EXTENDING CENTERING THEORY FOR THE MEASURE OF ENTITY COHERENCE

by

Milan Tofiloski

B.Sc. Honours (Physics and CS), University of Windsor, 2007

B.Math Honours, University of Windsor, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the School
of
Computing Science

© Milan Tofiloski 2009
SIMON FRASER UNIVERSITY
Fall 2009

APPROVAL

Name: Milan Tofiloski

Degree: MSc

Title of Thesis: Extending Centering Theory For The Measure Of Entity Coherence

Examining Committee: Dr. Lou Hafer
Chair

_____

Dr. Fred Popowich,
Professor, Computing Science
Simon Fraser University
Senior Co-Supervisor

_____

Dr. Maite Taboada,
Associate Professor, Linguistics
Simon Fraser University
Senior Co-Supervisor

_____

Dr. Francis Jeffrey Pelletier, Professor of Philosophy
Professor of Linguistics
Simon Fraser University
SFU Examiner

Date Approved: Nov 25, 2009

Abstract

This thesis extends Centering Theory by tracking all entities in an utterance to improve the measure of a text's coherence. This also eliminates the compromises in choosing between a sentence or a clause as the unit of analysis. Experiments show that tracking all entities improves the evaluation of coherence compared to Centering Theory's measures of coherence. The proposed coherence model does not require training, alleviating the need for costly training data. The model can be used to identify coherence gaps and propose candidate sentences to bridge such gaps. Our results approach the state-of-the-art in a sentence ordering experiment which identifies the original text amongst a collection of its permutations. We perform worse than the state-of-the-art in a summary coherence experiment that ranks output from automatic summarization systems such that the rankings correlated with human rankings. These results illustrate the need to incorporate lexical cohesion for a more complete coherence model.

The ten thousand questions are one question. If you cut through the one question, then the ten thousand questions disappear.

Zen proverb

Acknowledgments

Thank you to everyone (you know who you are).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1   What is Coherence?

This thesis investigates **coherence** of a **discourse** through its **entities**. *Discourse* can refer to a written text, a spoken conversation, or anything that conveys information either explicitly or implicitly. Since it is *texts* that will be the focus of this study, the definition of discourse in this thesis is restricted to written texts. *Coherence* describes the process of how elements of the text combine to form a unified whole. It is the study of how sentences combine to form paragraphs, how paragraphs combine to form chapters, and how these elements combine to form a text. An *entity* is any thing which can be referred to (see Section 2.1.4 for further elaboration on entities). Hence, this thesis explores the function of the change in salience of entities in a text in order to determine the text's coherence.

The term '**cohesion**' is often used synonymously with coherence as well as being used to describe a kind of coherence. Renkema [2004, pg. 103] refers to cohesion as something that is *discourse internal*, whereas coherence is *discourse external*. In other words, cohesion refers to connections that can be made within the text, and coherence refers to connections that are made to outside the text in the 'real world'. To further complicate terminology, both ideas, cohesion (discourse internal) and coherence (discourse external), are subsumed under *Coherence* (capitalized C), which is the overall generalization that captures the discourse external and discourse internal under one category.

Coherence can be viewed as the process of combining multiple semantic elements of a text into a unifying interpretation, an attempt at joining the elements into a coherent whole. Coherence is an invisible glue that binds multiple meanings with the intent of forming a

single unified meaning. Starting with the simplest element, the word can be considered a basic unit of meaning, where above the word comes the phrase, next the clause, with the sentence still above that, and so on until the text is observed as a whole. The way the individual units combine to form ever larger units is the main effect of coherence. Coherence can be viewed as the canvas that holds the various colours of a painting, whose ultimate effect is the finished image. In this sense, coherence is not something that can be directly observed or measured, but something that exists behind the scenes ('behind the painting'). Coherence should not to be confused with the text's meaning; it is how the individual meanings combine and form a single unified meaning. Continuing the analogy, the meaning of the text corresponds to what the painting is about, whereas coherence is what binds the paint together upon which an interpretation can be made.

Hovy defines coherent text as [1988, pg. 164]:

> text in which the hearer knows how each part of the text relates to the whole; i.e.,
> (a) the hearer knows why it is said, and (b) the hearer can relate the semantics
> of each part to a single overarching framework.

Determining coherence is not a black and white assessment. Asher and Lascarides note that "coherence is not a yes/no matter, but rather the quality of coherence can vary" [2003, pg. 20]. There is thus a degree of coherence to a discourse and not merely a simple classification of being either coherent or incoherent.

1.2 Coherence vs. Cohesion

This section discusses the distinctions between cohesion and coherence. **Cohesion** occurs between elements within a discourse, whereas **coherence** occurs between an element external to the discourse and an element within the discourse. Cohesion, as defined by Halliday and Hasan [1976], is the dependence of one element of a text on another part. Fleshing the definition further, it is also the "relations of meaning existing within a text, and that define it as a text... (that) occurs where the interpretation of some element in the discourse is dependent on that of another" [Halliday & Hasan, 1976, pg. 4]. Cohesion deals with parallelism, narration, lexical relatedness between words, etc. Coherence deals with world-knowledge, logic and logical relations (reasoning, consequence, result, inference, induction, causation, effect, etc.).

Cohesion's import to language lies in its function of constructing sense and meaning between parts of a text. Without cohesion, the arrangement and order of sentences would not matter, and the consequent meaning can be extracted independent of sentence order [Harris, 1952]. Hence, cohesion is to utterances what syntax is to words. Permuting the words of a sentence not only possibly modifies the meaning, but modifies it in such a way that the sentence may become non-grammatical or nonsensical. Similar to how words combine to form a sentence, cohesion could be viewed as a grammar for the way utterances combine together to form a single utterance. Halliday & Hasan mention properties of cohesion "that distinguish a text from a disconnected sequence of sentences" [Halliday & Hasan, 1976, pg. 1], with the caveat that "a text is not something that is like a sentence, only bigger" [Halliday & Hasan, 1976, pg. 2].

The boundary between cohesion and coherence is not necessarily distinct. Explicit discourse markers can be used in cohesion that would signal logical relations, which in turn would be utilized for coherence, markers such as "if..., then...", "because...", etc. In coherence, the *Result* relation can be signalled by "and" or "therefore"; a *Cause* relation can be signalled by the word "because", etc. [Kehler, 2002]. The coherence relations can be determined solely from the text vis a vis the discourse markers. VP ellipsis would be another example of cohesion since the elliptical gap can be resolved solely from within the text, as in the example "George voted Republican, and Barbara did too". The absence of explicit logical relations would be the domain of coherence, with the objective in such a case being to determine how to recover the implied relations.

