and one of the remaining types, and between each pair of the remaining types, such as the previously discussed intersection between *purpose* and *inducement*. The following are the most prominent. A common set of intersection would be *inducement* with *entailment*, since influence on the mental state of the intermediary *causer* entity is a dynamic process that often requires a state transition of some type, such as the moon peeks out from under the cloud, which inspired John to write a poem about the scenery at night. *Enablement* and *purpose* is also another common pair, since the provision of a necessary precondition often involves the intervention of a sentient being in e_1 such as Mary attended college, so that she might have a better profession in the future. There are also frequent cases of *outcomes* containing *efficient cause*, *final cause* or *enablement* as sub-sequences, particularly in domains that require specialist knowledge, such as ❶ mount Tambora underwent volcanic eruption → [❷ the volcanic plume carried much sulfur dioxide high into the atmosphere → ❸ the sulfur dioxide becomes distributed with stratospheric jets → ❹ the sulfur dioxide reacts with water present there to form sulfuric acid in minute droplets → ❺ sulfuric acid laden droplets reflect a fraction of solar radiation entering the stratosphere] → ❻ the earth's climate cools for the following months or years. In this case, the entire *outcome* sequence is established by looking at the individual sub-sequences, and multiple instances of both *efficient cause* and *enablement* are present to establish the causality, where the sub-sequences within [ ] are implicit and require additional knowledge to ferret out.

9.4.4 Classification of observed samples

Given the outlined perceptual classification above, we will discuss observations of the output data within this classification scheme. From a sampling of the highest ranked quantile from each dataset in each experiment, we see that all of these classes are well represented in the testing data; the classes are roughly ordered in increasing difficulty for human to detect and identify. For each of the classes, we will first briefly remind the reader of its definition by presenting a generic illustration. The paired examples number (a) and (b) are those in the top quantile the adjunctive causality module's results; while the numbered examples (i, ii, iii, ...) with the addition of original text with highlights are from the top quantile of embedded causality module's results. While the adjunctive results are pairwise, and therefore the adjunctive example included in each category only contains that type, the embedded results potentially have longer chains, thus additional annotation of sub-types are placed between any adjacent pair of longer chains.

While there is some probability that all different sub-types of causality to be contained in either the ranked adjunctive structures as well as embedded structures, adjoined causality is observed in this study to have a more even distribution of different types, where as adjunctive structures are skewed toward expressing *entailment*, *implication*, and *purpose* sub-types of causality. We need to bear in mind that in terms of the logical system used in these definitions, it is the logical system within people's minds (who make judgment about causality) that is paramount, and thus is not required to correspond to some external logical system that is rigorously and empirically tested. Thus this subsection is primarily a discussion over the data-set extracted with adjunctive causal structures in mind.

9.4.4.1 Efficient cause (entailment)

Semantic *entailment* can be illustrated by an example such as, *"John reached his bare hand into the smelting furnace, and suffered severe burns"*. In this case, there is no other possible outcome than his hand being burnt in any real-world scenario, and the e_2 is fully entailed by the occurrence of e_1 in the process. Not all the real-world situations are as clear cut as the above example, since there are usually other unlikely choices and remote possibilities of some configurations of Γ_ι that could be taken into account. But the *entailment* in most cases refers to some overwhelmingly most likely resulting e_2 from e_1, aside from some low probability scenarios as perceived by the individual.

1. (a) *when the nazis took power, he went under-ground*

   (b) *he worked in the communist resistance, was arrested*

2. (a) *the dissident minister to buck cabinet responsibility*

   (b) *[him/her to] risk the sack*

3. (a) *eurotunnel is already in default of its credit agreement with the bank syndicate*

   (b) *it is seeking an extra xx billions on top of the xx billions raised so far*

4. eurotunnel is already in default of its credit agreement with the banks , [it] has in effect been given until the end of the year to settle its differences with the contractors to permit a viable financing strategy to be put in place .

   i *eurotunnel is already in default of its credit agreement with the banks,*

   ii $\xrightarrow{entails}$ *[it] has in effect been given until the end of the year to settle its differences with the contractors*

   iii $\xrightarrow{purpose}$ *to permit a viable financing strategy to be put in place*

5. [Laurie feeling] that out of the grave of a boyish passion there had risen a beautiful, strong friendship to [the friendship] bless them both (Laurie and Jo)

   i *out of the grave of a boyish passion*

   ii $\xrightarrow{eventually\leadsto}$ *there had risen a beautiful, strong friendship*

   iii $\xrightarrow{entails}$ *[the friendship] bless them both*

6. before the housewives could rest several people called and there was a scramble to get ready to [the housewives] see them (receive them with hospitality)

   i several people called

   ii $\xrightarrow{entails}$ there was a scramble to get ready

   iii $\xrightarrow{enables}$ [the housewives] see them

9.4.4.2 Cause without sufficiency (enablement)

We can see an *enablement* relation in *"John put up the carosel in the back-yard, Mary sat on the wooden horse and rode around the carosel"*. The e_2 in this case, Mary using and enjoying the merry-go-around in the back-yard, is only possible given that someone put up the amusement ride in the first place, which was accomplished by John. We can see that in some highly-ranked examples from the corpus:

1. (a) *the odd militant supporter appears at the conference rostrom*
   (b) *he is easy to spot in the wrong place*
2. (a) *how information is gathered*
   (b) *how information is acted upon*
3. (a) *I lifted my chin as he (Augustus Caesar) stared at me (Cleopatra Selene II, a.k.a. Cleopatra VIII)*
   (b) *[I (Cleopatra Selene)] Let the emperor look at me an think of my mother (Cleopatra VII Philopator)*
4. this year it was to be a plantation of sun flowers the seeds of which cheerful land aspiring plant were to feed Aunt Cockle-top and her family of chicks
   i *it was to be a plantation of sun flowers*
   ii $\xrightarrow{eventually\rightsquigarrow}$ *the seeds of which cheerful land aspiring plant were to feed Aunt Cockle-top and her family of chicks*
5. she tries to find high-born women [high-born women] to bear him a son that she can take in as her own
   i *she tries to find high-born women*
   ii $\xrightarrow{enables}$ *[high-born women] to bear him a son*
   iii $\xrightarrow{enables}$ *she can take (a son) in as her own*

9.4.4.3 Formal cause (implication)

A simple illustration of *implication* would be *"John ordered Mary to kill Patrick; Mary did not act completely on her own accord"*. In this case, part of the circumstances that is required for e_1 to be true also means that e_2 is also necessarily the case. Similar cases can be found with the top-ranked datasets:

1. (a) *you can be more extravagant*

   (b) *you can go for more dramatic, impressive displays for special occasions*

2. (a) *the blade glides easily over my skin*

   (b) *it leaves it very smooth*

3. (a) *the more stubbornly East Germany's old men resist the changes*

   (b) *greater becomes the danger that the whole edifice they have constructed will collapse around them*

4. the more stubbornly East Germany's old men resist the changes, greater becomes the danger that the whole edifice they have constructed will collapse around them, threatening stability in europe and raising the question of reunification for which no one is prepared

   i *the more stubbornly East Germany's old men resist the changes*

   ii $\xrightarrow{implication}$ *greater become the danger that*

   iii $\xrightarrow{eventually\leadsto}$ *the whole edifice they have constructed will collapse around them*

   iv $\xrightarrow{implication}$

      i *threatening stability in Europe*

      ii *raising the question of re-unification (with West Germany)*

9.4.4.4 Final cause (purpose)

A simple example of a *purpose* type causality can be *"John gave some false information to incite Mary, so as to encourage Mary to kill Patrick"* These are samples that involve some active volition on the part of the *causer*.

1. (a) *exchange rate is held up artificially high by interest rate*

   (b) *it prejudices exports and encourages imports*

2. (a) *his fledgling company was enabled to buy a license from western electric*

   (b) *to develop transistor technology*

3. (a) *forward planning and good communication are the two foundations that must be in place*

   (b) *to guaranteed that people with HIV have good choices so that they continue to enjoy the very good quality of life wherever they choose to be*

4. akia moria recorded that his fledgling company sony was enabled to buy a license from western electric to develop transistor technology

    i *akia maria recorded that his fledgling company was enabled*

    ii $\xrightarrow{enables}$ *to buy a license from western electric*

    iii $\xrightarrow{purpose}$ *to develop transistor technology*

5. forward planing and good communication are the two foundation stones [that] must be in place to guarantee that people ill with HIV have the good choices and continue to enjoy the very good quality of life wherever they choose to be .

    i *forward planning and good communication are the two foundation stones [that] must be in place*

    ii $\xrightarrow{purpose}$ *to guarantee that people ill with HIV have good choices*

    iii $\xrightarrow{purpose}$ *[so that they] continue to enjoy the very good quality of life wherever they choose to be*

9.4.4.5 Cause with intermediary volition (inducement)

An example of causality with intermediate volition could be *'John handed the admission letter from her top school to Mary. She leapt for joy and went out and celebrated with her friends*; where the inducement comes where *John*'s action of giving the letter provides a trigger that allows *Mary* to transition to a new mental disposition to bring about the subsequent events. Some examples from our data are:

1. a succulent hash arrived and Mr. Wolfsheim, forgetting the more sentimental atmosphere of the old Metropole [Mr. Wolfsheim] began to eat with ferocious delicacy

    i a succulent has arrived

    ii $\xrightarrow{induces}$ [Mr. Wolfsheim] forgetting the more sentimental atmosphere of the old Metropole

    iii Mr. Wolfheim, ..., began to eat with ferocious delicacy

2. the old gentleman knew that perfectly well and particularly desired to prevent it for the mood in which he found his grandson assured him that it would not be wise to leave him to

his devices so stifling a natural regret at the thought of the home comforts he would leave behind he said stoutly: bless your soul, I'm not superannuated yet. I quite enjoy the idea (of traveling to London)

i the old gentleman know that perfectly well

ii $\xrightarrow{induces}$ particularly desired to prevent it

iii $\xrightarrow{induces}$ stifling a natural regret at the though of the home comforts he would leave behind

iv he said stoutly: .... I quite enjoy the idea (to express that he is willing to travel)

3. I'd asked the emperor to send us (Juba and Cleopatra Selene II) to Mauritania to [caesar] allow us to [Juba and Cleopatra Selene II] persuade the people there to build a port

i *I'd asked the emperor to send us (Juba and Cleopatra Selene II) to Mauritania*

ii $\xrightarrow{induces}$ *[caesar] allow us*

iii $\xrightarrow{induces}$ *[Juba and Cleopatra Selene II] persuade the people there to build a port*

9.4.4.6 Material cause (constitution)

A simple example of a *constitution* relationship would be *"John completed his degree at the university, so he passed all of his classes"* . In which both e_1 and e_2 are composite formulae that can be analyzed as a conjunction of multiple simple events, where the set of events corresponding to e_1 contains all the necessary events that constitute e_2 . Here are some top ranked examples that fit well with this type:

1. (a) *it is very low in mineral, it suits a wide range of people*
   (b) *it is ideal for families with young children*
2. (a) *the wicker screen is ideal for dividing your lounge*
   (b) *it conceals unwanted clutter*
3. (a) *retirement is a time of great change*
   (b) *it is also a time for development, either from choice or because they cannot find employment*

4. jimmy knapp, leader of the national union of railway-man, argues that his members take action in support of the miners during the coal strike having a genuine interest in the fight to keep pits open .

   i *jimmy knapp, ..., argues that his members take action in support of the miners during the coal strike*

   ii $\xrightarrow{constitute}$ *having a genuine interest in the fight*

   iii $\xrightarrow{purpose}$ *to keep pits open*

5. and if difficulties were necessary to increase the splendor of the effort what could be harder for a restless, ambitious girl to [the restless and ambitious girl] give up her own hope, plans, and desires and [the restless and ambitious girl] cheerfully live for others

   i *difficulties were necessary to increase the splendor of the effort (Counterfactual causality)*

   ii $\xrightarrow{constitute}$ *what could be harder (which would be the hardest) for a restless, ambitious girl*

   iii *[the restless and ambitious girl] give up her own hope, plans, and desires*

   iv $\xrightarrow{induces}$ *[the restless and ambitious girl] cheerfully live for others*

9.4.4.7 Latent causal chain (outcome)

It is often the case that complex causal chain is the least obvious type of causal structure to detect in the text by human judgment, since a lot of context might be required to ascertain the contents of Γ_ι at the time when the event described by the *protasis* takes place. An illustration can be seen in an example such as *"Mary walked out of the door with a long-range rifle, later that day, Patrick was found dead in the center of the city square"* . The connection between the two events as described is not always obvious to humans making a judgment on causality, it requires some advanced inference about things like: what a rifle is normally used to accomplish, the time frame of the events and the time it takes to travel between locations of e_1, e_2, any personal relationship there might be between Mary and Patrick, and a variety of other real-world issues.

1. (a) *your boyfriend is finding it hard to come to terms with the prospect of fatherhood*

   (b) *he is taking out his resentment on you*

2. (a) *granting relief to an applicant who had delay would cause substantial prejudice or hardship to some person*

   (b) *it would be detrimental to good administration*

3. (a) *the cutting of equivalent to a p reduction in the basic rate of income tax*

   (b) *to be replenishing some of the consumer depleted ammunition*

4. [Laurie feeling] that out of the grave of a boyish passion there had risen a beautiful, strong friendship to [the friendship] bless them both (Laurie and Jo)

   i *out of the grave of a boyish passion*

   ii $\xrightarrow{eventually\rightsquigarrow}$ *there had risen a beautiful, strong friendship*

   iii $\xrightarrow{entails}$ *[the friendship] bless them both*

5. the cutting of equivalent to a p reduction in the basic rate of income tax is expected to be replenishing some of the consumer depleted ammunition .

   i *the cutting of equivalent to a p reduction in the basic rate of income tax*

   ii $\xrightarrow{eventually\rightsquigarrow}$ *to be replenishing some of the consumer depleted ammunition*

   (Background: the cutting of income tax level provides more disposable income, which increases the amount of purchase power (the ”ammunition”) to sustain demand)

Chapter 10

Conclusion and Future Direction

For this study, we were able to achieve enhancement of the discover process of causal structures, by focusing on certain complex causal structures that are difficult to access through traditional methods. There are also numerous ways in which we may improve on the current system and extend into related applications.

10.1 Summary of current findings

For this study, we designed and demonstrated a set of procedures to rank the likelihood of causality from complex linguistic structures that are in forms of adjoined pairs of full clauses, as well as deeply embedded structures of multiple clauses. For adjunctive causal structures, the process takes two adjacent clauses as a single sequences. Using a standard HMM model, and a set of modifications that are specifically designed for this task, our algorithm was able to integrate some important lexical and hierarchically structural information into the training of individual HMMs. And the set of resulting HMMs were able to rank all adjacent clausal pairs, with very high probability of truly causal pairwise relations between clauses in the top quantiles.

For embedded causal structures, the process takes lexico-semantic as well as morpho-syntactic information in the expressions into a single form of representation; a collection of which then is extended into a *diffuse prototype*, a composite cognitive categorization model, for a complex multi-modal description of causality. An evolutionary algorithm, with a graph theoretic focus, is developed specifically to obtain the *diffuse prototype* from a limited number of training samples. The output model then can be used to score unseen samples according to a variegated notion of causality. Due to the nature of the model representation and the GA-like procedure, it is adaptable for a wide variety of human definitions of causality.

10.2 Further improvements in current system

There are certain directions in which in can make improvements to the various components, we will briefly discuss some of the most obvious next steps. Each of these proposed improvement is potentially an independent project, but built on the procedures of the current study.

10.2.1 Improvements in adjunctive causality

Going forward, we will continue to explore the possible improvements (some of them already evident after first run on large dataset) that can be made to the adjunctive causality procedure. With some incremental improvements to the system, we will apply that to BNC as well as another dataset, and then annotate the scored and ranked outputs based on the same type of quantile-based evaluation metric.

Given that our methodology so far ignores the bulk of the information in the wider context (outside of the pair), and ignores the information that is in the core components of the extracted semantic frames of the pair, we can place some simple pieces of these types of information back into the system, to see whether it affects the results. We can incorporate these additional sources of information into the procedure, information such as the presence or absence of certain entities shared arguments in the semantic frame pairs. We can also obtain other ready made systems that rely primarily on either contextual information, or frame internal essential information on the same datasets, and see what the performances are like, and whether there are intersections of the results, as well as the relative strengths or weaknesses of each.