The distinctions between coherence and cohesion are subtle yet important within studies of discourse and coherence. For the rest of this thesis the term *coherence* will primarily be used to signify *entity coherence*, which follows the terminology used in other studies [Barzilay & Lapata, 2008, Fillipova & Strube, 2004, Karamanis, 2007]. The rest of this chapter will continue to differentiate between coherence and cohesion, with later chapters using the term *coherence* instead of *cohesion* with the understanding that the type of coherence under discussion is one that is restricted to elements explicitly mentioned within the text.

1.3 The Different Types of Coherence

1.3.1 Coherence Within Utterances

The type of coherence studied in this thesis occurs between sentences and within sentences (between clauses), which are referred to as **utterances**. Sub-utterance coherence may deal with verb semantics, selectional restrictions on the arguments that verbs can take and the roles that nouns can play, lexical cohesion or adherence to grammatical/syntactic structure. Since most of the types of coherence that exist *within* the utterance can also exist *between* utterances, the next section will describe the different types of coherence more fully to avoid repeating definitions and to concentrate all definitions in one section.

1.3.2 Coherence and Cohesion Between Utterances

Halliday & Hasan [1976] identify five different types of cohesion that can exist in a text. For the most part, these types of cohesion occur between utterances[1]. The five types of cohesion according to Halliday & Hasan are (examples and descriptions taken from Renkema [2004, pg. 103]):

reference - semantic substitution, where the dummy word is typically a pronoun "John finished the race. He came in first." The word *He* refers to *John*.

substitution - replacement of a word or group of words by a "dummy" word ("These biscuits are stale. Get some fresh ones."), where *ones* is substituted for *biscuits*; similar to reference but the replacement is grammatical[2].

ellipsis - occurs when words or parts of a sentence are omitted, as in the example "John went to the movies, but Steve didn't." Part of the subordinate clause has been omitted (*go to the movies*), where the omitted part can be found in the main clause.

conjunction - how one clause connects to an adjacent clause; Renkema also refers to these as connectives; the connectives can be *additive*, *temporal*, *causal*; the connectives

---

[1]There are some exceptions, as in the sentence '*Steve loves himself*', where *himself* is a reference to *Steve* and resolved within the utterance.

[2]For simplicity, the term "referent" is used to capture both substitutive and referential coherence. Without loss of generality, the first two types, reference and substitution, can be collapsed into one type: *referential coherence*.

can be hypotactic (where a main clause is combined with a subordinate clause) or paratactic (two main clauses that do not exhibit a subordinating relationship are combined, which is exemplified by the sentence "He quit school and is moving to California.")

lexical cohesion - this type of coherence deals with links between nouns, verbs, adjectives, and adverbs; the links are manifested through synonymy, hyponymy, hyperonymy, meronymy, antonymy

The last type of coherence, *lexical cohesion*, also encapsulates a type of coherence that is the focus of this study, which is **entity coherence**. Entity coherence also overlaps with referential coherence, since the majority of the time the referents are referring to entities[3].

An example illustrating coherence that is discourse external versus a coherence that is discourse internal are the texts "John ran out of milk. So John's sister Susan went to the store." and "John ran out of milk. So Susan went to the store." The first text is discourse internal because the relationship between *John* and *Susan* is made explicit through the modifier *John's* in *John's sister*. In the second text, which is an example of discourse external coherence, the relationship is not made explicit. The information that Susan is somehow related to John cannot be inferred directly from the text. An external knowledge source is required to bridge the two sentences, or some sort of *common ground* between the participants must have been established (i.e., it is somehow implied that both speaker and hearer are aware of each other's knowledge that an explicit mention can be omitted).

Discourse external coherence is not a part of this study. The scope of this paper investigates entity coherence. Other types of coherence, and their relationship to entity coherence, are left for future work. Entities are discussed in more detail in Section 2.1.4.

1.4 Applications of Coherence

Coherence can be used to assist natural-language generation (NLG) systems as well as document summarization systems [Barzilay & Lee, 2004, Madnani et al., 2007, Mani, 2001, Marcu, 2000]. In natural-language generation, as well as document summarization, after

---

[3]A software system known as a *coreference resolver* will be used as a preprocessing stage before the analysis of entity coherence is carried out. Thus, the entity coherence system is isolated from the referential coherence system, in line with our desire to make a distinction between the two types of coherence even though the two are strongly coupled.

the sentences have been generated or extracted, a system incorporating coherence can be used to determine the most coherent order of sentences. This is identical to the sentence ordering task, which is one of the tasks used to evaluate the system proposed in this thesis (see Section 5.1). The objective in both cases is to determine which of the permutations of the sentences would be the *best* (i.e., most coherent) way of presenting such information.

Miltsakaki and Kukich [2004] also look at the automatic evaluation of student essays using entity coherence[4]. They examine "whether local discourse coherence, as defined by a measure of Centering Theory's Rough-Shift transitions, might be a significant contributor to the evaluation of essays. Rough-Shifts within students' paragraphs often occur when topics are short-lived and unconnected, and are therefore indicative of poor topic development" [Miltsakaki & Kukich, 2004, pg. 25]. Essentially, Miltsakaki and Kukich are correlating the type of topic transitions with the coherence of a text under the assumption that a text exhibiting a large number of drastic topic transitions corresponds to an incoherent text. The essays "are scored on a scale of 1-6 points, where a score of 1 indicates an extremely poor essay and a score of 6 indicates an excellent essay" [Miltsakaki & Kukich, 2004, pg. 26]. In their experiment, they manually tagged coreferring expressions rather than use coreference software. Automatically determining student essay quality saves human resources, which means the automation allows one to save both time and money. Thus, more essays can be evaluated in a shorter time.

An application of coherence which has received little attention [Cherry, 2008] is the effects of coherence measures on machine translation systems. In this case one would compare two different translations of a text to see how similar their coherence measures or entity movement patterns would be. The premise behind using coherence for machine translation evaluation is that accurate translations would contain the same type of coherence patterns between the original text and its translation.

1.5 Overview

The objective of this thesis is to describe a methodology that tracks all entities and all utterances in a text in order to determine the coherence of a text and to also illustrate the

---

[4]We did not have access to any GMAT essays written by students due to the confidentiality involved with the data. Their student essay coherence experiment is essentially reproduced in our text coherence experiment.

benefits of tracking all entities. The most common approach to evaluating entity coherence is encapsulated by Centering Theory, which tracks a single unique "center" (similar to a *topic* or *focus*) between adjacent utterances. The theory is based on linguistic principles that correlate the salience and grammatical role of an entity to its topicality in the discourse: entities that occupy prominent roles in an utterance are considered prominent in the discourse. In the representation proposed in this thesis, utterances are represented as vectors whose elements are the entities of the text, with their values corresponding to the *salience* of the entity within the utterance. Transitions between sentences are calculated by computing a 'difference' in entity salience between the adjacent utterance vectors. These transitions range from coherent utterance transitions all the way to incoherent transitions. The vector representations allow the assignment of a value according to the type of transition. This value, in turn, corresponds to the measurement of a coherence gap.

The questions addressed in this thesis are the following:

- What are the benefits in tracking all the entities in every utterance instead of the movement of a single entity as is done in Centering Theory?

- How do the four standard Centering transitions (*continue, retain, smooth-shift, rough-shift*), which are coarse-grained measures of entity transitions that track a single entity (Centering), translate into a fine-grained measure that tracks all entities? Basically, is there a way to map the coarse-grained measures of Centering Theory into the fine-grained ones proposed in this thesis.

- How can one identify a coherence gap? And after a coherence gap has been identified, is there a way to bridge such a gap?

Any model of entity coherence must address a small number of parameters. The main factors in such a model of entity coherence are:

- identifying entities

- ranking entities according to their prominence in the discourse

- computing coherence transitions between utterances

The second and third points are the main aspects of this thesis. For the identification of entities, we will use conventional definitions and principles as to what constitutes an entity, while employing standard software that simplifies their extraction from a text.
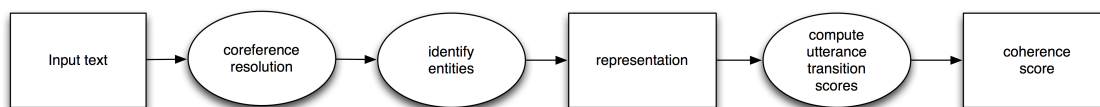
Figure 1.1: Overview of the system.

The thesis is outlined as follows: definitions and motivating concepts, along with a literature review of other theories and approaches to coherence, are explained in Chapter 2. Chapter 3 formulates a representation of utterances as vectors, with an eye toward capturing the representations used in Centering Theory. Chapter 4 introduces the idea of coherence trees, which are used to bridge coherence gaps. Coherence gaps are identified using coherence trees along with the sentence vector construction of Chapter 3. Experiments are performed and discussed in Chapter 5. Conclusions and a summary are given in Chapter 6.

The contributions of this thesis are as follows:

- Formulate a vector representation of sentences which allows the leveraging of the tools and manipulations associated with vectors. The vector representation allows the tracking of all entities in a text rather than a single 'most topical' entity. The representation also allows a more complete formulation of entity salience transitions between sentences using all the entities instead of a single 'most topical' entity.

- Minimize the compromises associated with choosing the unit of analysis (utterances which typically take their form as a sentence or a clause). This alleviates a common point of contention within the discourse community by shifting the focus to the weighting associated with the clauses as they form sentences. The model proposed allows one to distinguish between sentences and clauses in a natural way using a main/subordinate clause weighting as well as a *recency* weight that allows one to capture clause/sentential relationships in a linguistically principled way.

- Extending Centering Theory by tracking all entities over the entire text instead of a single entity between adjacent entities. This extension allows for Centering to overcome its shortcomings and compromises associated with utterances, unique centers, and local versus global scope.

- Constructing an utterance transition function which computes the degree of salience change between utterances. The set of transition functions when compared with those

of Centering Theory are finer-grained, and are able to distinguish and rank every possible transition possible in a text. The sentence transition function can also be used to compute the transitions over a history of utterances and not just the immediate predecessor using the notion of *recency*.

- Formulate a tree representation of the global entity coherence structure. This allows one to both visualize the text in terms of entity relations as well as propose a way to fill coherence gaps. The entity coherence gaps and proposed candidates for bridging such gaps are performed using a *sentence transition* function. The sentence transition function computes the size of the gap between utterances and then selects a candidate utterance to insert between the gapped utterances in order to minimize the coherence gap.

Chapter 2

Coherence Models

Various attempts at computationally determining entity coherence in text incorporate syntactic, lexical, or discourse approaches [Barzilay & Lapata, 2008, Barzilay & Lee, 2004, Elsner et al., 2007, Graesser et al., 2004], and combinations thereof. These approaches span the spectrum from local coherence (e.g., coherence between adjacent sentences) to global coherence (e.g., coherence over the entire text). This section presents the ideas and defines terminology associated with coherence. We also provide a related work section that provides an overview of previous approaches to coherence. The approach proposed by this thesis extends and modifies many of the existent ideas that are covered in the section on related work.

2.1 Discourse Definitions

Global coherence (Section 2.2.1) refers to the overall intent of the discourse and whether the individual **discourse segments** cohere in their individual purposes to support that overall intent. **Local coherence** (Section 2.2.2) focuses on adjacent discourse units (also known as **utterances**, described in Section 2.1.3) which typically occur between adjacent sentences. Computational approaches typically use the sentence as their **discourse unit** due to the difficulty in decomposing sentences into their constituent clauses [Barzilay & Lapata, 2008]. There has been some work done on identifying discourse units in a sentence [Soricut & Marcu, 2003, Tofiloski et al., 2009], but automatically identifying the clause hierarchy in terms of the actual type of subordinations has not had much investigation in the research community. Clause hierarchy has been analyzed indirectly through the

study of rhetorical relations [Marcu, 2000, Soricut & Marcu, 2003], but studies of rhetorical relations do not distinguish amongst the different types of subordinate clauses (relative, adverbial, etc.).

This section elaborates on such distinctions and definitions, illustrating both the benefits and disadvantages of various instantiations. Where applicable, we also mention some of the difficulties and bottlenecks when trying to automate such ideas. In some cases, definitions need to be revised such that they better address the complexity and difficulty involved in their implementation. In these cases (e.g., *discourse segments*), it would best to define such ideas in terms of what is a tractable implementation (i.e., define discourse segments in terms of lexical chains or functional zones rather than in terms of authorial intent or purpose due to their simpler implementation). Such refinements are not the focus of this thesis, but are mentioned for future work.

2.1.1 Definition of Discourse

A **discourse** can be referred to as a text, conversation, etc.[1]; in essence, anything that has an intention to convey or have a social purpose can be considered a discourse [Grosz & Sidner, 1986, Schiffrin, 1994, Renkema, 2004]. Schiffrin [1994, pg. 20] also defines discourse in two ways: as a "unit of language (above the sentence)", and as "having a particular focus". The 'above-the-sentence' aspect is a natural extension of syntax, which focuses on the structure contained within a sentence. The structure between sentences is the domain of discourse [Harris, 1952]. Yet, there is a subtle difference between syntax and discourse. As Halliday & Hasan [1976, pg. 1] explain in their book *'Cohesion In English'*:

> A text is sometimes envisaged to be some kind of super-sentence, a grammatical unit that is larger than a sentence but is related to a sentence in the same way that a sentence is related to a clause, a clause to a group and so on: by **constituency**, the composition of larger units out of smaller ones. But this is misleading. A text is not something that is like a sentence, only bigger; it is something that differs from a sentence in kind. A text is best regarded as a

---

[1]For the types of experiments and studies performed in this thesis, the type of discourse that will be focused on is the text. [Halliday & Hasan, 1976] define the text as "any passage, spoken or written, of whatever length, that does form a unified whole". *'Text'* referred to in this study will only involve the written text, not spoken text or transcripted text. The text can consist of a single document or a collection of documents (multi-document text). Thus, the focus of this study is to investigate coherence within written text. Coherence regarding conversations or dialogues is outside the scope of this study.