We will make further analyses on the output, and use these new observations to also improve on the classes of causal structures that we have obtained so far. It is possible to add to the classification that we have thus far, also possible that we would need to refine these definitions to take into account new facts about adjunctive causal structures. In addition, we suspect that the n-gram baseline for adjunctive causality would perform better and much closer to our system, given that it has a more appropriate, and larger data-set. We would like to obtain a better data-set for the baseline, one with adjacent pairs of clauses in a corpus, that are specifically labeled to be causally related, to train an n-gram or similar well known model on sequential structure, in order to obtain a better baseline performance for this.

10.2.2 Improvements on embedded causality

One potential area of improvement here is giving individual substructure types different importance. Currently, all substructures of the same size would provide the same amount of contribution during prototype consolidation. The consolidation phase relies on subgraph isomorphisms between test samples and the prototypes, and add to the score of a test sample when an isomorphic region is found, so only $n(T_s)$ and the frequency of a particular substructure type. But other properties of substructures are also potentially important in determining their appropriate contribution in the ranking of test samples. One such would be certain semantic classes of tokens that are prevalent among embedded causality, such as the *manner-of-causation* class of lexical verbs (Section 6.1.5). In the study we used the *lift* metric to determine the usefulness of individual lexical types among the terminals, so their presence caused the contribution of their host sub-structures to be weighed differently during the consolidation phase. There are reason to believe that presence of certain pairs, or k-tuples of terminals (not necessarily adjacent) in the substructures of the *diffuse prototype* could even be more telling with respect to the host substructures' importance to discovering their causality, and thus should be weighed with yet more importance; these would require more computational complexity, but could be explored further in follow on studies.

10.2.3 Potential classification task

During our study, we explored the possibilities of using a classification of scheme that came originally from classical sources, adapted by behavioral psychology to investigate cognitive processes involved in recognizing causality. For the embedded causal procedure, it can be adapted to discover a variety of different types of semantically sequential phenomena in complex lexico-syntactic structures, give that the *diffuse prototype* is capable of representing any set of substructures frequently occurring for a specific semantic class, and the *evolutionary algorithm* can be used to build the diffuse prototype from set of pairs of complex linguistic structure. So if there are multiple set of training data for each of the sub-classes of causal structure, then it would be possible to use the procedure to discover what is unique about each. There would be substantial overlap among these sub-categories, given that they all would share some features common to causality, but it would be a worthwhile investigation to see whether such a procedure can be adapted to find finer grained semantic classes using relatively small amounts of training data.

10.3 Extension into other linguistically expressed complex relations

Beyond the two major types of semantic relations among events discussed in our study (*reciprocity* and *causality*), there are other important relational classes such as *cooperativity*, *explanation*, *elaboration*, etc. The automatic extraction of information about these relations would be critical for constructing a complete representation of a network of individuals and events in social network or mobile applications. Each of these would present us with its own set of challenges, we will discuss the issues among *cooperatives* as an example. Among cooperative relations, for example, there are a number of major subtypes. These need distinct treatment in terms of the topology of the subgraphs representing each, as in the following:.

1. *John, Kate, Leo, and Mary moved a piano across the concert hall*
2. *John, Kate, Leo, and Mary carried a notebook to the lecture*
3. *John, Kate, Leo, and Mary met in the school building*
4. *John, Kate, Leo, and Mary finished six pot pies during dinner*
5. *John completed the choreography with Kate, Leo, and Mary*
6. *John completed the choreography, having Kate, Leo, and Mary at his side*

First, we must qualify that the interpretation of each of these depend in part on pragmatics, and without extensive context, there is no way to be certain of the correct topology meant by the original speaker, but we can make reasonable inferences about what is the likely scenario of each. Among these examples: (1) clearly prefers the interpretation that all of the involved parties performed a single action together of moving one piano across the hall. (2) strongly prefers the interpretation that each of them performed the action of 'carrying' on his or her own notebook. (3) has two distinct readings with significant probability, The first (3a) interprets all of them as meeting at the same place and the same time, while another (3b) slightly less likely reading places some subgroups of them meeting somewhere in the building at different times, but the union of all of the memberships of the meetings comes to be this group of four. (4) clearly has a prefers the meaning of the sum of the number of pies that they each finished comes to six. (5) has a number of different interpretations, differing in the roles that *Kate*, *Leo*, and *Mary* played during the choreography, likewise (6) also have various and an even wider range of interpretations. All of these are subtypes of cooperative relations, sometimes termed as *collective*, *distributive*, *cumulative*, *comitative*, and *applicative* readings of the surface forms, the first three of which are formally readings of pluralities, while the others are analytic constructions with some of these having their own sub-classification schemes. All of these will require separate treatment in their conversion into representative topology as a part of the graph, in order to remain faithful to the most likely meaning by the speaker;the detailed analyses of these cooperatives as well as those of the other major feature types we will explore in appropriate sections.

Once the relevant features are converted into their corresponding graph regions, there needs to be a set of well defined graph operations in iteratively composing the regions together into a single (possibly multi-component, if the social network has disconnected regions) structure. There are several requirements in designing this set of graph operations. It should minimize the representation of a single entity within the network as multiple vertices, and minimize the representation of an identical event by multiple instantiations of graph regions. It should have time complexity of no more than some low order polynomial of the $order * density$ of the final graph representation. The order in which the individual linguistic features are discovered and processed should not substantially affect the topology of the eventual structure.

10.4 Automatic retrieval of SN structure from linguistic information

Our ultimate aim is to build a system of event and relation discover that is able to reconstruct a social network (SN) from the linguistic data produced there within. This would allow for the construction of the graph-theoretic representation of an SN, as a relational structure among entities and events. The data-base of extensive and real-time analysis of complex entity and event relations of many semantic types, such as *reciprocity*, *causality*, *cooperativity*, etc, must be automatically acquired to enable robust applications. We can see their important by observing recent trends in both the proliferation of mobile applications and the prevalence of SN usage.

The evolution of the interaction between the web and the end user have taken the shape of several major iterations. For the early days of the Internet, shortly after becoming accessible to the general population (as client machines and network infrastructure became pervasive in the work place and at home), the web consisted of a collection of documents, with a few content creators, and a relatively large number of content consumers. At this stage, web sites were largely static and did not interact with the end users on a regular basis, which is widely known as the pre-dotcom bubble Web 1.0. The reader might remember the now defunct geocities.com as a prime example of the type of interaction during this period.

As Information Technology entered the web enters the Web 2.0 era, the web became regarded more as a collection of user communities each evolving over time. The predominant interaction between the end-users and the web gradually became much more fluid and bi-directional, and everyone participating in such communities is a potential producer and consumer of content. As a result, a plethora of web services built on this model rose to prominence, including web forums, blogs and on-line journals, comment pages of news and on-line media sharing sites, on-line wikis, dedicated social networking sites, and most recently real-time microblogging services.

In the mature form of Web 2.0 infrastructure as we know web today, the most visited web sites a set of blank canvas and organizational tools where the users can share their knowledge and creativity, and the predominant mode of this sharing takes place as text. The content within the web services are all uploaded, managed, and viewed explicitly by individual end-users (aside from web administrative tasks). This model bestows much more freedom on the user interactions compared to the previous one, but also requires constant attention and deliberate action on the part of the end users.

At the current time, we see the gradual emergence of yet another iteration of this development, from the Web 2.0 model of web sites being a mere sharing tool to the next version of the Web (some call version 3.0)'s mantra of the web being a collection of smart and interconnected applications that is capable of themselves creating and re-organizing content on behalf of the end user, with knowledge inferred from users' on-line activities and contributions. This moves the web from being a mere reflection of user's expressions to one that is an extension of the user's intent. This also requires that multiple on-line applications interact continuously with regard to all of the content that pertains to some common user identity across the web. One of the central technologies and necessary component of the future web, in order for this to fully come to fruition, is procedure(s) of automated discovery of user entities on-line and the construction of an accurate graph theoretic representation of the social network they compose.

There already exists a plethora of *explicit* information provided by the users themselves in most of these networks, such as profile information, visual and location tags, attendance at events, likes and dislikes, external links, etc, each of which is readily encoded into the network structure. The difficulty in providing something close to a complete representation is the *implicit* information that are hidden in the user produced data within the network or associated services / devices. These include unannotated visual data that may contain individuals, audio or video streams that indicate actions by individual users, geoalignment or tachometric data available to client applications of social networks on devices that the users carry, connectivity data from hardware sources like networked devices, peripherals, and near field communication protocols, and a number of other possibilities. The largest source of data which may contain *implicit* social network information is likely linguistic, copious amount of which is contained in each social network of significant size. Some of these networks (web-forums, Twitter) would be more specialized in linguistic communication, but the use of language in on-line social networks is nearly universal. It is here that a great deal of *implicit* information about the social network can be inferred. The extraction of this information, such as those semantically deep relations discovered using techniques in this study, result in a useful form for mobile services and social network applications.

References

Abe, S., Inui, K., & Matsumoto, Y. (2008). Two-phrased event relation acquisition: Couplingthe relation-oriented and argument-oriented approaches, *Proceedings of the 22nd International Conference on Computational Linguistics (COLING2008)*, pp. 1 – 8

Alur, R., & Chaudhury, S. (2006). Branching pushdown tree automata, *Proceedings of the 26th International Conference on the Foundation of Software Technology and Theoretical Computer Science (FSTTC)*, pp 393 - 404, Springer Verlag, Berlin / Heidelberg

Alvarez, M. P. (2009). The four causes of behavior: Aristotle and Skinner, *International Journal of Psychology and Psychological Therapy*, Volume 9, Number 1, pp 45 - 57

Alvarez, M. P., & Montes, G. (2006). Person, behavior and contingencies (an aesthetic view of behaviorism), *International Journal of Psychology*, Volume 41, pp 449 - 461, International Union of Psychological Sciences

Alvarez, M. P., & Sass, L. (2008). Phenomenology and behaviorism: a mutual readjustment, *Philosophy, Psychiatry and Psychology*, Volume 15, Association for the Advancement of Philosophy and Psychiatry (AAPP)

Ἀριστοτέλης. (before 350 B.C.) *φυσικὴ ἀκρόασις* (Physics); Translated by R. P. Hardie and R. K. Gaye, MIT Press 1994

Ἀριστοτέλης. (before 322 B.C.) *τὰ μετὰ τὰ φυσικά* (Metaphysics); Translated by J. Verner. Oxford Classical Texts: Oxford University Press, 1957

Baker, J. E. (1985). Adaptive selection methods for genetic algorithms, In *International Conference on Genetic Algorithms and Their Applications*, pp. 101 - 111

Baker, J. E. (1989). *An Analysis of the Effects of Selection in Genetic Algorithms*, Ph.D. Thesis, Vanderbilt University, Nashville

Bally, C.. (1920). *Impressionisme et grammaire, Mèlanges d'histoire littéraire et de philology offerts à M. Bernard Bouvier*, 261-279. Geneva: Sonor

Bargh, J. A. & Chartrand, T. L. (1999). The unbearable automaticity of being, *American Psychologist*, Issue 54, pp. 462 – 479, American Psychological Association

Barricelli, N. A. (1962). Numerical testing of evolution theories : Part I Theoretical introduction and basic tests, *Acta Biotheoretica*, Issue 16 (1-2), pp. 69 – 98

Bungum, L. & Gambäck, B. (2010). Evolutionary algorithms in natural language processing, *Proceedings of Norwegian Artificial Intelligence Symposium (NAIS).* Tapir Akademisk Forlag,

Baum, L. E., & Eagon, J. A. (1966) An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology, *Bulletin of the American Mathematical Society*, Volume 37, Number 6, pp. 1554 - 1563

Baum, L. E., & Petrie, T. (1967). Statistical Inference for Probabilistic Functions of Finite State Markov Chains, *the Annals of Mathematical Statistics*, Volume 73, Number 3, pp. 360 - 363

Baum, L. E., Petrie T., Soules, G., & Weiss, N. (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains, *the Annals of Mathematical Statistics*, Volume 41, Number 1, pp. 164 - 171

Beamer, B., & Girju, C. R. (2009). Using a bigram event model to predict causal potential, *In Proceedings of Conference on intelligent text processing and computational linguistics 2009*, Mexico City, Feb 24-28

Berthard, S. & Martin, J. H. (2008). Learning semantic links from a corpus of parallel temporal and causal relations, *Proceedings of ACL-08: HLT, Short Papers (Companion Volume)*, pp. 177 – 180, Association for Computational Linguistics

Bethard, S., Corvey, W., Klingenstein, S., Martin, J. H.. (2008). Building of a temporal-causal structure, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 08)*

Boas, H. (2007). Construction grammar in the 21st Century, *English Language and Linguistics 11 volume 3*, pp. 569 - 585

Bodländer, H. (1988). Dynamic programming on graphs with bounded tree-width *Proceedings of 15th International Colloquium on Automata, Languages, and Programming, volume 317 of Lecture Notes in Computer Science*, pp 105–118. Springer-Verlag, Berlin Heidelberg

Bora, T. C., Lebensztajn, L., Coelho, L. D. S.. (2012). Non-Dominated Sorting Genetic Algorithm Based on Reinforcement Learning to Optimization of Broad-Band Reflector Antennas Satellite, *IEEE Transactions in Magnetics*, Volume 48, Number 2, Supplement 4, Part 1, pp. 767-770, Piscataway, NY

Brennenstuhl, W. (1975). *Handlungstheorie und Handlungslogik* Scriptor Verlag, Krŏnberg

Brennenstuhl, W., & Wachowicz, K. (1976). On the pragmatics of control, *Proceedings of the 2nd annual meeting of the Berkley Linguistics Society*, pp. 396–397. Berkeley Linguistics Society, Berkeley CA

Brindle, A. (1981). *Genetic algorithms for function optimization*, Technical Report, pp. 81 - 82, University of Alberta, Canada

British National Corpus, (2007) University of Oxford Press, Longman Publishing, W

& R Chambers Publishing, in conjunction with British Library, University of Oxford, and Lancaster University, url:http://www.natcorp.ox.ac.uk/ , 2007

Brown, P. & Levinson, S. C. (1987). *Politeness: Some universals in language usage*, Cambridge University Press, Cambridge

Brown, P. F., de Souza, P. V., Mercer, R. L., del la Pietra V. J., & Lai, J. C.. (1992). Class-based n-gram models of natural language, *Journal of Computational Linguistics*, Volume 18, Issue 4, pp. 467 - 479

Burks, A. (1951). The logic of causal propositions, *Mind (1951) LX (239)*, pp 363 - 382, Oxford Journals on behalf of Mind Association, Oxford University Press

Bush, G., Luu, P., Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex, *Trends in Cognitive Sciences*, Issue 4, pp 215 – 222

Butt, M. & Ahmed, T. (2007). Non-canonical argument marking: beyond volitionality, specificity and animacy, *Proceedings from Workshop on Case, Word Order and Prominence*, Nijmegen

Carlson, M., Marcu, D., & Orkurowski, M. E. (2003). Building a discourse tagged corpus in the framework of rhetorical structure theory, In J. Kuppevelt & R. Smith (Ed.) *Current directions in discourse and dialogue*, pp. 85 - 112, Kluwer Academic Publishers

Cartwright, N. (1989). Causal laws and effective strategies, *Noûs 13*, pp 419 - 237

Caston, V. (2006). Aristotle's Psychology, In M. Gill & P. Pellegrin (Ed.) *A Companion to Ancient Philosophy.* Wiley-Blackwell Publishing, Hoboken, pp. 316 - 46

Chakraborty, B., & Chaudhuri, P. (2003). On The Use of Genetic Algorithm with Elitism in Robust and Nonparametric Multivariate Analysis, *Austrian Journal of Statistics*, Volume 32

Chang, D. S., & Choi, K. S. (2006). Incremental cue phrase learning and bootstrapping method for causality extraction using cue phrase and word pair probabilities, *Information Processing and Management*, Issue 42, Volume 3, pp 662 – 678

Chomsky, N. (1981) *Lectures on government and binding*, Dordrecht Foris

Chudasama, C., Shah, S. M., Panchal, M. (2011). Comparison of parents selection methods of genetic algorithm for TSP, *International Conference on Computer Communication and Networks CSI- COMNET-2011*, published by International Journal of Computer Applications (IJCA)

Cohen, J. (1960) A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, Volume 20, Issue 1, pp 37 – 46

Cole, P. (1983). The grammatical role of the causee in universal grammar, *International Journal of American Linguistics 49*, pp 115 - 133

Comrie, B. (1976) *The syntax of causative constructions: cross-linguistic similarities and divergences*, Academic Press

Cook, K. & Emerson, R. (1978). Power, equity, and commitment in exchange net-works.