> *semantic* unit: a unit not of form but of meaning. Thus it is related to a clause
> or sentence not by size but by *realization*, the coding of one symbolic system in
> another. A text does not consist of sentences; it is realized by, or encoded in,
> sentences. If we understand it this way, we shall not expect to find the same
> kind of *structural* integration among the parts of a text as we find among the
> parts of a sentence or clause.

Halliday & Hasan further explain how coherence is the glue that binds the structure of
sentences, and in turn the structure of discourse. In summary, with this important distinc-
tion made between syntax and discourse, it can be said that syntax is the structure *within*
sentences, and discourse is the structure *between* sentences.

The "having-a-particular-focus" aspect mentioned by Schiffrin can be viewed as what
the discourse is about, i.e., the discourse intent (or purpose). The overall intent of the
discourse is sometimes referred to as the **discourse purpose** (**DP**) [Grosz & Sidner, 1986].
Grosz & Sidner [1986, pg. 178] describe discourse purpose as follows:

> Although typically the participants in a discourse may have more than one aim
> in participating in the discourse (e.g., a story may entertain its listeners as
> well as describe an event; an argument may establish a person's brilliance as
> well as convince someone that a claim or allegation is true), we distinguish one
> of these purposes as foundational to the discourse. We will refer to it as the
> discourse purpose (DP). From an intuitive perspective, the discourse purpose is
> the intention that underlies engaging in the particular discourse. This intention
> provides both the reason a discourse (a linguistic act), rather than some other
> action, is being performed and the reason the particular content of this discourse
> is being conveyed...

Hence, there is an underlying reason, or purpose, for the linguistic act (i.e., the text) to be
performed. Since purpose in a discourse is not the focus of this study, we can simplify the
terminology used so that any subsequent mention of a discourse's purpose will be referred
to as its **intent** or **intentions**.

Beaugrande and Dressler [1981] list seven criteria for a text to qualify as a discourse:

cohesion connections from one element of a text to another

coherence connections between elements of a text that are connected by elements outside the text, i.e., a 3rd element is used to connect the two

intentionality the authorial intent of the text's objective

acceptability sequence of sentences are accepted by the audience e.g., in the city of Milan, I can make the claim that I own the city since it shares the same name as me, but no one else would accept this

informativeness discourse must contain new information, and the reader must understand what is in the discourse

situationality context in which the discourse has been produced

intertextuality sentence sequences related to other sentence sequences by a form or meaning

Of the seven criteria listed above, most attention has been paid to cohesion and coherence [Renkema, 2004], which are the primary focus of this study. The other criteria are not considered further in this thesis.

Coherence is a way of determining the quality of the discourse in terms of its attempt at clarity and ability to convey intent. From Zellig [1952, pg. 3]:

> Arbitrary conglomerations of sentences are indeed of no interest except as a check on grammatical description; and it is not surprising that we cannot find interdependence among the sentences of such an aggregate. The successive sentences of a connected discourse, however, offer fertile soil for the methods of descriptive linguistics, since these methods study the relative distribution of elements within a connected stretch of speech.

Zellig lays the framework for the symbiotic relationship between coherence and discourse. One cannot have a discourse without coherence, nor can there be coherence without a discourse. The two are closely coupled in such a way that their essence is dependent on the other. The "connected stretch of speech" is due to coherence, which is what provides the *connections* between parts of the text.

2.1.2   Definition of Discourse Segment

A **discourse segment** is a smaller unit of text which comprises the discourse. The purpose (or intent) that a discourse segment conveys is referred to as the **discourse segment purpose** (**DSP**) [Grosz & Sidner, 1986]. The discourse segment purpose is a "smaller" purpose intended to assist in conveying the overall discourse purpose. A discourse segment itself is composed of a sequence of smaller units of expression, called utterances (which are discussed further in Section 2.2.2). Grosz and Sidner [1986, pg. 177] describe the relationship between utterance and discourse segment as such:

> The utterances in a segment, like the words in a phrase, serve particular roles with respect to that segment. In addition, the discourse segments, like the phrases, fulfill certain functions with respect to the overall discourse. Although two consecutive utterances may be in the same discourse segment, it is also common for two consecutive utterances to be in different segments. It is also possible for two utterances that are nonconsecutive to be in the same segment.

Discourse segments can be viewed as parts of the discourse that when adjacent have distinct purposes (if the purposes of adjacent segments were identical, then the two segments could be collapsed into one single discourse segment).

2.1.3   Definition of Utterance

A discourse segment is comprised of **utterances**. An utterance is an **elementary discourse unit**, which is defined to be the simplest unit in a discourse that cannot be decomposed into a simpler unit. Discourse units can be manifest in many different ways, taking the form of clauses, sentences, multiple sentences, etc. [Poesio et al., 2004a].

Different instantiations of Centering Theory (a theory based on using entity structure to analyze coherence, which will be discussed fully in Section 2.3.2) use different units for their utterance [Poesio et al., 2004a]. An utterance can take the form of a clause, simple sentence, complex sentence, paragraphs, etc. (other units may be adopted in spoken language that better represent and capture speech discourse). Givon [1983, pg. 7] describes how clauses combine to form complex discourse units:

> The clause ('sentence') is the basic information processing unit in human discourse. A word may have 'meaning', but only the proposition - grammaticalized as a clause - carries information. Human discourse, further, is multi-propositional. Within it, chains of clauses are combined into larger thematic units which one may call thematic paragraphs... These may further combine into larger yet discourse units (such as 'paragraphs', 'sections', 'chapters', 'parts', or 'stories').

In a study of different instantiations of Centering Theory, Poesio et al. [2004a, pg. 323] used the definition of **clause** as "sequences of text containing a verbal complex, all its obligatory arguments, and all postverbal adjuncts".

From a computational point of view, the unit of utterance is restricted by the performance and quality of existing software in its ability to segment a text into whichever definition of utterance one wishes to use. For the purposes of this study, the sentence will be the primary unit for an utterance. Current clause segmenters introduce noise [Barzilay & Lapata, 2008], and, to our knowledge, identifying the clause hierarchy (e.g., identifying the main clause and subordinate clauses) has not had much attention in the research community (Marcu [2000] is the only study we are aware of), and existing discourse segmenters decompose sentences into discourse units that are not necessarily equivalent to clauses [Carlson and Marcu, 2001, Tofiloski et al., 2009]. The process of identifying main and subordinate clauses is akin to identifying nucleus and satellite relations between discourse units. Future work will explore the performance of a dependency parser for such a task.

The relationship between main and subordinate clauses is important in understanding both the benefits and shortcomings of Centering Theory, and has received some study within Centering Theory literature [Kameyama, 1998, Miltsakaki, 2003, Poesio et al., 2004b]. Regarding subordinate clauses, Miltsakaki [2003] says "the term 'subordinate clause' covers a wide variety of clause types, ranging from non-finite verb forms such as infinitives and participles, *to* complement clauses, free relatives, conditionals, relatives, and adverbial clauses".

Deciding on the unit to be used for utterance is still something that has yet to achieve consensus within the discourse (and Centering Theory) community. What is generally agreed upon is that the sentence should not be the fundamental discourse unit (although Miltsakaki [2002] supports the sentence as the unit of utterance). Discourse units can be

difficult to identify, especially in spoken language. Mosegaard Hanse [1998] has also investigated different instantiations for the unit of utterance. Thompson and Couper-Kuhlen [2005] lend their support to the clause as the basic unit of discourse, stating that clauses are "understood as [predicate + phrases that accompany it], while sentence' is a term reserved for a unit that can consist of either a clause or a combination of clauses". Hence, by Thompson and Couper-Kuhlen's definition, the sentence is not necessarily an atomic representation. Kameyama [1998] and [2004a] discusses the problems involved when dealing with sentences as the unit of utterance in comparison to utterance units at the sub-sentential level. The different units that have been proposed by them are sentences and finite clauses. Poesio et al. perform an experimental analysis of the effects of using the sentence as a clause in comparison to finite clauses and their effects on Centering Theory results. They note that "identifying utterances with sentences may produce better Centering structures, but it is in disagreement with other theories of discourse" [Taboada & Hadic-Zabala, 2008, pg. 25]. Taboada & Hadic Zabala [2008] also make the argument for using the clause as a unit for the utterance.

The issue of sentences versus clauses as units of utterance is explored further in Section 3.3.1, where a model is proposed that is able to support both definitions simultaneously.

2.1.4   Definition of Entity

An entity is essentially anything which can be referred to. Typically it is a noun object or noun phrase. Poesio et al. [2004a] discuss some of the disagreements associated with what could be considered an entity. They also provide an example of the resulting coherence due to considering deictic words as well as ignoring deixis. An example is the following taken from Poesio et al. [2004a, pg. 13]:

(2.1)   You should not use PRODUCT-Z

(2.2)   if you are pregnant or breast-feeding.

(2.3)   Whilst you are receiving PRODUCT-Z....

If the second-person pronoun *you* is ignored then there is an incoherent entity transition between the three utterances.

For a computational system that is designed to automatically determine coherence, it is thus dependent upon the performance of the individual modules (e.g., recognizing entities and then identifying whether or not entities corefer), similar to the performance of discourse segmenters mentioned in Section 2.1.3. In order to determine what the entities in a text are, a part-of-speech tagger is first run on the input text. Coreference resolution is then applied, which resolves all referents (such as pronouns) with their references (coreference refers to two expressions being "referentially identical" [Strube and Hahn, 1999]). Thus, in an automated system, one's definition of entity is tightly coupled to the performance of existing systems. For the experiments performed in this paper, the definition of entity that was used were noun phrases and any referent resolved to a noun phrase[2]. Individual tokens that were tagged as noun phrases (NP) or nouns (N*) were treated as entities. Compound nominals that were not tagged as being a single coreferent unit were treated as distinct entities. Thus, due to software limitations, if a compound nominal is not tagged to corefer with another entity, such as 'Microsoft Corporation', then it is considered two separate entities: 'Microsoft' and 'Corporation'.

The core of this thesis revolves around entity coherence. It is thus important to be clear about what entity coherence does and does not entail. Entity coherence is to be distinguished from lexical coherence. Karamanis [2007, pg. 446] describes entity coherence as "the way NP referents relate subsequent clauses in the text." Lexical coherence involves lexical relationships between words (synonymy, antonymy, hyponymy, etc.). Entity coherence deals with entities that are considered to be distinct in reference from each other i.e., no further measurable relation exists between two entities. The lexical aspect of entities can be used in combination with entity coherence as both a preprocessing and a post-processing step. In a preprocessing stage, lexical coherence can be used to relate the entities that are not distinct, as in a mereological usage of the word "door" for "house" [Poesio et al., 2004a]. In a post-processing stage, lexical coherence can be used to aid in determining the relational coherence between utterances. The focus of this thesis will be entity coherence without any lexical coherence involved. Hence, all entities will be assumed to be distinct unless they've been identified to corefer by the coreference resolver. Entity coherence will be looked at further in Section 2.3.2, when Centering Theory is presented.

---

[2]The software OpenNLP was used for coreference resolution. http://OpenNLP.sourceforge.net

2.2 Definition of Coherence

This section discusses the differences between the two different scopes of coherence: *global* and *local* coherence. Also discussed are *coherence gaps* that result due to an incoherent text (specifically, entity incoherence), as well as what is involved in *bridging* such gaps.

2.2.1 Definition of Global Coherence

Grosz et al. [1995] define **global coherence** as coherence occurring between discourse segments. In the case of a text, where the text is considered to be the discourse with smaller parts of the text considered discourse segments (such as chapters, paragraphs or sentences), the global coherence would be the coherence that occurs above the level of utterances. In contrast to global coherence, coherence at the level of utterances is known as *local coherence.* Using the example of a text and its paragraphs, the text has an overall intention that it is trying to convey (discourse purpose), and each individual discourse segment is conveying something as well (discourse segment purpose). The effectiveness of the individual discourse segments on the overall discourse purpose would be a measure of the text's global coherence [Grosz et al., 1995, pg. 204]. In summary, Taboada and Hadic Zabala [2008, pg. 73] state:

> A discourse segment is recognizable because it always has an underlying intention associated with it. Discourse segments can also be embedded. They exhibit local coherence (among the utterances in the segment), and global coherence (with other segments in the discourse).

Embedded segments are also outside the scope of this thesis. A full study of how the intentions of the individual discourse segments combine to convey the intention of the overall discourse is not the focus of this paper. Such a study would involve aggregating the various types of coherence (relational, lexical, entity, inferential, etc.), whereas the focus here involves entity coherence at the global and local level.

Global coherence is *not* just the tracking of coherence over multiple utterances (beyond the restricted window of adjacent utterances): the global coherence is determined by the individual discourse segments and how they cohere to convey the overall purpose of the discourse. This presents a slight problem in relating global coherence with *global entity coherence* since the tracking of entities over a text does not involve an "entity transformation" in the same way that utterances in a discourse segment collapse to form a unique

discourse segment purpose. Since the focus of this thesis is entity coherence, reference to *global coherence* will involve tracking entity coherence over the entire text, rather than in the transformative way that utterances are no longer referenced when discussing overall discourse purpose (or global coherence).