American Sociological Review 43:721–39

Coppock, E. (2009). *The Logical and Empirical Foundations of Baker's Paradox*, Dissertation, Linguistics, Stanford University

Cortez, C. (1995). Argument structure: English verbs of suiting, *Proceedings of Revista Alicantina de Estudios Ingleses 8*, pp 69 - 78

Cramer, N. L. (1985). A Representation for the adaptive generation of simple sequential programs, *Proceedings of International Conference on Genetic Algorithms and their Applications*, Carnegie-Mellon University

Croft, W. A. (1991). *Syntactic categories and grammatical relations*, University of Chicago Press, Chicago IL

Damasio, A. R. (1994). The brain binds entities and events by multiregional activation from convergence zones, In H. Gutfreund & G. Toulouse (Ed.) *Biology and computation: A physicist's choice. Advanced series in neuroscience*, Volume 3, pp. 749 – 758, World Scientific Publishing, River Edge, NJ

Darwin, C. (1859). *on the Origin of species: (by Means of Natural Selection, The Preservation of Favoured Races in the Struggle for Life)* John Murray Publishing

Davidson, D. (1967). The logical form of action sentences, In N. Rescher (Ed.) *The logic of decision and action*, University of Pittsburgh Press, Pittsburgh

Davidson, D. (1980). *Essays on Actions and Events*, Oxford: Clarendon Press

Decadt, B., Hoste, V., Daelemans, W., & den Bosch, A. V. (2004). GAMBL, Genetic algorithm optimization of memory-based WSD. In R. Mihalcea & P. Edmonds (Ed.) *Proceedings of the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (Senseval-3)*, pp. 108 – 112

Delmonte, R., Bristot, A., Piccolino Boniforti, M. A., & Tonelli, S. (2007) Entailment and anaphora resolution in RTE3, *Proceedings of ACL Workshop on Text Entailment and Paraphrasing*, pp. 48-53, Prague, Association of Computational Linguistics

Delmonte, R., Tonelli, S., & Tripodi, R. (2009). Semantic processing for text entailment with VENSES, *Proceedings of the TAC 2009 Workshop on TE*, Gaithersburg, Maryland

Do, Q. X., Chan, Y. S., Roth, D. (2011). Minimally supervised event causality identification, *Proceedings of Empirical methods in natural language processing 2011*, Special Interest Group Linguistic Data: ACL

Dotlačil, J., Nilson, O. (2008). Null theory of long-distance reciprocity in English, *proceedings of NELS 39*, Ithaca NY

Dowty, D. (1979). *Word meaning and Montague grammar: the semantics of verbs and times in generative semantics and in Montague's PTQ*, D. Reidel Publishing Company, Dordrecht Netherlands

Dowty, D. (1991). Thematic protoroles and argument selection, *Proceedings of LSA (Lin-*

guistic Society of America)

Dryer, M. (1986). Primary objects, secondary objects, and antidative, *Language 62* pp808 - 845, Linguistic Society of America

Edgington, D. (1997). Review: Mellor on chance and causation, *British Journal of Philosophy and Science, Issue 48*, pp 411 - 433, Oxford University Press, Oxford

Everaert, M., Marelj, M., & Siloni, T. (2012). *The theta system: Argument structure at the interface*, Oxford Studies in Theoretical Linguistics, Oxford University Press

Fehr, E., & Găchter, S. (1998). Reciprocity and economics: the economic implications of *homo reciprocants*; *Fourty Second European Economic Review*, pp. 845 - 859

Fehr, E., & Găchter, S. (2000). Cooperation and punishment in public goods experiments; *American Economic Review*, Issue 90, pp. 980 - 994

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

Fillmore, C. (1968). The case for case, In B. Emmon; H. Emmon (Ed.) *Universals in linguistic theory*, Holt Rinehart and Winston Publishing

Fillmore, C. (1970). The Grammar of hitting and breaking, In R. A. Jacobs & P. S. Rosenbaum (Ed.) *Readings in English Transformational Grammar* pp 120 - 133, Ginn Publishing, Waltham MA

Fillmore, C. (1971). Types of lexical information, In D. Steinberg & L. Jacobovitz (Ed.) *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology*, Cambridge University Press

Fillmore, C. (1984). Frames and the semantics of understanding, *Quanderi di Semantica 6*, pp 222 - 254

Fillmore, C., Baker, C. (2009). A frames approach to semantic analysis, In B. Heine & H. Narrog (Ed.), *The Oxford Handbook of Linguistic Analysis* Oxford University Press

Fleck, D. (2002). Causations in Matses, *Grammar of causation and interpersonal manipulation*, edit: M. Shibatani, John Benjamins Publishing

Fleischer, N. (2006). The origin of passive *get*, *English Language and Linguistics, Volume 10, Issue 2*, pp 225 - 252, Cambridge University Press

Fleischer, N. (2008). Passive get, causative get, and the phasehood of passive VP, In R. L. Edwards, P. J. Midtlyng, C. L. Sprague, K. G. Stensrud (Ed.) *Proceedings of CLS 41*, pp 59 - 67, Chicago Linguistic Society, Chicago, IL

Fodor, J. (1970). Three Reasons for Not Deriving Kill from Cause to Die, *Linguistic Inquiry 1*, pp. 429 – 438

Fogel, D. B. (1998). Evolutionary computation: the Fossil record, IEEE Press, Piscataway, NJ

Gessner, S. (2001). Object marking and agentivity in Navajo causatives, *Proceedings of the Annual Meeting of Berkeley Linguistic Society, Volume 27*, University of California at

Berkeley Publishing

Girju, C. R. (2003). Automatic detection of causal relations for question answering, *The 41st Annual Meeting of the Association for Computational Linguistics (ACL 2003), Workshop on Multilingual Summarization and Question Answering - Machine Learning and Beyond*

Givón, T. (2001). *English grammar: A function based introduction*, John Benjamins Publishing, Amsterdam, Netherlands

Givón, T. (2001). *Syntax: an introduction*, John Benjamins Publishing, Amsterdam, Netherlands

Givón, T. (2008). The ontogeny of complex verb phrases: How children learn to negotiate fact and desire, In T. Givón, M. Shibatani (Ed.) *Syntactic complexity: Diachrony, acquisition, neuro-cognition, evolution . 2009. vi*, pp. 311–388, John Benjamins Publishing, Amsterdam, Netherlands

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument Structure*, University of Chicago Press, Chicago IL

Goldberg, A. (2002). Surface generalizations: An alternative to alternations, *Cognitive Linguistics 13*, pp 327 - 356, Walter de Gruyter

Goldberg, A. (2009). The nature of generalization in language, *Cognitive Linguistics 20*, pp 93 – 127, Walter de Gruyter

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing, Reading MA

Goldberg, D. (1990). A note on Boltzmann tournament selection for genetic algorithms and population-oriented simulated annealing, *Complex Systems 4*, Volume 4, pp. 445 - 460

Goldberg, D. & Deb, K. (1993). A comparative analysis of selection schemes used in genetic algorithms, *Foundations of Genetic Algorithms*, pp. 69 - 93

Goldberg, D., Deb, K., & Clark, J. (1991). Genetic algorithms, noise, and sizing of population, *Complex System*, Volume 6, pp. 333 - 362, 1991

Graffin, N. F., Ray, W. J., & Lundy, R. (1995). EEG concomitants of hypnosis and hypnotic susceptibility, *Journal of Abnormal Psychology*, Issue 104, pp. 123 – 131

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*, Wiley Publishing, New York

Gruber, J. (1976). Lexical structures in syntax and semantics, *North Holland Linguistic Series*, Volume 25. North-Holland, Amsterdam

Guessarian, I. (1983). Pushdown tree automata, *Math System Theory 16, Volume 4*, pp 237 - 263

Guigno, R. (2002). *Searching algorithms and data structures for combinatorial, temporal, and probabilistic databases*, Dottore di Ricerca (Ph. D Thesis), Universitá Degli Studi di Catania, Catania Italy

Guigno, R., & Shasha, D. (2002). Graphgrep: A fast and universal method for querying graphs, *Proceedings of the Fifteenth International Conference on Pattern Recognition (ICPR 2002)*, IEEE Computer Society

Project Gutenberg, Hart, M. (2005). *Gutenberg corpus*, http:www.gutenberg.org

Haegeman, L. (1985). The get passive and Burzio's generalization, *Lingua (An International Review of General Linguistics) 66*, pp 53 – 77, Elsevier Publishing

Hale, K. & Keyser, S. J. (2002). Prolegomenon to *a Theory of Argument Structure*, The MIT Press, Cambridge, Massachusetts

Hall, B. (1965). *Subject and Object in English*, PhD dissertation, MIT

Hardcastle, V. G. (1991). *Partitions, probabilistic causality, and Simpson's paradox*, *Synthese, Volume 86, Number 2*, pp 209 - 228, Springer Publishing

Harley, H. B. (1995). *Subjects, events, and licensing*, Doctor of Philosophy Thesis, Department of Linguistics and Philosophy, Massachusetts Institute of Technology

Harley, H. B. (2004). *Wanting, having and getting*, *Linguistic Inquiry 35*: pp 255 – 267, MIT Press Journals

Hashimoto, C., Torisawa, K., Kloetzer, J., Sano§Istvan, M., Oh, V. J. H., & Kidawara, Y. (2014). Toward future scenario generation: extracting event causality exploiting semantic relation, context, and association features, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*

Haspelmath, M.. (1993) More on the typology of inchoative/causative verb alternations, *Causatives and Transitivity*, eds B. Comrie & M. Plinsky. John Benjamin Publishing

Hobbes, J. R. (2005). Toward a useful concept of causality for lexical semantics, In C. Condoravdi and S. Kaufmann (Ed.) *Journal of Semantics*, Special Issue on Modality and Temporality,

Hogan, J. A. (1994). The concept of cause in the study of behavior, In J. A. Hogan & J. J. Bolhuis (Ed.) *Causal mechanisms of behavioral development*, pp 3 - 15, Cambridge University Press, Cambridge

Holland, J. D. (1975). *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor, MI

Hollmann, W. (2003). *Passivity of English periphrastic causatives*, Doctoral dissertation, University of Machester, Machester

Holmes, J. (1999). *The syntax and semantics of causative verbs*, Thesis, University College London

Irwin, T. (1998). *Aristotle's first principles*, Clarendon Press, Oxford University Press

Jackendoff, R. (1972) *Semantic interpretations in generative grammar*, MIT Press, Cambridge Mass

Jackendoff, R. (1987). The status of thematic relations in linguistic theory, *Linguistic Inquiry Issue 18, Volume 3*, pp. 369 – 411

Jackendoff, R. (1990). *Semantic structures*, The MIT Press, Cambridge Mass

de Jong, K. A. (1975). *an Analysis of the behavior of a class of genetic adaptive systems*, Ph. D. Dissertation, University of Michigan Press, Ann Arbor, MI

Joshi, A., Prasad, R., Webber, B. (2006, July). Discourse annotation: Discourse connectives and discourse relations, *Tutorial at the Association for Computational Linguistics*, Sydney, Australia.

Karimi, K. (2010). A brief introduction to temporality and causality, *Proceedings of CoRR*

Kay, P., & Fillmore, C. (1999). Grammatical Constructions and Linguistic Generalizations: "What's X doing Y?" Construction *Language 75*, pp. 1 - 34

Kemmer, S., & Verhagen, A. (1994). The grammar of causatives and the conceptual structure of events, *Cognitive Linguistics 5-2*, pp. 115 - 156

Khoo, C., Myaeng, S., Oddy, R. (2001). Using cause-effect relations in text to improve information retrieval precision, *Information Processing and Management* Issue 37, pp. 119 – 145

Killeen, P. R. (2001). The four causes of behavior, *Current Directions in Psychological Science*, Volume 10, pp. 136 - 140

Killeen, P. R., & Nash, M. R. (2003). The four causes of hypnosis, *The International Journal of Clinical and Experimental Hypnosis*, Volume 51, pp. 195 - 231

Kirkpatrick, S., Gelatt, J. C. D., Vecchi, M. P. (1983). Optimization by simulated annealing, *Science*, Volume 220, pp. 671 - 680

Kiryakov, A., Popov, B., Terziev, I., Manov, D., Ognyanoff, D. (2004). Semantic annotation, indexing, and retrieval, *Web Semantics: Science, Services and Agents on the World Wide Web*, Volume 2, Issue 1, pp. 49 – 79

Kjell, J. S. (2001). *An analysis of the anticausative alternation*, Working report online: http://folk.uio.no/kjelljs/

Kodama, K. (2004). The English caused-motion construction revisited – a cognitive perspective, *Papers in Linguistic Science 10*, pp. 41 - 54

Koontz-Garboden, A. (2009). Anticausativization, *Proceedings of Natural Language and Linguistic Theory 27*, pp. 77 – 138

Kremer, U. (1999). Ergebnisse der wissenschaftlichen Begleitung und Auswertungvon regionalwirtschaftlichen Vorhaben, *Rahmen und im Umfelddes Projekts Regionalwirtschaftliche Kooperation und arbeitsorientierteStrukturpolitik in Nordrhein-Westfalen (REKON)*. ISAConsult, Bochum

Lakoff, G. (1965) *On the nature of syntactic irregularity*, Doctoral dissertation, Indiana University

Lakoff, G., & Peters, S. (1969). Phrasal conjunction and symmetric predicates, *In Modern studies in English: Readings in transformational grammar 113-142*, eds. D. Reibel & S. Schane. Englewood Cliffs, NJ

Langacker, R. (1988). An overview of cognitive grammar, In B. Rudzka-Ostyn (Ed.) *Topics in Cognitive Linguistics*, pp 3 - 48, John Benjamins Publishing, Amsterdam Philadelphia

Langacker, R. (1991). Transitivity, case and grammatical relations, *Concept, Image and symbol: The Cognitive Basis of Grammar*, pp 209 - 260 Berlin, New York: Mouton de Gruyter

Langacker, R. (2002). *The cognitive basis of grammar: Concept, Image, and Symbol*, 2nd edition, Mouton de Gruyter, Berlin, New York

Lauer, S. (2010). *Periphrastic causative verbs in English: What do they mean? The expression of causal necessity and causal sufficiency in ordinary English*, Department of Linguistics Qualifying Paper, Stanford University

Leech, G. N. (1983). *Principles of Pragmatics*, Longman Publishing, New York

Lewis, D. K. (1973). *Counterfactuals*, Harvard University Press, Cambridge, MA

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*, University of Chicago Press, Chicago IL

Levin, B., & Rappaport-Hovav, M. (1994). A preliminary analysis of causative verbs in English, *Lingua 92*, pp. 35 – 77

Levin, B., & Rappaport-Hovav, M. (1995). *Unaccusativity: At the syntax–lexical semantics interface*, MIT Press, Cambridge MA

Lewis, D. K. (1979). Counterfactual dependence and the arrow of time, *Noûs, Volume 13*, pp. 455 - 467

Lewis, D. K. (1986). Counterfactual dependence and time's arrow, *Philosophical papers, Volume II*, Oxford University Press, Oxford

Lewis, D. K. (2000). Causation as influence, *Journal of Philosophy 97*, pp. 182 - 198, The Journal of Philosophy Inc.

Li, C., & Girju, C. R. (2010). *Language and social networks in user generated on-line data*, Technical Report of University of Illinois at Urbana-Champaign

Lin, Z., Kan, M. Y., & Ng, H. T. (2009, August 6 - 7). Recognizing implicit discourse relations in the Penn Discourse Treebank, *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 343 – 351, Singapore

Mihaila, C., Ohta, T., Pyysalo, S., Ananiadou, S. (2013). Biocaus: Annotating and analyzing causality in the biomedical domain, *BMC Bioinformatics 2013*, The National Center for Text Mining, School of Computer Science, The University of Manchester, UK

Maldonado, R. (1988). Energetic Reflexives in Spanish, *Berkeley Linguistics Society 14*, pp. 153 - 165

Maldonado, R. (2004). Basic voice patterns in Tarascan (P'orhepecha), In M. Achard & S. Kemmer (Ed.) *Language, Culture and Mind*, CSLI Publications

Maldonado, R. (2007). Soft causatives in Spanish, In B. Cornillie & N. Cornillie (Ed.) *From Action and Motion to Transitivity and Causality. On interpreting Construction Schemas*, Mouton de Gruyter, Berlin

Maldonado, R. (2011). Reciprocal constructions in Olutec, In N. Evans, A. Gaby, S. Levinson, A. Majid (Ed.) *Reciprocals and semantic typology*, pp. 265 - 276, John Benjamins Publishing

Mandelblit, N. (1997). Grammatical Blending: Creative and Schematic Aspects in Sentence Processing and Translation (Doctoral Thesis in Cognative Science), *Part Two: A grammatical blending account of Hebrew binyanim, grammatical blending, creative and schematic aspects.* University of California, San Diego

Mandelblit, N. (2000). The grammatical marking of conceptual integration: From syntax to morphology, *Cognitive Linguistics*, Issue 11 (3 / 4), pp. 197 - 251

Mann, T. P. (2006). Numerically stable Hidden Markov Model implementation

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit, *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55 - 60

Mashohor, S. (2005). Elitist selection schemes for genetic algorithm based printed circuit board inspection system, *Evolutionary Computation, 2005. The 2005 IEEE Congress*, Volume 2, pp. 974 - 978

de la Maza, M., & Tidor, B. (1992). Increased flexibility in genetic algorithms: The use of variable Boltzmann selective pressure to control propagation, *New ORCA CSTS Conference: Computer Science and Operations Developments in Their Interfaces*, pp. 425 - 440

de la Maza, M., & Tidor, B. (1993). An analysis of selection procedures with particular attention paid to proportional and Boltzmann selection, *Proceedings of Fifth International Conference on Genetic Algorithms.* Morgan Kaufmann Publishing, San Mateo

McCawley, J. D. (1978). *Conversational implicature and the lexicon*, In P. Cole (Ed.) *Syntax and Semantics 9: Pragmatics*, pp. 245 - 259. Academic Press

McKoon, G., & MacFarland, T. (2000). Externally and Interally Caused Change of State Verbs, *Language 76*, pp. 833 – 858

Mellor, D. H. (1995). *The facts of causation*, Routledge, London

Menzies, P. (2009). Counterfactual theories of causation, In E. N. Zalta (Ed., web published) *Stanford Encyclopedia of Philosophy, Fall 2009*, plato.stanford.edu/entries/causation-counterfactual , Stanford University Department of Philosophy

Miltsakaki, E., Prasad, R., Joshi, A., & Webber, B. (2004). The Penn Discourse TreeBank, *Proceedings of the Language Resources and Evaluation Conference*, Lisbon, Portugal

Molm, L. (2010). The structure of reciprocity, *Social Psychology Quarterly* June 2010 Volume 73 Number 2, pp. 119 - 131

Moufatich, F. E. (2008). lowest common ancestor, *Course 1: Trees - The Ubiquitous Structure in Computer Science and Mathematics*, Joint Advanced Student School (JASS Workshops), St. Petersberg

Murray, S. (2007). Dynamics of reflexivity and reciprocity, In M. Aloni, P. Dekker, F. Rölofsen (Ed.) *Proceedings of the Sixteenth Amsterdam Colloquium*

Murray, S. (2008). Reflexivity and reciprocity with(out) under-specification, In A. Grønn (Ed.) *Proceedings of Sinn und Bedeutung 12*, Oslo: ILOS

Nakao, C. (2002). *A Study on Japanese Reciprocal Constructions*, MA thesis of Department of Linguistics, University of Tokyo

Ogiermann, E. (2009). Politeness and indirectness across cultures: A comparison of English, German, Polish, and Russian requests, *Journal of Politeness Research, Issue 5*, pp 189 - 216, Walter de Gruyter

Oh, J. H., Torisawa, K., Hashimoto, C., Sano, M., de Saeger, S., Ohtake, K. (2013). Why-question answering using intra- and inter-sentential causal relations. *Proceedings of the 51st Annual Meeting of the Association for Com-putational Linguistics (ACL 2013)*, pp. 1733 – 1743

Otto, E., & Rojas, M. C. R. (2007). EDA: An Evolutionary Decoding Algorithm for Statistical Machine Translation, *Applied Artificial Intelligence*, Issue 21, Volume 7, pp. 605 – 621

Pakray, P., Neogi, S., Bhaskar, P., Poria, S., Bandyopadhyay, S. (2010). A Textual Entailment System using Anaphora Resolution, *Proceedings of 3rd International Conference on ICACTE*, Volume 6, Advanced Computer Theory and Engineering

Parsons, T. (1979). *An analysis of mass terms and amount terms*, In F. J. Pelletier (Ed.) *Mass terms, Some philosophical problems*, pp 137 - 166, Reidel, Dordrecht

Parsons, T. (1985). Underlying events in the logical analysis of English, In E. LePore et. al. (Ed.) *Actions and events: Perspectives on the philosophy of Donald Davidson*, MIT Press, Cambridge, MA

Parsons, T. (1990). *Events in the semantics of English: A study in subatomic semantics*, Massachusetts Institute of Technology

de Pascalis, V., Ray, W. J., Tranquillo, I., D'Amico, D. (1998). EEG activity and heart rate during recall of emotional events in hypnosis: Relationships with hypnotizability and suggestibility, *International Journal of Psychophysiology*, Issue 29, pp. 255 – 275

Paul, M., Girju, C., Li, C. (2009). Mining the web for reciprocal relationships *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL)*, Boulder CO

de Pauw, G. (2003). Evolutionary computing as a tool for grammar development, *Proceedings of the 2003 International Conference on Genetic and evolutionary computation*, pp. 549 – 560, Springer-Verlag, Berlin Heidelberg

Pearl, J. (1999). Probabilities of causation: Three counterfactual interpretations and their identification, *Synthese: An International Journal for Epistemology, Methodology, and Philosophy of Science*

Pearl, J. (2000). *Causality: Models, reasoning, and inference*, Cambridge University Press, New York

Pitler, E., Louis, A., & Nenkova, A. (2009, August 2 - 7). Automatic sense prediction in discourse relations in text, *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, pp. 683 – 691, Suntec, Singapore

Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure*, MIT Press, Cambridge MA

Piñón, C. (2001). Modeling the causative-inchoative alternation, *Linguistische Arbeitsberichte 76*, pp. 273 - 293

Piñón, C. (2001). A Finer look at causative-inchoative alternation, *Proceedings of Semantics and Lingtuistic Theory*

Powers, D. (1998). Applications and explanations of Zipf's law, *Association for Computational Linguistics*, pp. 151 – 160

Prasad, R., Miltsakaki, E., Joshi, A., Webber, B. (2004). Annotation and Data Mining of the Penn Discourse TreeBank, *Proceedings of the ACL Workshop on Discourse Annotation*, Barcelona Spain

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E, Robaldo, L, Joshi, A., Webber, B. (2007). The Penn Discourse Treebank 2.0, *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC).* Marrakech, Morroco

Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L., Webber, B. (2007). *The Penn Discourse Treebank 2.0 Annotation Manual*, PDTB Research Group, University of Pennsylvania, University of Torino, University of Edinburgh

Prasad, R., Webber, B., Joshi, A. (2014). Reflections on the Penn Discourse TreeBank, comparable corpora and complementary annotation, *Association for Computational Linguistics*, MIT Press, Cambridge MA

le Priol, F., Djioua, B., Garcia, D. (2007). Automatic annotation of discourse and semantic relations supplemented by terminology extraction for domain ontology building and information retrieval, *FLAIRS Conference*, pp. 374 - 379. AAAI Press

Rabiner, L. R. (1989). A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE 77*

Rabiner, L. R. & Juang, B. H. (1993). *Fundamentals of speech recognition.* Englewood Cliffs, Prentice Hall, New Jersey

Rachlin, H. (1992). Teleological behaviorism *American Psychologist*, Volume 47, pp 1371 - 1382, American Psychological Association

Radinsky, K., Davidovich, S., Markovitch, S.. (2012). Learning causality for news-events prediction, *Proceedings of International World Wide Web Conference 2012 (WWW 2012)*, pp. 909 – 918

Radinsky, K., & Horvitz, E. (2013). Mining the web to predict future events, *Proceedings of Sixth ACM International Conference on Web Search and Data Mining (WSDM 2013)*, pp. 255 – 264

Rappaport-Hovav, M., & Levin, B. (1988) What to do with theta-roles, In W. Wilkins (Ed.) *Thematic Relations, Syntax and Semantics*, pp 21: 7 -36, Academic Press

Rappaport-Hovav, M., & Levin, B. (2011) Lexicon Uniformity and the Causative Alternation, In M. Everaert, M. Marelj, and T. Siloni (Ed.) *The theta system: argument structure at the interface*, pp. 150-176, Oxford University Press, Oxford UK

Ray, W. J. (1997). EEG concomitants of hypnotic susceptibility, *International Journal of Clinical and Experimental Hypnosis*, Issue 45, pp. 301 – 313

Rechenberg, I. (1973). *Evolutionsstrategic: Optimierung technicher System nach Principen der Biologischen Evolution*, Frommmann-Holzboog Verlag, Stuttgart

Reichenbach, H. (1947). *Elements of symbolic logic*, Free Press, New York

Reinhart, T. (1996). Syntactic effects of lexical operations: reflexives and unaccusatives, *OTS Working papers in Linguistics, TL-97-002*, University of Utrecht, 1996

Reinhart, T. (2000). The theta system: syntactic realization of verbal concepts, *OTS Working papers in linguistics*

Reinhart, T. (2002). The theta system: an overview, *Theoretical Linguistics 28*, pp229-290, edit: M. Krifka, Mouton de Gruyter, 2002.

Reinhart, T., & Siloni, T. (2003). Thematic arity operations and parametric variations, *OTS working papers in linguistics, TL-03-001*, University of Utrecht

Riaz, M., & Girju, C. R. (2010). Another look at causality: Discovering scenario-specific contingency relationships with no supervision, *In Proceedings of ICSC (CERN School of Computing)*

Richards, N. (2001). An idiomatic argument for lexical decomposition *Linguistic Inquiry Volume 32*, pp 183–192, MIT Press

Robins, J. M., & Greenland, S. (1989). The probability of causation under a stochastic model for individual risk, *Biometrics, Volume 24*, pp 1125 - 1138, International Biometric Society

Rodriguez, L., Garcia-Varea, I., & Gamez, J. A. (2008). On the Application of Different Evolutionary Algorithms to the Alignment Problem in Statistical Machine Translation, *Neurocomputing*, Volume 71(4 - 6), pp. 755 – 765

Rozwadowska, B. (1987). Are thematic relations discrete? , In R. Corrigan, E. R. Eckman & M. Noonan (Ed.) *Linguistic Categorization: Proceedings of an International Symposium in Milwaukee*, pp 115 - 130, Wisconsin

Russell, B. (1948). *Human Knowledge: Its Scope and Limits*, Taylor & Francis

Saad, G. N., & Bolozky, S. (1980). Theoretical implications of morphological causitivization in Arabic and Hebrew, *In Proceedings of North American Conference on Afroasiatic Linguistics, April 13-14*, San Francisco CA

Saad, G. N., & Bolozky, S. (1984). Causitivization and transitivization in Arabic and Modern Hebrew, *Afroasiatic Linguistics 9*

Sachs, J. (2001) *Aristotle's On the Soul and On Memory and Recollection*, Green Lion Press

Sagae, K., & Tsujii, J. (2007). Dependency parsing and domain adaptation with LR models and parser ensembles, *Proceedings of the CoNLL 2007 Shared Task. Joint Conferences on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL'07).* Prague, Czech Republic

Salmon, W. C. (1980) Probabilistic causality, In M. Tooley (Ed.) *Causation, in Oxford Readings for Philosophy*, pp 137 - 153, Oxford University Press UK

Sasaki, T. (1987). *On the aspectual opposition of passives in modern Hebrew*, Kyoto University Research and Information Repository

Schimpf, K. M., & Gallier, J. H. (1985). Tree pushdown automata, *Journal of Computer Sciences 30, Volume 1*, pp. 25 - 40

Schaffer, J. (2004). Counterfactuals, causal independence and conceptual circularity, *Analysis 64, Volume 4*, pp. 299 - 309

Scheffler, U. (1992[a]). Events as shadow entities, *Proceedings of Sixth International Symposium LOGICA '92*, Institute of Philosophy of CSAS, Praha, Czechoslovakia

Scheffler, U. (1992[b]). On the logic of event causation: Part I, Fundamental reflections, *Konstanzer Berichte zur Logik und Wissenshafttheorie, Volume 34*, Universität, Zentrum Philosophie und Wissenschaftstheorie

Schmidhuber, J. (1987). *Evolutionary principles in self-referential learning: On Learning now to learn: the meta-meta-meta...*, Doctoral Thesis, Technische Universität München, Germany

Shasha, D., Wang, J. T. L., Giugno, R. (2002, June). Algorithmics and applications of tree and graph searching, *Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*, Madison Wisconsin

Schimbera, J., Schimbera, P. (2010). *Determination des Indeterminierten: Kritische Anmerkungen zur Determinismus- und Freiheitskontroverse*, Verlag Dr. Kovac, 2010

Shibatani, M. (1976). The grammar of causative constructions: A conspectus, In M. Shibatani (Ed.) *Syntax and Semantics 6: The Grammar of Causative Constructions*, pp. 1 - 40, Academic Press, New York

Shibatani, M., & Pardeshi, P. (2002). The causative continuum, *The Grammar of Causation and Interpersonal Manipulation*, pp 85 - 126, 2002

Shields, D. (2010). Aristotle's Psychology, In E.N. Zalta (Ed.) *The Stanford Encyclopedia* *of Philosophy*

Simon, H. (1952). Logic of causal relations, *Cowell commission papers, New series, volume 70*, Cowell Commission for Research in Economics: University of Chicago, Chicago IL

Skinner, B. F. (1938). *The Behavior of organisms: An experimental analysis*, B.F. Skinner Foundation, Cambridge, Mass

Skinner, B. F. (1953). *Science and human behavior*, Macmillan Publishing, New York

Slavcheva, M (2007). Linking reflexive verb structure to verb meaning in a cross-lingual lexical setting *26th conference on Lexis and Grammar*, Bonifacio

Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search*, MIT Press, Cambridge Mass

Stalnaker, R. C. (1975). A theory of conditionals, In E. Sosa (Ed.) *Causation and conditionals*, Oxford Readings in Philosophy, Oxford University Press, London

Starr, W. B. (2010). *Computing counterfactual assumptions*, Web Published, part of Ph. D. Dissertation, The State University of New Jersey

Stefanowitsch, A. (2001). *Constructing causation: A construction grammar approach to analytic causatives*, Ph.D. dissertation, Rice University

Suppes, P. (1970). A probabilistic theory of causality, *Acta Philosophica Fennica, Volume 24*, North Holland Publishing

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

Talmy, L. (1996). The windowing of attention in language, In M. Shibatani & S. Thompson (Ed.) *Grammatical constructions: their form and meaning*, pp. 235 – 287, Oxford: Oxford University Press

Tarjan, R. E., & Vishkin, U. (1984). Finding biconnected components and computing tree functions in logarithmic parallel time, *Proceedings of FOCS*, pp. 12 - 20

Tian, J., & Pearl, J. (2000). Probabilities of causation: Bounds and identification, *Annals of Mathematics and Artificial Intelligence, Volume 28*, pp 287 - 313

Tullo, C., & Hurford, J. (2003). Modeling Zipfian distributions in language, In S. Kirby (Ed.) *Proceedings of Language Evolution and Computation Workshop at ESSLLI*, pp. 62 - 75, Vienna

Turing, A. (1950). Computing machinery and intelligence, *Mind: a Quarterly Review of Psychology and Philosophy*, Volume LIX, Number 236, pp. 433 - 460, The Mind Association

van Valin, R. D., & Wilkins, D. P. (1996). The case for 'effector': case roles, agents and agency revisited, In M. Shibatani & S. Thompson (Ed.) *Grammatical constructions: their form and meaning*, pp. 289 – 322, Oxford University Press, Oxford, UK

Veltman, F. (2005). Making counterfactual assumptions, *Journal of Semantics, Volume 22*, Oxford Journals: Humanities, Oxford University Press

Vendler, Z. (1957). Verbs and times, *The philosophical review 66*, pp 143 - 160, Duke

University Press

Vendler, Z. (1967). *Linguistics in philosophy*, Cornell University Press

Webber, B., & Joshi, A. (1998, August). Anchoring a lexicalized tree-adjoining grammar for discourse, *ACL/COLING Workshop on Discourse Relations and Discourse Markers*, Montreal, Canada

Weitzenhoffer, A. M. (1978). Hypnotism and altered states of consciousness, In A. Sugarman & R.E. Tarter (Ed.) *Expanding dimensions of consciousness*, pp. 183 – 225, Springer Verlag, New York

Welch, L. R. (2003). Hidden Markov Models and the Baum-Welch Algorithm, In L. C. Perez (Ed.) *IEEE Information Theory Society Newsletter December 2003*, IEEE Publishing

Wierzbicka, A. (1988). *The semantics of grammar*, John Benjamins Publishing, Amsterdam, Philadelphia PA

Wierzbicka, A. (1998). The semantics of English causative constructions in a universal-typological perspective, In M. Tomasello (Ed.) *The new psychology of language: cognitive and functional approaches to language structure*, pp 113 - 153, Lawrence Erlbaum Publishing, Mahwah NJ

Wierzbicka, A. (2002). English causative constructions in an ethnosyntactic perspective: Focusing on LET, In N. Enfield (Ed.) *Ethnosyntax*, pp 162 - 203, Oxford University Press, Oxford UK

Williams, E. (1989). The anaphoric nature of $\theta-$roles, *Linguistic Inquiry* Volume 20, pp 425 - 456, the MIT Press, Cambridge Mass

Williams, E. (1994). *Thematic structure in syntax*, The MIT Press, Cambridge MA

Williamson, J (2009). *Probabilistic theories of causality*, *Oxford Handbook of Causation*, Oxford University Press, Oxford UK

Wolff, P. (2003). Direct Causation in the Linguistic Coding and Individuation of Causal Events, *Cognition 88*: pp. 1 – 48

Wolff, P. (2007). Representing Causation, *Journal of Experimental Psychology, General*, pp 82 - 111, American Psychological Association

Wolff, P., Song, G., & Driscoll, D. (2002). Models of causation and causal verbs, In M. Andronis, C. Ball, H. Elston, S. Neuval (Ed.) *Papers from the 37th meeting of the Chicago Linguistics Society, Main Session, Volume 1*, pp. 607 – 622, Chicago Linguistics Society, Chicago IL

Wolff, P., & Zettergren, M. (2002). A vector model of causal meaning, *Proceedings of the twenty-fifth annual conference of the Cognitive Science Society*, Erlbaum Publishing, Hillsdale NJ

Wolff, P., & Song, G. (2003). Models of causation and the Semantics of casual verbs, *Cognitive Psychology*

Wright, S. K. (2001). *Internally Caused and Externally Caused Change of State Verbs*, PhD dissertation, Northwestern University, Evanston, IL, 2001.