2.2.2 Definition of Local Coherence

Grosz et al. [1995] describe local coherence as coherence occurring between utterances within a discourse segment. The focus of Centering Theory (discussed in Section 2.3.2) was to attempt to capture the relationship between local coherence and the discourse participants' focus of attention ("attentional state"). Local coherence attempts to determine the level of coherence between utterance pairs, where an incoherent utterance pair would represent something akin to non-sequitirs (e.g., "I went to the store. Canada is part of North America.").

The main focus of this thesis involves entity coherence at the local level, between adjacent utterances. At this level we utilize the grammatical role of the entity and how it changes role between utterances as a way to determine local coherence. The overall coherence of the text is thus determined by the sum of all the local coherence values. Again, this is not to be confused with the global coherence of the text, which is coherence between discourse segments and their discourse intentions. Global entity coherence is further discussed under *recency*, in Section 3.3.1.

2.2.3 Definition of Coherence Gaps

A **coherence gap** occurs between two utterances when the pair does not exhibit **maximal coherence**. Maximal coherence can be defined as something which does not leave the reader attempting to fill gaps in the text i.e., there is nothing in the text the reader does not understand. The notion of maximal coherence is not difficult to define, but attempting to measure or recognize it is not as straightforward. It is also reader dependent, since the reader's knowledge is used as part of the evaluative basis involved in *understanding* the text. This also illustrates the contrast between coherence and cohesion, where the kind of gaps that are *cohesive* can be resolved from within the text, and the kind of gaps that are *coherent* are to be resolved using information found outside the text. As Graesser et al. further articulate [2004, pg. 1]:

> The coherence relations are constructed in the mind of the reader and depend on the skills and knowledge that the reader brings to the situation. If the reader has adequate world knowledge about the subject matter and if there are adequate linguistic and discourse cues, then the reader is likely to form a coherent mental representation of the text. A reader perceives a text to be coherent to the extent that the ideas conveyed in the text hang together in a meaningful and organized manner.

Again, for simplicity, the terms 'cohesion' and 'coherence' are used interchangeably in this thesis and are used only in the cohesive sense, where the information used to resolve any gaps comes from the text itself.

The 'gap' between the two different kinds of coherence gaps (*within* and *without* the text) is not as simply demarcated as we are making it appear. As Hirst explains regarding the reader/writer interface [2008, pg. 5]:

> Ironically, in the traditional, more logic-based approaches of the mid-1970s to the mid-1980s, the dominant view was the most postmodern one, the reader-based view: language understanding was viewed as a knowledge-based enterprise, and so a text was understood by a system in the light of its own particular knowledge. This was so at both the sentence-meaning level and the text-meaning level; in the latter case, exemplified by the work of Schank and his colleagues... to understand language was to use knowledge and inference to 'fill in the gaps' in what was said. The reader-based view became most explicit in Corriveau's computational model of time-constrained text comprehension.

Here, interrogating the gaps are manifest in the questions "What is the author trying to tell me?" (writer-based), and "What does this text mean to me personally?" (reader-based)[3]. Negotiating these gaps would possibly take the form of an interplay between the two orthogonal views. An answer to one of the questions may reveal answers to the other, and vice versa.

---

[3]Hirst actually mentions three avenues available for answering the posed questions. They are: the reader, the writer, and the text itself. For simplicity, the text and the writer have been combined into one. The assumption here is that the "interrogation" does not have access to the original author and thus cannot discover its intent. Hence, reducing the writer to what can be found in the text seems a reasonable reduction when dealing with the discourse of written texts. See Hirst for a further discussion.

There are different types of coherence gaps each corresponding to the different types of coherence one is investigating. The different kinds of coherence are: grammatical, temporal [Lascarides and Oberlander, 1992, Kameyama et al., 1993], syntactic [Kehler, 2002], inferential [Asher & Lascarides, 2003, Fais, 2004], referential [Mann and Thompson, 1988], entity-based [Grosz et al., 1995, Poesio et al., 2004a], and lexical [Fais, 2004]. Since this thesis focuses on entity coherence, only entity coherence gaps are given attention. With entity coherence gaps being much more straightforward to detect than the other types of gaps, such as inferential gaps or reasoning gaps, detection of entity coherence gaps can be used as an initial stage in identifying the type of gap involved.

In their simplest form, entity coherence gaps arise when two utterances have no mentioned entities in common. Utterances having at least one entity in common can be considered to contain less of a gap. The more entities two utterances have in common, the less of a gap there exists between them. By itself, entity coherence gaps are not enough to fully characterize the coherence gap, but they are the simplest kind to observe computationally (no reliance on world-knowledge inference engines or lexical relational measures), provided the parameters and criteria for computing the size of the gap have been determined.

2.2.4 Definition of Entity Bridging

A **bridge** is a structure used to minimize a coherence gap. There are many different kinds of coherence gaps (lexical, grammatical, inferential, etc.), with each having their corresponding bridge. Discourse literature [Clark, 1977, Asher & Lascarides, 2003] has primarily focused on *inferential bridging*, where "two objects or events are related in a particular way that isn't explicitly stated" [Asher & Lascarides, 2003, pg. 18] using the *logical* notion of inference. Asher and Lascarides believe that bridging inferences is dependent on rhetorical structure and rhetorical relations. Their definition of inferential bridging is actually a general definition of bridging which they assume resolves all coherence gaps through inference. This notion of inferencing can be interpreted as *anything capable of bridging the gap*. They mention the co-dependence on relational coherence and generating bridging inferences, though the bridging is not dependent on relations and logic alone. We thus drop the qualifier *inferential*, and refer to this general relationship between two sentences solely as a *bridge*.[4] The focus in this thesis will be on **entity bridges**.

---

[4]An example of a gap not involving inference for its bridge is *'John already saw the movie. Steve didn't, but they still went to see another one instead.'* In this example, part of the sentence is omitted from explaining

Entity bridges have also been referred to as **cohesive ties** [Halliday & Hasan, 1976], which are a more specific type of entity bridging. Halliday and Hasan [1976, pg. 9] observe that

> ... cohesive ties between sentences stand out more clearly because they are the only source of texture, whereas within the sentence there are structural relations as well. In the description of a text, it is the intersentence cohesion that is significant, because that represents the variable aspect of cohesion, distinguishing one text from another.

They define a tie as "a term for one occurrence of a pair of lexically cohesive items" [Halliday & Hasan, 1976, pg. 3]. The types of ties they describe are: referential (pronouns, anaphor), repetition of words, substitution, ellipsis, conjunction and lexical cohesion [Halliday & Hasan, 1976, pg. 4]. Ties that fall under their classification of *repetition of words* are, for all intents and purposes, equivalent to the notion of entity bridges. When entities are repeated in adjacent utterances, it can be said that the entity in common is part of what binds the two utterances together. Fais [2004] also discusses lexical cohesion in terms of bridging, through the notion of bridging inferables. Fais says "bridging references, however, unlike the usual case of anaphora, may be mediated not only by a strict identity condition, but also by any number of other semantic relationships (is-a, has-a, made-of, at-time, etc.)" [2004, pg. 126]. In Fais' terms, we only concern ourselves with entity coherence, those entities which are explicit (i.e., non-inferable). The other types of ties, along with their associated gaps, will not be further discussed.