Wright, S. K. (2002). *Transitivity and Change of State Verbs*, *Proceedings of BLS 28*, pp. 339 – 350

Yamaguchi, T. (1998). Lexical Semantic Analysis of Causative/Inchoative Alternation in Japanese: A preliminary investigation of subclasses of verbs, *Essex graduate student papers in language and linguistics, volume II*, Department of Language and Linguistics, University of Essex

Yaman, F., Yilmaz, A. E. (2012). Elitist genetic algorithm performance on the uniform circular antenna array pattern synthesis problem, *PRZEGLAD ELEKTROTECHNICZNY (Electrical Review)*

Yang, S. (2007). Genetic algorithms with elitism-based immigrants for changing optimization problems, *EvoWorkshops 2007, LNCS 4448*, pp. 627 – 636, Springer Verlag, Berlin

Yeo, S. (2005). Distinguishing periphrastic causatives from morphological causatives, *Studies in generative grammar, Volume 15*, pp. 231 - 249, Korean Generative Grammar Circle

Zavala, R. (2002). Olutec causatives and applicatives, In M. Shibatani (Ed.) *Grammar of causation and interpersonal manipulation*, John Benjamins Publishing

Zipf, G. K. (1949). *Human behavior and the principle of least effort*, Addison-Wesley Publishing, Cambridge Mass

Zubizarreta, M. L. (1992). The Lexical Encoding of Scope Relations among Arguments, *Syntax and the Lexicón: Syntax and Semantics 24. Ed. Eric Wehrli and Tim Stowell*, pp 251 - 279, Academic Press, New York

Appendix A: Annotation Directions

The following is the annotation guidelines used during the study for the causal modules. It consists of a brief version of the guidelines for quick reference during annotation; a full version of guidelines for initially familiarizing the annotator with the annotation procedure; and a set of guidelines for each of the semantic subclasses to enforce reasonably uniform understanding of the concept of semantically causal relation in language. We will use the embedded causal guidelines here, the adjunctive causal guidelines are the same with some small changes.

A.1 Abbreviated annotation guidelines

Here is the brief outline of the guild-lines of embedded causal structures. In the web-based version of the annotation tool, these are always present on a part of the annotation page:

1. Use the format 'XYi' (for each entry if some clear relation(s) exist) or 'N' (if no such relation exists)
2. Contextual information furnished when needed for appropriately annotating
    - [...] contains the information that completes the meaning of the segment
    - (...) contains contextual info outside of the segment that's highly relevant
3. Only potential relations BETWEEN segments (NOT within) are considered
4. Some relations may not be between immediately adjacent segments, but may 'skip over' one or more segments
5. Similarly, relations may not always occur in the straightforward order (maybe reversed, such as D $\xrightarrow{caus}$ B
6. Sometimes entire logical event sequences can occur in hypothetical or imaginary worlds

7. The sequential relations must be meaningful (not tautology) and positive (not negation, such as stating some relationship does not exist); and they need to be a real sequence of events/states/properties in some world (real or imaginary)

A.2 Full annotation guidelines

These are the guidelines in finding whether or not real-world relation(s) likely exist between individual segments, among a sequence of segments in text, discovering sequential relations among events/states/properties/etc. Each of these events/states/properties is represented by one of the segments in the sequence.

A.2.1 Annotation Format:

1. If some clear relation is found in the sequence, each annotation entry (each page) should be annotated as 'XYi':
   - the 'X' and 'Y' are the names of the segments, which should up in the sample on top as 'A', 'B', 'C', 'D', ...
   - the 'i' indicates the type of relation that this falls into (the overview of the types is on the left bottom of the page)
   - use one line for each entry, meaning hit 'enter' between any two entries in the text-box
2. If NO clear relation is found in the sequence:
   - Mark the text-box with 'X', then hit 'enter'

A.2.2 Markers of additional information:

1. things between square brackets [...] : additional content that can complete the forms of utterances/writings
   - — this can be viewed as components that makes the meaning of the segment whole
2. things between parentheses (...) : supplemental contents that can provide more explanation, or identify certain individuals in text
   - this supplemental information is for background only, and should NOT be construed as part of the meaning of the segment that contains it.

A.2.3 Segments of utterances:

The original text samples are divided into "segments", e.g., A, B, C, D, ... etc.

1. Each segment, as much as possible, has a self contained meaning; and when necessary, complete with supplemental content in [...] to form a complete meaning.
2. We should NOT include relations contained things within a single segment, such that the following is NOT causal between the segments (we will use '—' to indicate separation between segments):
    - "he had given her some interesting information, — that the expedition had given rise to the experimental results."
    - "The distance between the locales made traveling difficult, — which was said to be about 20 miles"
    - "The man walked by the street corner brusquely, — as if he failed to see her on the sidewalk"
3. We also do NOT include possible relations with anything outside the sequence of frames.

A.2.4 Relations not necessarily between adjacent segment in sequence:

1. sometimes an frame within the structure may not be included in the causal chain, so the chain may 'skip' one or more of the frames, but yet be a logical sequence, such as (again, with '—' as separation between segments):
    - "the wash machine broke down, — they realized in the morning, — so the clothes have to be washed by hand" (here the 1st and 3rd segments have a causal relation, but the 2nd segment is not involved)
    - "the planet is locked in gravitational embrace of the star, — while the the solar system formed eons ago; — today the planet still orbits around the star" (here similarly, the 1st and 3rd have a logical relation, not with the 2nd as an intermediary)

A.2.5 Ordering not always canonical (not always straightforward):

1. sometimes there are instances of reverse ordering, where the logical protasis comes after the apodosis in the linear surface order, such as:
    - "the house's foundation was sinking each day, — since it was built on an unstable foundation"

- ”the Welsh countryside is dotted with castles, — for the English occupation during the middle ages needed strongholds to control the region”

A.2.6 Real and hypothetical logical sequences:

1. some sequences of causation can occur within hypothetical, or imaginary world

   - counterfactual (conditionals that can potentially lead to causal relations) are also regarded as causal, such as:
     - ”Mike dreams that, when he is able finish his degree, he should be able to begin a prosperous career”
     - ”in case when a nuclear holocaust happens, — we should all remain under ground for several months”
     - ”should Mary have attended classes, — she would have gotten an A”
   - sometimes also hypothetical, such as ”X did something AS IF to accomplish something else” ”I want to put the pot on the stove, and boil the water in it”

A.2.7 The logical relation must be meaningful:

1. the content of two neighboring segments must have some meaning that lead the reader to believe in a real-world relations between two events/states/properties/etc.

   - so tautological sequences are not causal.
   - e.g. ”she hates him, — so she is full of hatred for him”, which does not tell us anything new.

A.2.8 Need to be positive causal relations:

1. make sure some negative causal relations are NOT counted:

   - ”he was too busy in the middle of his work — to take her to the play”
     - where the protasis provides a condition that prevents the apodosis to occur in sequence.
   - ”John made a hose trying to spray Mary with water, — but only succeeding in watering the lawn”
     - failing his intending objective of getting her wet

A.2.9 The relation between the events/states/properties need to be real in some world:

1. the link between the events/states/properties needs to be A leads to B, with actual relations between the two, or some causal relation where the entire A $\xrightarrow{caus}$ B sequence is hypothetical (again, '—' indicates separation between segments):
    - the relation should NOT be one of desire, hope, conjecture, or any other non-real relation.
    - thus the following examples would be EXCLUDED:
        - "he has the right — to decide his daughter's school attendance"
        - "I am trying to persuade her — to take part in the play"
        - "she longs for the her achievement — to be recognized as top of class"

A.3 Class annotation guidelines

The following are the per class guidelines for recognizing causal structures. The theoretical details of these categories are discussed in 9.4.1.

A.3.1 Efficient cause:

We mean by **entailment** a basic meaning that some latter event/state/condition logically follows some former as a consequence. This can be illustrated by an example such as,

- *"John reached his bare hand into the smelting furnace, and suffered severe burns"*
- *"the rivets holding up the walkway were rusted away, and at some point it broke away (as people stepped on it)*

In this case, there is no other possible outcome than his hand being burnt in any real-world scenario, and the second event is fully entailed by the occurrence of first event in the process. Not all the real-world situations are as clear cut as the above example, since there are usually other unlikely choices and remote possibilities of some configuration of current set of events that could be taken into account.

So rather than strict logical consequence, entailment is applied in terms of probability in the real world, where the first event makes it more likely that the second will occur; in most cases that increase in the probability of the later event is very significant, as in the above example, by John reaching into the furnace, he greatly increases the chance of him getting burnt (as opposed to the the chance of him randomly getting burnt by some sequence of causation with remote possibility, such as the furnace tipping over).

A.3.2 Cause of necessity:

In every day terms, **enablement** means that the presence of the first event/state/property provides a presupposed property for the latter event/state/property, without which the second event is never likely to be true in the system. By its nature, an enablement relation predicates nothing about whether the second event actually occurs, only that now one of the pre-conditions has been fulfilled by the first; So the first event/state/property provides some necessary but not sufficient condition for the second to take place.

We can see an enablement relation in

- *”John put up the carosel in the back-yard, Mary sat on the wooden horse and rode around the carosel”*
- *”John turned around at the driveway, Mary looked into his face at the door”*

The latter event in this case, Mary using and enjoying the merry-go-around in the back-yard, is only possible given that someone put up the amusement ride in the first place, which was accomplished by John in the former. Similarly, John turning around toward the door enables Mary to look him in the face in the second event.

A.3.3 Formal cause:

We mean by **implication**, that given some present static condition that exists in the world, which can be some state of a person/object (e.g. John being asleep) or a collection of properties after some event having occurred (e.g. after a tsunami), the existence of some other event/state/property likely is also implied. This differs from entailment in that for implication, no passage of time between the former and the latter or a temporal causal sequence is required.

Some simple examples of implication, as defined here, would be

- *”Mary had planned a detailed trip to Aruba, she became familiar with the tourist hot-spots on the island”*
- *”John ordered Mary to kill Patrick; Mary did not act completely on her own accord”*

Here Mary having planned a vacation implies that she knows about the locales there; and part of the circumstances that is required for Mary to take orders from John with regard to the killing implies that she wasn’t acting completely on her own.

A.3.4 Final cause:

A **purpose** relation since it requires some ¡strong¿volition¡/strong¿ (intent) on the part of some individual who is involved in the ¡b¿first¡/b¿ event, with the knowledge that the first event/state/property likely leads to the second. in order to be classified as such, and intentionality is not a logically well defined concept. Since it is primarily based on volition, it does not require that B already occurred, but only that the intention of the *causer* is there to set events into motion to bring it about.

This type of causality is also very often expressed using specific structures, especially a number of prominent patterns between segments (each representing a single event/state/property), such as:

- "for the purpose of"
- "so as to"
- "in order to"
- ... ...

But not all of them have such explicit patterns, some more subtle forms also occur in real language, such as:

- *he crawled on his hands and knees in the bushes, and avoided being seen by the sniper in the tower*
- *she purchased an elaborate gown for this weekend, and then impressed everyone at the prom dance with her style*

A.3.5 Cause with intermediate volition:

An **inducement** relation, for our purposes, is defined as a trigger that brings about the change in someone's disposition which leads to action; so the some individual that is an actor in the ¡b¿second¡/b¿ of the two events in sequence has his/her mental disposition affected by the first event, and thus it induces some new behavior in that individual. It is in that way in symmetry with ¡strong¿purpose¡/strong¿ that the mental state of some individual is involved, except the individual in this case takes part in the latter event. This type of relation can be expresses as the former event/state/property mentally prepare someone, prompts someone or waking the desire in someone to take part in the latter event.

Some examples in language would be:

- *"John handed the admission letter from her top school to Mary, who leapt for joy and went out and celebrated with her friends"*
- *"the magnificent view of the ocean can be seen from the cliff, the poet (who was enjoying the scenery) there was inspired to write an epic song"*

Where Mary was induced by receiving the admission letter to celebrate; and the poet was induced by the ocean view to make a new composition.

A.3.6 Material cause:

A constitution relation here means that the former state/event/property in some manner or sense encompasses the latter event/state/property. Such relations could be: playing hockey is to handling a stick; shopping in the mall is to paying for an outfit; becoming a martial artist is to obtaining a particular belt; etc. In other words, the latter events/states/properties in the pair may be viewed as components that are a part of a longer process or description encompassed by the former.

Some examples in language include:

- *"he has always been a free spirit, and being stuck in is home village is a difficult prospect for him"*
- *"she was a noble in the land of East Anglia, she had many titles and much land under her name"*

Here, someone being a free spirit encompasses the property that it is difficult for him to stay and live out his life in a small community; and someone being a noble woman (normally) encompasses the fact that she possesses titles and land.

A.3.7 Latent causal chain:

An **outcome** basically the expression of a longer sequence of events, but only the first and last are expresses in language. It is often the case that this is the least obvious type of causality to detect in the text by human judgment, since a lot of context might be required to ascertain the contents of final event at the time when the initial takes place.

An illustration can be seen in an example such as

- *"they prepared a sumptuous feast for the party; and afterwards had the same food for dinner for the next three nights"*
- *"Mary walked out of the door with a long-range rifle; later that day, Patrick was found dead in the center of the city square"*

In the first example, the connection between the two events as described is not always obvious to humans making a judgment on causality: the large feast prepared by the family is not necessarily all eaten (or the guests never arrived in numbers), so much of it was left over, and in not wanting to see a large amount of food go to waste, they decide to have the leftovers for the next few days. Even though most of the intermediate states/events are conjectures, these are reasonable inferences based on the known initial and final events in the sequences. In the second example, similar human judgment and knowledge are required to see the likely connection; it requires some advanced inference about things like: what a rifle is normally used to accomplish, the time frame of the events and the time it takes to travel between locations, any personal relationship there might be between Mary and Patrick, and a variety of other real-world issues.