The idea behind a **bridge** is to close a gap occurring between two utterances. The purpose of an *entity bridge* is to modify the transition type occurring between two utterances from one that is less preferred (i.e., less coherent) into a more preferred transition. In the language of Centering Theory (see Section 2.3.2), the objective would be to move from a

---

that *Steve didn't see the movie*, and thus falls under ellipsis and not under inference. An example of a lexical bridge is in the sentences *'Steve didn't buy the house. He didn't like the kitchen.'*, where there is a mereological relationship between *house* and *kitchen*. The *kitchen* is understood to be a part of the *house*. This can be contrasted with the sentences *'Steve didn't buy the house. He didn't like the car.'*, where there is no lexical relationship between *house* and *car*. The example which Asher & Lascarides provide [2003, pg. 20] illustrates this further: *'John loves to collect cars. But his favourite car is a 1999 Ford Mondeo.'* which is more coherent than *'John loves to collect cars. But he hates football.'* Both examples use a *Contrast* relation, but the first is considered a stronger coherence due to its stronger lexical coherence. They refer to this stronger coherence as being part of "an interpretation which maximises the quality of its rhetorical relations."

rough-shift or *NOCB* transition to a *continue* or *smooth-shift* transition. The bridging of the gap that occurs between two utterances can be implemented in one of two ways (or a combination of the two):

- insert a new utterance between the gapped utterances such that the resultant sequence of transitions are preferred over the original transition

- modify one (or both) of the utterances such that the transition, and hence coherence, is improved

The former approach will be investigated as a way to bridge coherence gaps as measured by entity transitions. Inserting a new utterance between two gapped utterances in an attempt to minimize the gap will be explored in Section 4.

2.3 Related Work

In this section we provide an overview of entity-based coherence models. We also provide some other theories that do not explicitly deal with coherence, but indirectly through the way anaphora is resolved. These theories provide motivation for different aspects of our model, such as salience (Section 3.1.3) and recency (Section 3.3.1).

2.3.1 Discourse Focus Stacks

Grosz & Sidner describe the structure of discourse as being "a composite of three interacting constituents: a linguistic structure, an intentional structure, and an attentional state" [1986, pg. 177]. The **linguistic structure** refers to the sequence of utterances in the discourse, which are the basic elements of the linguistic structure. The linguistic structure is the surface aspect of the relations between utterances based on the intentional structure [Grosz & Sidner, 1986]. This basically amounts to the relations that are involved between the linguistic elements (utterances). The study of this aspect of discourse is primarily the domain of relational coherence. The **intentional structure** refers to the overall intentions or purpose of the discourse.

The focus within this thesis is best captured by the final constituent, that of the **attentional state**. The attentional state is intended to capture the "participant's focus as

the discourse unfolds" [Grosz & Sidner, 1986, pg. 179]. We discuss the ideas behind the
attentional state, with the linguistic structure and intentional structure being outside our
study[5].

One might assume that the attentional state is a product of what is important to the
participant's attention, but Grosz & Sidner clarify this point [Grosz & Sidner, 1986, pg.
179]:

> The attentional state is a property of the discourse itself, not of the discourse
> participants. It is inherently dynamic, recording the objects, properties, and
> relations that are salient at each point in the discourse. The attentional state is
> modeled by a set of focus spaces; changes in attentional state are modeled by a
> set of transition rules that specify the conditions for adding and deleting spaces.
> We call the collection of focus spaces available at any one time the focusing
> structure and the process of manipulating spaces focusing. The focusing process
> associates a focus space with each discourse segment; this space contains those
> entities that are salient - either because they have been mentioned explicitly
> in the segment or because they became salient in the process of producing or
> comprehending the utterances in the segment.

This approach allows one to minimize the difficulty involved in having to model both the
structure of the text as well as the variables involved in modeling the participants as well.
This restricts the study to the text itself and alleviates one from the complexities involved
in having to model the participants. The transition rules that occur between utterances
will be further discussed in Section 2.3.2, which introduces and discusses Centering Theory.
Grosz & Sidner provide a further overview of the attentional state [1986, pg. 177]:

> Attentional state contains information about the objects, properties, relations,
> and discourse intentions that are most salient at any given point. It is an ab-
> straction of the focus of attention of the discourse participants; it serves to sum-
> marize information from previous utterances crucial for processing subsequent
> ones, thus obviating the need for keeping a complete history of the discourse.

---

[5]Even within the definitions, there exists some overlap between the linguistic structure and the attentional
state in the way that utterances are related to each other through their entities. For simplicity, we demarcate
between the linguistic structure of a discourse and its attentional state by stating that the tracking of entities
across utterances is solely the domain of attentional state.

The important aspect of this quotation occurs in the last line with the implication of a difficulty or complexity involved with maintaining the entire history of the discourse. We have interpreted this line as illustrating that one of the advantages of their proposed focus stack is how it simplifies the tracking of multiple potential foci (as well as the focus history), and not a comment on the impossibility of tracking all potential foci and focus histories. One of the contributions of this thesis is in being able to capture the discourse's entire history of entity salience as well as accounting for a single most-focused entity if need be.

A **focus stack** (also referred to as a focus space stack) is used to represent **information** from a discourse. *Information* refers to the "objects, properties, and relations" of the discourse. The data structure used to capture the information is a stack, where items on the stack can be pushed and popped. Information on higher spaces is more accessible than information on lower spaces.

Clarifying the terms also requires attention. As Arnold explains [Arnold, 1998, pg. 5]:

> 'Focus of attention' is not the same as the term 'focus' which is used in descriptions of information structure, and roughly refers to the new, highlighted information in a sentence. Instead, 'focus of attention' refers to the item which is currently most salient in the discourse in terms of pronoun reference, which corresponds more to the discourse topic than the focus.

In this thesis, the words "focus", "attention", "salience", and "discourse prominence" are used interchangeably. The caveat being that there are a few different instantiations of "salience", based on the context under consideration. This includes salience within a sentence (the most salient entity within an utterance), local salience (the most salient entity at a given point in the discourse within a window of utterances), and global salience (the most salient entity over the entire discourse). These notions will be captured using a combination of salience (Section 3.1.3) and recency (Section 3.3.1).

The focus space model structures the interaction of discourse segments in terms of their purpose. Defining and identifying a focus space is non-trivial. The next section discusses the constituents of such discourse segments, namely *entities*, and how they contribute to the focus of the discourse segments.