Appendix B: Top Quantile Samples

B.1 Description of adjoined causals

Here are some initial positive samples, confirmed with human annotation, within the top quantile of the BNC testing data; each pair of ((a), (b)) is presumed to be causal in the direction of $e_a \longrightarrow e_b$. Since not all causalities are easily assessed without context, some of required an careful reading of the extended context, and some required additional information (outside of the corpus) to decide, so the included ones below are those whose causalities are easier to observe in non-contextual forms:

B.2 Top quantile examples of adjoined causal structures:

B.2.1 Positive Samples from testing part of BNC

There is a significant number of samples from the partial BNC corpus, where the causality could be understood by the speaker without extensive contextual information:

- 1. the net change was reduced by interest on the x billions of outstanding reserves probably worth ...
  2. the net change is xx millions a month
- 1. advance in the field of cosmetics means that, today, superfluous ingredients and allergens can be identified and substituted
  2. this is great news indeed for anyone with sensitive skin
- 1. regular exercise lowers ldl cholesterol level, yet raises hdl cholesterol level
  2. it reduces the risk of heart disease

- 1. organizations like acet need this support to enable people like myself to retain maximum control
  2. [so that we can] continue to live at home as independently as we can
- 1. granting of relief to an applicant who had delay would cause substantial prejudice or hardship to some person
  2. it would be detrimental to good administration
- 1. a discussion note will be prepared on the implications of the health service commissioner's decision
  2. it should be available by the end of October on the receipt of xx
- 1. when plants die down reduce watering
  2. [then] store the tuber in its pot in a dry, warm place
- 1. the design is open ended
  2. it could be adapted to separate passengers from their cars
- 1. forward planning and good communication are the two foundation stones that must be in place to guarantee that people ill with HIV have good choices
  2. [so that] they enjoy the very good quality of life wherever they choose to be
- 1. an acetlink person, possibly an existing volunteer, would keep their church informed about our work
  2. [he/she] would encourage people to consider becoming volunteers helping with fundraising ideas
- 1. our scheme is designed to ensure that lwt is in a position to make the program
  2. [our scheme is designed to ensure that lwt is in a position] to broadcast the program at the weekend from xx:xx onwards
- 1. for an informal dinner party, do not hide your guests from one another with a giant display
  2. rather use a series of small container groups that suit the shape of your tables
- 1. how information is gathered
  2. how information is acted upon
- 1. you can be more extravagant [for special occasions]
  2. [you can] go for more dramatic, impressive displays for special occasions

- 1. under telephone license regulation, they must keep a list of such requests (I think this refers to "do not call" list)
  2. [under telephone license regulation, they] must respect them in [the] future
- 1. my husband left me
  2. [he] went back to his ex-wife
- 1. yesterday barklays bank announced an increase in its base rate
  2. it is quickly followed by the other clearing banks
- 1. he is trapped in a body [that] is severely disabled
  2. [he] will spend the rest of his life in a wheel chair
- 1. a basic grant pack is also in preparation (a package for the organization age concerned England)
  2. [it] will form a part of this series (a series of publications from the organization)
- 1. the vehicles will be given to the national association of boys' club
  2. [they] will tour rundown inner-city areas
- 1. it is very low in mineral, it suits a wide range of people
  2. it is ideal for families with young children (this refers to spring water from Britain)
- 1. he would not stay more than a day, so we spend it together entirely
  2. [we] play music (referring to a visit by Chopin)
- 1. the wicker screen is ideal for dividing your lounge
  2. [it] conceals unwanted clutter
- 1. the odd militant supporter appears at the conference rostrum
  2. [he] is easy to spot in the wrong place (an expensive sea-front hotel)
- 1. the government has shied away from forcing unions to discipline members
  2. [it] has put proposals to curb strikes in essential services on the back burner
- 1. the club plans to sell its present grounds in east london
  2. [it plans] to move half mile down the road to create a leisure and community center
- 1. when the nazis took power, he went underground
  2. [he] worked in the communist resistance, was arrested

- 1. when the nazis took power, he went underground and worked in the communist resistance, was arrested
  2. [he] did eight years penal servitude
- 1. your boyfriend is finding it hard to come to terms with the prospect of fatherhood
  2. [he] is taking his resentment out on you
- 1. your son resents his step-brother for taking up your time
  2. [he] is worried that you do not love him any more
- 1. it is the first time our national and international networks have gathered together in one place
  2. [it] has made us all realize just how much the work has grown
- 1. [the] dissident minister to buck cabinet responsibility
  2. [to] risk the sack (sacking refers to dismissal from the P.M.)
- 1. the blade glides easily over my skin
  2. [it] leaves it very smooth
- 1. retirement is a time of great change
  2. [it] is also a time for development, either from choice or because they cannot find employment, people are retiring early
- 1. exchange rate [is] held up artificially high by interest rate
  2. [it] prejudices exports and encourages imports
- 1. [Sandy Lister] who was a doubtful starter at the beginning of the week, has recovered from a bout of tonsillitis
  2. [Sandy Lister] will play
- 1. consumers reject inferior products (while discussing economic theory)
  2. [consumers] make them unsellable
- 1. the positive consent system [was] proposed by the european commission
  2. [the positive consent system was] agreed by the council of ministers last month
- 1. [the change in the legislation] would revive uncertainty antagonizing people who have escaped the BR route
  2. [the change in the legislation] would have caused an acceptable delay in the legislation

- 1. the team selected by the scottish commonwealth game council is the smallest scottish contingent since the game in jamaica
  2. [the team selected ....] brings an immediate response from david lease, scotland's national athletics coach
- 1. ten thousand pounds will build you the highest column in the world
  2. [the column] will produce an astonishing effect
- 1. lwt is in a position to make [the program]
  2. [lwt is in a position] to broadcast a program
- 1. this type of display is fool-proof
  2. [this type of display] takes only minutes to arrange (blooms on souffle dish)
- 1. an acetlink person, possibly an existing volunteer, would keep their church informed about our work
  2. [he/she] would encourage people to consider becoming volunteers helping with circulated newsletters
- 1. (Imperative) dry fry or boil mince
  2. pour off resulting fat
- 1. the odd militant supporter appeared at the conference rostrum
  2. [the militant support] was easy to spot in the wrong place (appearing that he didn't belong there)
- 1. exchange rate [was] held up artificially high by interest rate prejudiced exports
  2. [the artificially high interest rate] encourages imports
- 1. he [mansell] illegally reversed his ferrari in the pit
  2. [his ferrari] failed to stop at a black disqualification (result of his illegal maneuver) flag waved three times during the race in estoril
- 1. i became close to friend of a friend
  2. [i] gradually fell in love with her
- 1. once the question was raised, they looked for possible sexual abuse
  2. [they] found possible sexual abuse in siblings

B.2.2 Non-canonical order pairs:

Some pairs that were highly ranked are in fact causal, but in the reverse to the normal order, and occur as $\langle apodosis, protasis \rangle$ pairs, such as:

1. (a) selahaatin osberk, due to be put on a flight to istanbul today, should be granted refugee status
   (b) [he] should have a well-founded fear of persecution
2. (a) the elderly couple in question have been sold an unsuitable product
   (b) [they] have not been given adequate risk warnings

B.2.3 Positive Samples from novels

The samples from the novels corpus is on average much more difficult to convey without a significant context in the discourse, and sometimes even events long distance away in the corpus (somewhere much earlier in the plot of the novel). First, we will show some samples from the top quantiles that are relatively easy to see the causality in a stand-alone form, or with some minimal additional information from the immediate context:

- 1. .... realized that the merchant was Syrian ....
  2. .... [the merchant] spoke through an interpreter ....
- 1. .... it (a long talk with Meg) seemed to have made a man of him, given him the strength to fight his own way, ...
  2. .... [it has] taught him a tender patience with which to bear he natural longings and failures of those he loved, ...
- 1. I possessed not only the will, but also the power
  2. to cook wholesome food got my little girls, and help myself when I could no longer afford to hire help
- 1. (after opening the curtain) the moon broke suddenly from behind the clouds
  2. [the moon] shone on her a bright, benign face, ...
- 1. Half a dozen jovial lads were talking about skates in another part of the room
  2. she longed to go to join them, for skating was one of the joys of her life
- 1. [Meg] used to enjoy his masculine amazement at the queer things women wanted ....
  2. she always insisted on his doing so (him seeing the details within her private expense book)

- 1. Amy stirred and sighed in her sleep (just after Jo and Amy were fighting over the book)
  2. as if eager to begin at once to mend her fault, Jo looked up with an expression on her face which it had never worn before
- 1. Jo took Beth down to the quiet place, where she could live much in the open air
  2. [Jo] let the fresh see breeze blow a little color into her pale cheeks
- 1. when you feel discontented, think over your blessings
  2. and [you] be grateful
- 1. I lifted my chin as he (Augustus Caesar) stared at me (Cleopatra Selene II, a.k.a. Cleopatra VIII)
  2. [I (Cleopatra Selene)] Let the emperor look at me an think of my mother (Cleopatra VII Philopator)
- 1. time had appeased her (Aunt March) wrath
  2. [time] made her repent her vows (March stated that she would not give money to the couple if Meg married Brooke)
- 1. [mother] is going to stay quietly in her room all day (since she is tired and ill)
  2. [mother] let us (Meg, Beth, Amy, Jo, and the rest of the household) do the best we can (manage the household in her absence)
- 1. [Mrs. March] examined her presents
  2. [Mrs. March] read the little notes which accompanied them
- 1. Then father came to the rescue, quietly managed everything
  2. [Mrs. March] never [has] been able to get on without him since
- 1. [father] made himself so helpful that I saw my mistake
  2. [Mrs. March] never [has] been able to get on without him since
- 1. Then father came to the rescue, quietly managed everything
  2. [father] made himself so helpful that I saw my mistake
- 1. you (Amy) broke the rules
  2. you (Amy) deserve some punishment
- 1. Mrs. March was both surprised and touched
  2. [Mrs. March] smiled with her eyes full as she examined her presents ....

- 1. .... lance caught another boy by the shoulder
  2. [lance] threw him to the ground
- 1. .... the poor, dear fellow was going away to forget his trouble
  2. [the poor dear fellow] was going to come home happy
- 1. our boy (Laurie) was getting fonder than ever of Jo
  2. [Laurie] wouldn't hear a word on the subject and scolded violently if anyone dared to suggest it (because of complications in the relationship, including the presence of Beth)
- 1. [Laurie] to lose heart at the first failure
  2. [Laurie] to shut himself up in moody indifference
- 1. he (Tom) could once return to a certain starting place (as a metaphor for revisiting parts of his life past)
  2. [Tom] could go over it slowly (the elusive element in his life that can be fixed)
- 1. After studying himself (Laurie speaks of himself in the 3rd person) to a skeleton all week
  2. a fellow (Laurie) deserves petting and ought to get it (a metaphor for some intimacy with Jo on the sofa in the corner)
- 1. the rain poured down his (a man marveling over Gatsby's books) thick glasses, and he took them off
  2. [the man] wiped them to see the protecting canvas unrolled from Gatsby's grave
- 1. (imperative frame: Laurie speaking about some metaphorical "castle in the air" asking Jo to) wait
  2. see if it doesn't bring you something worth having
- 1. the man drove so fast that Flo was frightened
  2. [Flo] told me to stop him
- 1. it ("one forlorn fragment of dollanity") was rescued by Beth
  2. [it] was taken to her refuge
- 1. each (of the older girls) took one of the younger sisters into her keeping
  2. [each of the older girls] watched over her (younger sister) in her own way, "playing mother" they called it

- 1. March did not take so romantic a view of the case
  2. [March] looked grave
- 1. I(Jo) 've made up my mind to bear it (a match that Jo wasn't pleased with)
  2. [Jo] shall not say a word against it
- 1. the man peered doubtfully into the basket, plunged in his hand
  2. [the man] drew one up (a dog up from the basket)
- 1. Polly began to flap about in his cage, so I (Amy) went to let him out
  2. [Polly] found a big spider there (in the cage)
- 1. I (Jo) 've got to run in (into the town)
  2. [Jo] get some paper ....
- 1. He (father) would pull out maps
  2. [father] show us where he'd marched
- 1. she (Elizabeth's cousin) and her noble admirer may be aware of what they (family's objections) are about
  2. [cousin and her lover] not run into a marriage which has not been properly sanctioned
- 1. I (Mrs. March) should accept them (social status and money) gratefully
  2. [Mrs. March would] enjoy your (Jo's) good fortune (in marriage, speaking about Jo marrying into wealth and status)
- 1. any girl reader who has suffered like afflictions (those involved in running a household) will sympathize with poor Amy
  2. [any girl reader] will wish her well through her task
- 1. Laurie opened his mouth to ask another question, but remembering just in time that it wasn't manner to make too many inquiries into people's affairs
  2. he shut it again, and looked uncomfortable
- 1. Laurie, who had heard what she (Jo) said (when she was attempting to find a carriage, in order to help Meg, who sprang her ankle), came up
  2. [Laurie] offered his grandfather's carriage, which had just come for him
- 1. Gatsby's notoriety, spread about by the hundreds who had accepted his hospitality and so became authorities on his past

  2. [Gatsby's notoriety] had increased all summer until he fell just short of being news

- 1. the hundreds (Gatsby's acquaintances) who had accepted his hospitality
  2. [Gatsby's acquaintances] so became authorities on his past
- 1. discordant din from those in the car had been audible for some time
  2. [discordant din added to the violent confusion of the scene]
- 1. they (Meg, Mother, Beth, and others) all broke down (during the time they sat waiting for the carriage to arrive)
  2. [all of them] cried bitterly (sending mother on her journey)
- 1. [Meg] forgot herself entirely till something in the brown eyes looking down at her made her remember the cooling tea
  2. [Meg] led the way into the parlor, saying she would call her mother (to drink the brewed tea)
- 1. presently a lovely Jewess appeared at an interior door
  2. [the lovely Jewess] scrutinized me with black hostile eyes
- 1. [Some old people] can sympathize with children's little cares and joys
  2. [these old people] can hide wise lessons under pleasant plays, giving and receiving friendship in the sweetest way
- 1. Some old people can make them (little children) feel at home
  2. [these old people can be] giving and receiving friendship (with the children) in the sweetest way
- 1. [Some old people] can sympathize with children's little cares and joys
  2. [these old people can be] giving and receiving friendship (with the children) in the sweetest way
- 1. Meg helped Jo clear away the remains of the feast, which took half the afternoon and left them so tired
  2. that [Meg and Jo] agreed to be contented with tea and toast for supper (not having to do cooking)

After the initial round of testing on the procedure for adjunctive causality, several obvious improvements can be made. One of which is treating parallel clauses and coordinate structures that are scoped by *existential*, *universal*, and *neutral* quantifiers differently; this largely refers to the fact that *"and"* connectives between parallel clauses or coordinate structures results in a very different probability of resulting in causality, compared to *"or"* connectives, and both behave differently from those with neutral connectives such as *"then"*. We also found that using a product of WordNet based similarity measure, and PMI results in the best outcome when extracting counterfactuals, so this will be used in future trials.

B.3 Description of embedded causals

Here is a sample of embedded causal structures found within the first quantile of the ranking by the GA informed by diffuse prototyping, from the BNC testing dataset. Since not all causalities are easily assessed without context, some of required an careful reading of the surrounding context, and some required additional real-world information to decide, so the included ones are top-quantiles ones whose causalities are easier to observe in non-contextual forms:

Both the overall surface sequence is shown, as well as the individual links in the causal chain shown as (a, b, c, ...). We present the general cases of highly ranked samples that are also identified by humans as positive, those that either are obvious to a human reader, or requires some minimal amount of context outside the sample. Some examples that require much more wider context in the corpus or even outside knowledge are presented in a separate section in Section 10.4; and some samples that are causal, but having a different sequential order than the surface sequence are presented in Section 10.4 . The basic color-coded labels for frames within embedded structure is as follows, each color denotes the position of the semantic frame in terms of it as a link in the causal chain:

B.4 Top quantile examples of embedded causal structures

- **contents of outer-most frame as causal chain link**
- **contents of 2nd outer-most frame as chain link**
- **contents of any center frame chain link**
- **contents of 2nd inner-most frame as chain link**
- **contents of the inner-most frame as chain link**
- **any auxiliary frame not directly part of the chain**

B.4.1 Positive embedded samples from BNC

- the sport and television market appears to be a free-for-all, delighting the most darwinist of entrepreneur, [such that] a scrutiny of those top five private network soon reveal that, four of them are in the hands of a holding firm called fin-vest, being itself the broadcasting vehicle of the socialist-leaning businessman silvio berlusconi

1. the sport and television market appears to be a free-for-all
2. $\xrightarrow{implication}$ [this fact] delighting the most darwinist of entrepreneur,
3. $\xrightarrow{induces}$a scrutiny of those top five private network soon reveal that four of them are in the hands of a holding firm called fin-vest
4. $\xrightarrow{constitute}$ being itself the broadcasting vehicle of the socialist-leaning businessman silvio berlusconi

- given that a property has to be close to amenities, if it is that, it is excluded from the right to be bought

  1. a property has to be close to amenities
  2. $\xrightarrow{entails}$ it is excluded from the right to be bought

- a year later he switched to gwardia warsaw, where he made sufficient impact that it attracted the attention of widzew lodz .

  1. he switched to gwardia warsaw (BNC is incorrect, the actual name of the club is gwardia warszawa)
  2. $\xrightarrow{enables}$ he made sufficient impact
  3. $\xrightarrow{entails}$ it attracted the attention of widzew lodz

- it (the parliamentary majority) can legislate arbitrarily to deprive groups of citizens of their basic rights or freedom, to enlarge its own power at the expense of local governments to weaken the ability of the media to inform the public or to sap judicial independence .

  1. it (the majority) can legislate arbitrarily
  2. $\xrightarrow{purpose}$ to deprive groups of citizens of their basic rights or freedom
  3. $\xrightarrow{purpose}$ to enlarge its own power at the expense of local governments
  4. $\xrightarrow{constitute}$
     (a) to weaken the ability of the media to inform the public
     (b) to sap judicial independence

- forward planing and good communication are the two foundation stones [that] must be in place to guarantee that people ill with HIV have the good choices and continue to enjoy the very good quality of life wherever they choose to be .

  1. forward planning and good communication are the two foundation stones [that] must be in place
  2. $\xrightarrow{purpose}$ to guarantee that people ill with HIV have good choices
  3. $\xrightarrow{purpose}$ [so that they] continue to enjoy the very good quality of life wherever they choose to be

- it is clear that he has always been acting in an independent capacity in the offer for eagle and [that] it has no connection with braithwaite

  1. he has always been acting in an independent capacity
  2. $\xrightarrow{implication}$ it has no connection with braithwaite

- it is so low [such] that they have to claim additional means-tested benefits such as income support, housing benefits, and community charge

  1. it is so low (their income level)
  2. $\xrightarrow{implication}$they have to claim additional means-tested benefits such as income support, housing benefits, and community charge

- jimmy knapp, leader of the national union of railway-man, argues that his members take action in support of the miners during the coal strike having a genuine interest in the fight to keep pits open .

  1. jimmy knapp, ..., argues that his members take action in support of the miners during the coal strike
  2. $\xrightarrow{constitute}$ having a genuine interest in the fight
  3. $\xrightarrow{purpose}$ to keep pits open

- eurotunnel is already in default of its credit agreement with the banks , [it] has in effect been given until the end of the year to settle its differences with the contractors to permit a viable financing strategy to be put in place .

  1. eurotunnel is already in default of its credit agreement with the banks,
  2. $\xrightarrow{entails}$[it] has in effect been given until the end of the year to settle its differences with the contractors
  3. $\xrightarrow{purpose}$ to permit a viable financing strategy to be put in place

- the success of the initial pilot program has be recognized by the ministry of health and the institute of health education, and led to maurice, kate, and ana meeting with government officials and representatives from unicef to establish a longterm training program in five romanian regions starting in November of this year .

1. the success of the initial pilot program has been recognized by the ministry of health and the institute of health education
2. $\xrightarrow{eventually\rightsquigarrow}$led to maurice, kate, and ana meeting with government officials and representattives from unicef
3. $\xrightarrow{purpose}$ to establish a long term training program in five romanian regions
4. $\xrightarrow{constitute}$ [training programs] starting in November of this year

- the more stubbornly East Germany's old men resist the changes, greater becomes the danger that the whole edifice they have constructed will collapse around them, threatening stability in europe and raising the question of reunification for which no one is prepared

1. the more stubbornly East Germany's old men resist the changes
2. $\xrightarrow{implication}$ greater become the danger that
3. $\xrightarrow{eventually\rightsquigarrow}$ the whole edifice they have constructed will collapse around them
4. $\xrightarrow{implication}$
   (a) threatening stability in Europe
   (b) raising the question of re-unification (with B.R.D.)

- there is however a video of his performance, which he is encouraged to play from time to time by his two children .

1. there is however a video of his performance
2. $\xrightarrow{enables}$ which he is encouraged to play time to time by his two children

- it sees the practical development of its work as taking place mainly at local levels and to this end [it] encourages the formation of autonomous local groups which can respond to need in own areas of the IRS .

1. the practical development of its work as taking place mainly at local levels
2. $\xrightarrow{induces}$ [it] encourages the formation of autonomous local groups
3. $\xrightarrow{enables}$ which can respond to need in [their] own areas of the IRS

- [he] would again be absent when his country strives to acquire the far point they need in poland ensure qualification for the world cup finals next summer

1. his country strives
2. $\xrightarrow{entails}$ to acquire the far point they need in poland
3. $\xrightarrow{purpose}$ to ensure qualification for the world cup finals next summer

- the management team led by peter jansen originally put up xx per cent of the m equity elements of the buyout [that] have seen its investment increase x fold at the mb group offer price .

  1. the management team led by peter jansen originally put up xx percent of the m-equity elements of the buyout
  2. $\xrightarrow{eventually\leadsto}$ w [which] have seen its investment increase x fold at the mb group offer price

- shelford is selling to rugby club in england and wales by marketing agents, and according to the cardiff link in the sponsorship chain he is available for almost any kind of promotional activity .

  1. shelford is selling to rugby club in england and wales by marketing agents
  2. $\xrightarrow{constitute}$ he is available for almost any kind of promotional activity

- the next stage of our election preparation will be a leaflet to be used during the election campaign itself, entitling [you to] your vote count .

  1. the next stage of our election preparation will be a leaflet
  2. $\xrightarrow{purpose}$ to be used during the election campaign itself
  3. $\xrightarrow{implication}$ [t] entitling [you to] your vote count

- the section of the community care act which is to be implemented in April will require authority that they make some direct provision for residential care under section xx of the act

  1. the section of the community care act which is to be implemented in April
  2. $\xrightarrow{entails}$ [this section of the community care act] will require authority that
  3. $\xrightarrow{purpose}$ they make direct provision for residential care under section xx of the act

  The first two frames of the chain are intertwined such that, the predicate copula e_1 is embedded in e_2 in the external argument position. Semantically, the a legislation that is to be implemented will require authority of some type; which is an *entailment* relation. Alternatively, we can view this causal relation in the other direction of $e_2 \to e_1$, where it is an *enablement* relation.

- work on the launch and the focus of the appeal are close to completion and we will benefit from our good advertising office's industry contacts .

  1. work on the launch and the focus of the appeal are close to completion
  2. $\xrightarrow{eventually\leadsto}$ we will benefit from our good advertising office's industry contacts

- inevitably the increase in the connection charge will only serve to put telephone ownership out of the reach of more elderly people .

  1. the increase in the connection charge
  2. $\xrightarrow{implication}$ to put telephone ownership out of the reach of more elderly people

- [he/she] is able rise to the ball and head it out of the reach of the sprawling paul heald to register his xx th goal of the season

  1. [he/she] is able to rise to the ball
  2. $\xrightarrow{purpose}$ [to] head it out of the reach of the sprawling paul heald
  3. $\xrightarrow{implication}$ to register his xxth goal of the season

- after watching them work out, team captain ann jones decided to [t] put jo durie and clare wood in against the indonesians .

  1. [t] watching them work out
  2. $\xrightarrow{eventually\rightsquigarrow}$ team captain ann jones decided
  3. $\xrightarrow{entails}$ to put jo durie and clare wood in against the indonesians

- the case arose out of a letter sent by the attorney general to the booksellers handling spy-catcher warning them [that] they are in contempt of court, because an injunction has been obtained to stop publication of extracts of the book in several national newspapers

  I a the case arose out of a letter sent by the attorney general to the book-sellers handling spy-catcher
    b $\xrightarrow{purpose}$ [t] warning them that they are in contempt of court

  II a an injunction has been obtained
    b $\xrightarrow{purpose}$ to stop publication of extracts of the book in several national newspapers

  (This contains three separate causal relations: (1) the letter sent by the attorney brought about the warning of the contempt of court; (2) the injunction from the court precipitated the stopping in the publication; and the two events (1) and (2) are themselves involved in an encompassing causality, that event 2 brought about event 1; there may be a causal chain that runs through this entire embedded structure, but the second pair of events seem to precede the first pair, so the sequence would be $obtain(entity_x, injunction) \xrightarrow{caus} stop(injunction, publication) \xrightarrow{caus} send(AG, letter, book-sellers) \xrightarrow{caus} warn(letter, book-sellers)$)

- akia moria recorded that his fledgling company sony was enabled to buy a license from western electric to develop transistor technology

   1. akia maria recorded that his fledgling company was enabled
   2. $\xrightarrow{enables}$ to buy a license from western electric
   3. $\xrightarrow{purpose}$ to develop transistor technology

- tickets for the whole event is on sale price from the box-office of the empire leicester square and we have pairs of free tickets for the first five readers to arrive at the cinema on Sunday morning bearing a copy of the independent .
   1. tickets for the whole event is on sale price from the box-office of the empire leicester square
   2. $\xrightarrow{eventually\rightsquigarrow}$ we have pairs of free tickets for the first five readers
   3. $\xrightarrow{entails}$ to arrive at the cinema on Sunday morning bearing a copy of the independent (a major morning news outfit in Britain)

- eurotunnel is already in default of its credit agreement with the bank syndicate, [that it] is seeking an extra xx billions on top of the xx billions raise so far .
   1. eurotunnel is already in default of its credit agreement with the bank syndicate
   2. $\xrightarrow{entails}$ [it] is seeking an extra xx billions on top of the xx billions raised so far

- we have a very full agenda for our scheduled meeting on October xx so it is decided that we meet on the xxth with cricket as the sole topic of discussion .
   1. we have a very full agenda for our scheduled meeting on October xx
   2. $\xrightarrow{entails}$ we meet on the xxth with cricket as the sole topic of discussion

- it is in the public interest for challenges to the exercise of public power that it is made promptly so that the administration and citizens alike know what the law is .
   1. it is in the public interest for challenges to the exercise of public power
   2. $\xrightarrow{entails}$ it is made promptly so that
   3. $\xrightarrow{purpose}$ the administration and citizens alike know what the law is

- these will all be presented to interested donors in a common summarized format [that] will provide them with the information they are looking for to make their decisions .
   1. these will all be present to interested donors in a common summarized format
   2. $\xrightarrow{implication}$ [which] will provide them with the information they are looking for
   3. $\xrightarrow{purpose}$ to make their decisions

- [they] will have to be able to accumulate sufficient provisions to support themselves in retirement for example those people who have not worked for many years because they were unemployed or disabled or cared for relatives

  1. [they] will have to be able to accumulate sufficient provision
  2. $\xrightarrow{enables}$ to support themselves in retirement

B.4.2 Positive embedded samples from novels

- the old gentleman knew that perfectly well and particularly desired to prevent it for the mood in which he found his grandson assured him that it would not be wise to leave him to his devices so stifling a natural regret at the thought of the home comforts he would leave behind he said stoutly: bless your soul, I'm not superannuated yet. I quite enjoy the idea (of traveling to London)

  1. the old gentleman know that perfectly well
  2. $\xrightarrow{induces}$ particularly desired to prevent it
  3. $\xrightarrow{induces}$ stifling a natural regret at the though of the home comforts he would leave behind
  4. $\xrightarrow{enables}$ he said stoutly: .... I quite enjoy the idea (to express that he is willing to travel)

- after I'd seen as much of the world as I want to I'd like to settle in Germany and [I] have just as much music as I choose I'm to be a famous musician myself and all creation is to rush to hear me

  1. after I'd seen as much of the world as I want to
  2. $\xrightarrow{entails}$ I'd like to settle in Germany (Counterfactual causality)
  3. $\xrightarrow{constitute}$ [I] have just as much music as I choose
  4. $\xrightarrow{eventually\rightsquigarrow}$ I'm to be a famous musician myself
  5. $\xrightarrow{eventually\rightsquigarrow}$ all creation is to rush to hear me

- she tries to find high-born women [high-born women] to bear him a son that she can take in as her own

  1. she tries to find high-born women
  2. $\xrightarrow{enables}$ [high-born women] to bear him a son
  3. $\xrightarrow{enables}$ she can take (a son) in as her own

- Before we unpacked Julia and I explored every nook and cranny of our new apartment including the ivory-trimmed wardrobe filled with old-fashioned garments handed down by Julii women which we took turns trying on

  1. Julia and I explored every nook and cranny of our new apartment, including the ivory trimmed wardrobe filled with old-fashioned garments
  2. $\xrightarrow{enables}$ which we took turns trying on

-

- *...., I was grateful that the emperor's daughter was so absorbed with her image in the mirror that she didn't notice me slipping on my mother's serpent bracelet, wrapped carefully in an old bloody dress under my mattress, I buried it ....* ..., I was grateful that the emperor's daughter was so absorbed with her image in the mirror that she didn't notice me slipping my mother's serpent bracelet, wrapped carefully in an old bloody dress under my mattress, I buried it ....

  1. the emperor's daughter was so absorbed with her image in the mirror
  2. $\xrightarrow{implication}$ she didn't notice me slipping on my mother's serpent bracelet
  3. $\xrightarrow{enables}$ [so I] wrapped carefully in an old bloody dress under my mattress
  4. $\xrightarrow{purpose}$ I buried it

- he (Marcus Antony, Cleopatra Selene's father) would even get down on his knees [he (father)] pretending to stalk me like one of the great cats of the jungle

  1. he would even get down on his knees
  2. $\xrightarrow{purpose}$ [he] pretending to stalk me like one of the great cats of the jungle

- (after Laurie went to college) then he avoided the tender subject altogether, wrote philosophical notes to Jo, [he (Laurie)] turned studious and he gave out that he was going to 'dig' [he] intending to graduate in a blaze of glory

  1. [he (Laurie)] turned studious
  2. $\xrightarrow{purpose}$ [he] gave out that he was going to 'dig'
  3. $\xrightarrow{purpose}$ [he] intending to graduate with a blaze of glory

- [the professor] catching her up with a laugh and [the professor] holding her so high over his head that she had to stoop her little face to [she] kiss him

  1. [the professor] catching her up with a laugh
  2. $\xrightarrow{enables}$ [the professor] holding her so high over his head

3. $\xrightarrow{entails}$ she had to stoop her little face
4. $\xrightarrow{purpose}$ [she] kiss him

- this year it was to be a plantation of sun flowers the seeds of which cheerful land aspiring plant were to feed Aunt Cockle-top and her family of chicks

1. it was to be a plantation of sun flowers
2. $\xrightarrow{eventually\rightsquigarrow}$ the seeds of which cheerful land aspiring plant were to feed Aunt Cockle-top and her family of chicks

- the reluctant climbers of the ladder of learning found their way strewn with flowers as it were, come to [they] regard the gentle giver as sort of fairy godmother who (the fairy godmother) sat above there and [the fairy godmother] showered down gifts miraculously suited to their tastes and needs

I a the reluctant climbers of the ladder of learning found their way strewn with flowers

b $\xrightarrow{implication}$ [the reluctant climbers] regard the gentle giver as sort of fairy godmother

II a who (the fairy godmother) sat above there

b $\xrightarrow{enables}$ [the fairy godmother] showered down gifts miraculously suited to their tastes and needs

- [Amy] feeling that the neighbors were interested in her movements she wished to efface the memory of yesterday's failure by a grand success today so she ordered the 'cherry bounce' (a strong drink made with whiskey, cherries, and lemon juice) and [she] drove away in state (to achieve a state of mind) to [Amy] meet and escort her guests to the banquet

1. [Amy] feeling that the neighbors were interested in her movements
2. $\xrightarrow{induces}$ she wished to efface the memory of yesterday's failure by a grand success today
3. $\xrightarrow{eventually\rightsquigarrow}$ she ordered the 'cherry bounce'
4. $\xrightarrow{enables}$ [she] drove away in state
5. $\xrightarrow{enables}$ [Amy] meet and escort her guests to the banquet

- I (Cleopatra Selene II, a.k.a. Cleopatra VIII) was a Ptolemy princess (meaning descended from Hellenic-pharonic blood-line), [Cleopatra Selene was] a queen in exile who (the queen in exile) must bide her time until she (the queen in exile) could think of some plot, some plan to [some plot/plan] return her to her throne

1. I (Cleopatra Selene II) was a Ptolemy princess
2. $\xrightarrow{constitute}$ [Cleopatra Selene II was] a queen in exile
3. $\xrightarrow{implication}$ who (the queen in exile) must bide her time
4. $\xrightarrow{enables}$ she could think of some plot, some plan
5. $\xrightarrow{purpose}$ [some plot/plan] return her to her throne

- a succulent hash arrived and Mr. Wolfsheim, forgetting the more sentimental atmosphere of the old Metropole [Mr. Wolfsheim] began to eat with ferocious delicacy

1. a succulent has arrived
2. $\xrightarrow{inducess}$ [Mr. Wolfsheim] forgetting the more sentimental atmosphere of the old Metropole
3. $\xrightarrow{enables}$ Mr. Wolfheim, ..., began to eat with ferocious delicacy

- [Laurie feeling] that out of the grave of a boyish passion there had risen a beautiful, strong friendship to [the friendship] bless them both (Laurie and Jo)

1. out of the grave of a boyish passion
2. $\xrightarrow{eventually\rightsquigarrow}$there had risen a beautiful, strong friendship
3. $\xrightarrow{entails}$ [the friendship] bless them both

- she (Amy) decided to make her will as Aunt March had done so that if she did fall ill and [she did] die her possession might be justly and generously divided

1. she decided to make her will
2. she did fall ill (Counterfactual causal)
3. $\xrightarrow{eventually\rightsquigarrow}$ [she did] die (Counterfactual causal)
4. $\xrightarrow{enables}$ her possessions might be justly and generously divided

- so Laurie let the days pass [Laurie] enjoying every hour [Laurie] leaving to chance the utterance of the word that [the utterance] would put an end to the first and sweetest part of his new romance

1. Laurie let the days pass
2. $\xrightarrow{enables}$ [Laurie] enjoying every hour
3. $\xrightarrow{enables}$ [Laurie] leaving to chance the utterance of the word (Counterfactual)
4. $\xrightarrow{eventually\rightsquigarrow}$ [the utterance] would put an end to the first and sweetest part of his new romance

- and Laurie poked the fire to [Laurie] hide a little twitching of the lips that he could not control

  1. Laurie poked the fire
  2. $\xrightarrow{purpose}$ [Laurie] hide a little twitching of the lips

- but to [Amy] stay at home with three selfish sisters and a grown-up boy was enough to [the situation] try the patience of Boaz

  1. [Amy] stay at home with three selfish sisters and a grown-up boy
  2. $\xrightarrow{implication}$ [the situation] try the patience of Boaz (Counterfactual causal: had the *experiencer* been Boaz)

- by late afternoon, I (Cleopatra Selene II) joined the rest of the women of the household Lady Octavia took it upon herself to [Lady Octavia] teach me (Cleopatra Selene II) to [Cleopatra Selene II] spin wool

  1. I (Cleopatra Selene II) joined the rest of the women of the household
  2. $\xrightarrow{constitute}$ Lady Octavia took it upon herself
  3. $\xrightarrow{purpose}$ [Lady Octavia] teach me (Cleopatra Selene II)
  4. $\xrightarrow{purpose}$ [Cleopatra Selene II] spin wool

- I'd asked the emperor to send us (Juba and Cleopatra Selene II) to Mauritania to [caesar] allow us to [Juba and Cleopatra Selene II] persuade the people there to build a port

  1. I'd asked the emperor to send us (Juba and Cleopatra Selene II) to Mauritania
  2. $\xrightarrow{induces}$ [caesar] allow us
  3. $\xrightarrow{induces}$ [Juba and Cleopatra Selene II] persuade the people there to build a port

- before the housewives could rest several people called and there was a scramble to get ready to [the housewives] see them (receive them with hospitality)

  1. several people called
  2. $\xrightarrow{entails}$ there was a scramble to get ready
  3. $\xrightarrow{enables}$ [the housewives] see them

- if he (William Collins) is disposed to make them any amends I (Mr. Bennett) shall not be the person to [I] discourage him

  1. he (William Collins) is disposed to make them any amends

2. $\xrightarrow{constitute}$ I shall not be (such a) person
3. $\xrightarrow{enables(bynotdiscouraginghim)}$ [I] discourage him

- and if difficulties were necessary to increase the splendor of the effort what could be harder for a restless, ambitious girl to [the restless and ambitious girl] give up her own hope, plans, and desires and [the restless and ambiguous girl] cheerfully live for others

  1. difficulties were necessary to increase the splendor of the effort (Counterfactual causality)
  2. $\xrightarrow{constitute}$ what could be harder (which would be the hardest) for a restless, ambitious girl
  3. [the restless and ambitious girl] give up her own hope, plans, and desires
  4. $\xrightarrow{induces}$ [the restless and ambitious girl] cheerfully live for others

- here the public gathered to [the public] view the statue of the goddess [the statue] rising in stone before us I (Cleopatra Selene II) went soft to see her my heart squeezing with unexpected joy

  I a the public gathered
  b $\xrightarrow{purpose}$ [the public] view the statue of the goddess

  II a I (Cleopatra Selene II) went soft to see her
  b $\xrightarrow{implication}$ my heart squeezing with unexpected joy

- if you (Nick) want to kiss me (Daisy) any time during the evening, Nick, just [Nick] let me (Daisy) know and I(Daisy)'ll be glad to arrange it for you (Nick) just [Nick] mention my (Daisy's) name, or present a green card

  1. you (Nick) want to kiss me (Daisy) any time during the evening (Counterfactual causal)
  2. $\xrightarrow{entails}$ [Nick] let (Daisy) know
  3. $\xrightarrow{enables}$ I(Daisy) will be glad to arrange it for you (Nick)
  4. i [Nick] mention my (Daisy's) name
     ii [Nick] present a green card

- Octavia was impatient to get on with it I(Octavia)'m sending you (Cleopatra Selene II) to learn with the rest of the children I know you've barely recuperated from your ordeal but exhaustion (from the learning) is the best thing to [exhaustion] erase pain

  1. Octavia was impatient to get on with it

2. $\xrightarrow{implication}$ I(Octavia)'m sending you (Cleopatra Selene II) to learn with the rest of the children
3. $\xrightarrow{purpose}$ [exhaustion from the learning] erase pain

- one of the guards searched Euphronius he actually put his unclean hands on our wizard's hold person I (Cleopatra Selene II) watched, aghast, trying to ignore the curious motion within the basket an echo of fear that snaked around my heart then the ill-mannered Roman guard approached me and I (Cleopatra Selene II) held my basket out to him [Cleopatra Selene II] hoping he'd reach inside (Counterfactual causal) hoping that whatever evil spirit lurked there would fly out [evil spirit] strike him dead

  1. one of the guards searched Euphronius
  2. $\xrightarrow{entails}$
     (a) I (Cleopatra Selene II) watched
     (b) [Cleopatra Selene II was] aghast
     (c) [Cleopatra Selene II was] trying to ignore the curious motion within the basket
  3. $\xrightarrow{eventually\rightsquigarrow}$ the ill-mannered Roman guard approached me
  4. $\xrightarrow{induces}$ I (Cleopatra Selene II) held my basket out to him
  5. $\xrightarrow{purpose}$ he'd reach inside (Counterfactual causal)
  6. $\xrightarrow{entails}$ whatever evil spirit lurked there would fly out
  7. $\xrightarrow{entails}$ [evil spirit] strike him dead

- prone upon the floor lay Mr. March with his respectable legs in the air likewise prone was Demi [Demi] trying to imitate the attitude with his own short, scarlet-stockinged legs, both grovelers so seriously absorbed that they were unconscious of spectators till Bhaer laughed his sonorous laugh, and Jo cried out, with a scandalized face ....

  1. lay Mr. March with his respectable legs in the air
  2. $\xrightarrow{entails}$ [Demi] trying to imitate the attitude with his own short, scarlet-stockinged legs
  3. $\xrightarrow{constitute}$ both grovelers so seriously absorbed that they were unconscious of spectators
  4. $\xrightarrow{eventually\rightsquigarrow}$
     (a) Baher laughed his sonorous laugh
     (b) Jo cried out

- the first supper - there would be another one (second supper) after midnight (so Nick would stay at Gatsby's) - was now being served and Jordan invited me [Nick Carraway] to [Nick] join her (Jordan's) own party, who were spread around a table on the other side of the garden

  1. there would be another one [second supper] after midnight
  2. $\xrightarrow{enables}$ Jordan invited me [Nick Carraway]
  3. $\xrightarrow{purpose}$ [Nick] join her own party

- a smart shower at eleven had evidently quenched the enthusiasm of the young ladies who were to arrive at twelve for nobody came (those having accepted an invitation to spend a day there for lunch and sight-seeing along the river) and at two the (Jo's) exhausted family sat down in a blaze of sunshine to [Jo's family] consume the perishable portions of the feast (prepared in anticipation of the guests) that nothing might be lost

  1. a smart shower at eleven had evidently quenched the enthusiasm of the young ladies who were to arrive at twelve
  2. $\xrightarrow{implication}$ nobody (of those invited) came
  3. $\xrightarrow{implication}$ the exhausted family sat down in a blaze of sunshine
  4. $\xrightarrow{eventually\rightsquigarrow}$ [Jo's family] consume the perishable portions of the feast
  5. $\xrightarrow{purpose}$ nothing might be lost (no food spoiled)

- I (Mrs. March) had all my girls to comfort me at home and his (Mr. Lawrence's) son was waiting, miles away, to say goodbye to him, perhaps I felt so rich, [I felt] so happy thinking of my blessings that I made him a bundle, [I] gave him some money, and [I] thanked him heartily for the lesson he had taught me

  1. (a) I (Mrs. March) had all my girls to comfort me at home
     (b) i. his (Mr. Lawrence's) son was waiting
         ii. [Mr. Lawrence's son was] miles away
         iii. [Mr. Lawrence's son] say goodbye to him (hypothetical)
  2. $\xrightarrow{enables}$
     (a) I (Mrs. March) felt so rich
     (b) [Mrs. March felt] so happy thinking of my blessings
  3. $\xrightarrow{eventually\rightsquigarrow}$
     (a) I (Mrs. March) made a bundle
     (b) I (Mrs. March) gave him (Mr. Lawrence) some money
     (c) I (Mrs. March) thanked him (Mr. Lawrence) heartily for the lesson he had taught me

B.4.3 Examples of special cases where algorithm selected ones not identified by human

The following several examples at first glance are not causal in terms of their structure. With a brief look at their original forms and their context, it is judged to be unlikely, but with no firm conclusion reached. However, thereafter a survey of background literature on the web around these subject reveal that there are some causal connections among the frames of each one of these embedded structures.

- report in the journal housing finds that many local authorities are having great difficulties in exempting from the right to buy those properties they consider to be especially suitable for old people

  (The "exempting the right to buy those properties preserves those properties for old people; the exemption is exemption from the obligation to repay the discount (akin to a mortgage tax exemption), for the repayment discount under the Preserved Right to Buy law in U.K. A secure tenant under the Right to Buy law, if it is their principle residence and is self-contained. Under the law, someone who has rented the home for an extended period of time has the right to purchase at lower than market value.)

  1. report in the journal housing finds that many local authorities are having great difficulties
  2. [t] exempting from the right to buy those properties they consider
  3. to be especially suitable to old people

- the cutting of equivalent to a p reduction in the basic rate of income tax is expected to be replenishing some of the consumer depleted ammunition .

  1. the cutting of equivalent to a p reduction in the basic rate of income tax
  2. $\xrightarrow{eventually\rightsquigarrow}$ to be replenishing some of the consumer depleted ammunition

  (Background: the cutting of income tax level provides more disposable income, which increases the amount of purchase power (the "ammunition") to sustain demand)

- the philosophy that if it goes down in the US, it must be going up in Europe has meant that smurfit is looking to Europe as its main engine for growth in the months to come when it must decide how to spend the xx billion cash raise from the recent financial restructuring with morgan stanley .

  (This actually has a complicated underlying set of real-world events, that seems opaque to readers not well versed in finance and the history of these companies: There was a 3 part financial restructuring deal that culminated between the Irish firm Jefferson Smurfit Group and Morgan Stanley. The deal follows the steps (1) MS purchasing the outstanding public shares of the JSG's US majority owned subsidiary

Jefferson Smurfit Corp; both of them combined already owned CCA; (2) MS and JSG combine CCA and JSC into a single entity as a joint owned subsidiary; (3) they recapitalized the combined company with participation from private equity. The deal brought in over 1 bn of cash on hand in the following year to fuel the expansion of JSG in Europe, which eventually becomes Smurfit-Kappa after merging with Paris based Kappa Packaging)

1. the philosophy that if it goes down in the US, it must be going up in Europe has meant
2. smurfit is looking to Europe as its main engine for growth in the months to come
3. it must decide how
4. to spend the xx billion cash raised from the recent financial restructuring with morgan stanley

B.4.4 Examples of opposite the nominal linear order of events in surface form

The following examples appear to have some causal relation contained within the embedded structure, but the sequence of the causal chain seems to differ from the canonical order of the surface sequence.

- Beth smiled and felt comforted for the tiny thing (a little grey coated sandbird that looked at her with friendly eyes) seemed to offer its small friendship and [the little grey sandbird seemed to] remind her [Beth] that a pleasant world was still to be enjoyed
    1. the tiny thing (a sandbird) seemed to offer its small friendship
    2. [the sandbird] remind her that a pleasant world was still to be enjoyed
    3. Beth smiled and felt comforted
- she (Jo) had the strength of mind to [Jo] to hold fast to the resolution she (Jo) had made [the resolution] when she (Jo) decided that she did not love her boy, and never could
    1. she (Jo) decide that she did not love her boy, and never could
    2. she (Jo) had made [the resolution]
    3. she (Jo) had the strength of mind
    4. [Jo] hold fast to the resolution
- yet my (Cleopatra Selene II's) mother (Cleopatra VII)'s statues stood proud and unharmed for Euphronius knew the Romans always had a price with the help of wealthy friends he (Euphronius)'d helped to ransom my mother's statues with all the gold in the temple treasury

1. Euphronius knew the Romans always had a price
2. (wealthy friends helped Euphronius to this end)
3. he (Euphronius) had helped to ransom [Cleopatra VII]'s statues with all the gold in the temple treasury
4. [Cleopatra Selene II's] mother (Cleopatra VII)'s statues stood proud and unharmed

- the future of the channel tunnel was in doubt last night after confirmation that cost has escalated by xx per cent to at least xx billion and a warning that its bankers may not provide fresh loan unless the financial crisis surrounding the project is resolve by christmas .

  1. cost has escalated by xx percent to at least xx billion
  2. a warning that its bankers may not provide fresh loan unless the financial crisis surrounding the project is resolved by christmas
  3. the future of the channel tunnel was in doubt last night

- I (Cleopatra Selene II)'d have to do it all for my husband's glorious reign I knew that my crown was to be only symbolic and that Mauritania did not belong to me but I'd been raised to rule and wanted to learn everything I could about the kingdoms my intended bridegroom had been given

  1. I (Cleopatra Selene II) knew that my crown was to be only symbolic
  2. Mauritania did not belong to me
  3. I'd have to do it all for my husband's glorious reign

- so it was uncomfortable to see the stoicism with which Agrippa and Marcella now approached the altar Marcella held her lips tight [Marcella's lips are] sewn up, fastened down I (Cleopatra Selene II) wonder whether she had ever breathed a day in her life or whether she'd always been that girl arguing with me while we decorated boughs for the Saturnalia

  1. Agrippa and Marcella now approached the altar
  2. Marcella held her lips tight
  3. [Marcella's lips are] sewn up, fastened down
  4. (a) I (Cleopatra Selene II) wonder whether she had ever breathed a day in her life
     (b) I wonder whether she had always been that girl arguing with me ....
  5. it was uncomfortable to see the stoicism with which ....

- if the bbc is to be persuaded to buy more drama from the independent street trader outside the buildings, the merchandise will need to be a little better finished .

  1. the merchandise will need to be a little better finished
  2. the bbc is to be persuaded
  3. to buy more drama from the independent street trader outside the buildings

- to do this we would like to recruit people to be willing to be an acetlink in their church.

  1. we would like to recruit people to be willing
  2. to be an acetlink in their church
  3. to do this

- Brandon Ormsby's transfer from Leeds United to Cardiff City was called off yesterday, after the defender decided he does not want to play in the third division

  1. the defender (Ormsby) decided
  2. he does not want to play in the third division
  3. Brandon Ormsby's transfer from Leeds United to Cardiff City was called off yesterday

- it is asking readers to say if they have been misrepresented in a program that they expose the tricks and deceptions used by tv to deliberately mislead people .

  1. it is asking readers to say if they have been misrepresented in a program
  2. they expose the tricks and deceptions used by tv
  3. to deliberately mislead people

This last case is in particular interesting, since the frames fit well together in its present form. If it is presented in a logical sequence, it would have to be presented as something like: *'the tv studios employed tricks and deceptions, which allowed them to deliberately mislead people; so now the investigation is asking people whether they have been misrepresented in a program, so as to expose such tactic'*. It is considerably more verbose, primarily because the long nominal arguments and the initial frame $studios \xrightarrow{employ} \langle tricks \wedge deceptions \rangle$ cannot be shared and coindexed between the frames. So it is primarily an issue of verbal parsimony that leads to this construction as opposed to the far longer but in sequence one.