2.3.2 Centering Theory

Centering Theory [Grosz et al., 1995] is a theory that evaluates local coherence by tracking entity movement between sentences. It extends the ideas behind Grosz & Sidner's [1986] focus spaces and stacks discussed in the previous section. As Miltsakaki explains [2003, pg. 18]:

> The notion of focus space is, also, elusive. What is a focus space and how is it identified? Is the focus space equivalent to an abstract segment associated with a discourse purpose or is it an attentional update unit? A first attempt to model aspects of attentional structure yielded a reformulation of Centering as a model of local discourse coherence... Centering was developed as a model of the center of attention between speakers in natural language discourse. The model aimed at modeling the interaction between 'attentional state', inferential complexity and the form of referring expressions.

Centering Theory classifies the degree of coherence based on sequences of utterance transitions. The degree of coherence classifies the utterance transitions into four basic types: *Continue*, *Retain*, *Smooth-shift*, and *Rough-shift*. These and other transition types are discussed in detail in Section 3.2.

Centering Theory evaluates coherence at the *local* level, which is the level between adjacent utterances. Centering tracks the movement of a single entity (the most prominent entity, which could be different from utterance to utterance) as it changes in **salience** from utterance to utterance. *Salience* refers to the prominence an entity has in an utterance. Salience can manifest itself as the entity's grammatical role (e.g., subjects are ranked more salient than objects), its linear order (e.g., entities being mentioned first in an utterance are considered more salient irrespective of their grammatical role), or a combination thereof[6] (Section 3.1.3 provides a complete discussion on salience), as well as through the way an entity is realized (see Section 2.3.3).

In an utterance U, the most salient entity contained within it is referred to as the **preferred center**, abbreviated as C_p. Centering Theory tracks the movement in salience of

---

[6]In the case of conjoined subjects, such as *Tom and Mary are a happy couple*, both *Tom* and *Mary* assume the highest salience, and the tie can be broken by considering which entity is mentioned first (in this case, *Tom*).

	$C_b(U_n) = C_b(U_{n-1})$ or $C_b(U_n) = \emptyset$	$C_b(U_n) \neq C_b(U_{n-1})$
$C_b(U_n) = C_p(U_n)$	Continue	Smooth-shift
$C_b(U_n) \neq C_p(U_n)$	Retain	Rough-shift

Table 2.1: Table of Standard Centering Theory transitions.

the preferred centers from one utterance (U_{n-1}) to the next utterance (U_n). The backward-looking center, abbreviated as C_b, is the most salient entity in U_{n-1} that is also **realized** in U_n. When an entity is *realized* in an utterance, it means that it has been mentioned in the utterance. In addition to the backward-looking center is a list of entities in U_{n-1} that are candidates to become the C_b of U_n. This list of candidates are called the **forward-looking centers**, abbreviated as C_f, which is a list of all the entities from U_{n-1} that are ranked according to their salience in decreasing order.

We provide a quick overview of the traditional Centering Theory transitions. Section 3.2 discusses an extended set of transitions.

Continue - this occurs when an entity has been the preferred center for three consecutive utterances (including the current utterance). $C_b(U_i) = C_b(U_{i-1})$ and $C_b(U_i) = C_p(U_i)$

Retain - occurs when there is a new preferred center yet the previous preferred center occurs in the current utterance. $C_b(U_i) = C_b(U_{i-1})$ and $C_b(U_i) \neq C_p(U_i)$

Smooth-shift - occurs when the preferred center of the current utterance was the highest ranked entity in the previous utterance that is realized in the current utterance the backward-looking centers of the previous two utterances are not the same entity. $C_b(U_i) \neq C_b(U_{i-1})$ and $C_b(U_i) = C_p(U_i)$

Rough-shift - the preferred center was not the highest ranked entity in the previous utterance that also appears in the current utterance, as well as not having the same backward-looking centers. $C_b(U_i) \neq C_b(U_{i-1})$ and $C_b(U_i) \neq C_p(U_i)$

The above transitions can be summarized in Table 2.1.

There is a distinction between the preferred center (C_p) and the backward-looking center (C_b). The C_p can be viewed as the topic (also can be referred to as the focus or the most prominent entity) of an utterance, whereas the C_b can be viewed as the topic of two adjacent utterances. The backward-looking center is intended to capture the current topic based on

the previous utterances. There is no reverse notion in Centering Theory which has a center of the current utterance that is based on the entities realized in the next utterance.

The most preferred type of entity transition between utterances is one where the C_p is the most salient in both of the adjacent utterances (e.g., the same entity is the grammatical subject in both utterances) and also that the C_b is identical in the two utterances. The idea is that the most coherent transition between utterances U_{n-1} and U_n is one where the entity structures are identical between the two utterances i.e., the same entity maintains being the most salient across three consecutive utterances.

Centering Theory proposes using a single C_b. The justification for this is that every utterance has a "unique main link" with its previous utterance. This assumption "simplifies the complexity of the inferences required to integrate an utterance into the discourse" [Poesio et al., 2004a, pg. 311]. One of the contributions of this thesis is to propose a method of tracking all entities and illustrating the benefits gained from such an extension to Centering Theory. This is not a comment as to whether or not an utterance, or a discourse segment, contains a single entity as its focus. Rather, the purpose is to illustrate the benefits in tracking salience change of all entities in the spirit of Centering Theory in order to better evaluate the coherence of a text. Extending Centering Theory to track multiple entities allows Centering Theory to extend its utility into other applications and tasks, not only anaphora resolution.

The C_f *list* is a list of entities, whereas the C_b is a singleton. The justifications used by Grosz et al. for a unique C_b involves pronoun resolution, where the objective is to identify the entity which a pronoun refers to [Grosz et al., 1995, pg. 211]. The "link between pronominalization and the identity of the C_b has been used by Grosz et al. to support the claim... that utterances have a unique C_b" [Poesio et al., 2004a, pg. 312]. Further, after providing examples illustrating the effect of grammatical role on resolving referents, Grosz et al. state [1995, 214]:

> In summary, these examples provide support for the claim that there is only a single C_b, that grammatical role affects an entity's being more highly ranked in C_f, and that lower-ranked elements of the C_f cannot be pronominalized unless higher-ranked ones are. Kameyama [1985] was the first to argue that grammatical role, rather than thematic role, which Sidner used, affected the C_f ranking. Psycholinguistic research since 1986 [Hudson-D'Zmura, 1988, Gordon et al., 1993] supports the claims that there is a single C_b and that grammatical role plays

a determining role in identifying the C_b. It furthermore suggests that neither thematic role nor surface position is a determinant of the C_b. In contrast, both grammatical role and surface position were shown to affect the C_f ordering.

Others have argued for allowing utterances to contain either multiple topics, or to allow there to be a gradient measure of topicality associated with the entities in an utterance [Givon, 1983, Arnold, 1998, Lappin and Leass, 1994]. Arnold states that "the idea that some entities are more topical than others has its advantages, which might be retained if topicality is considered as a gradient property, and is not limited to a single entity per utterance" [1998, pg. 6]. Regarding unique C_b's, Arnold continues [1998, pg. 39]:

> Attributing special status to the top-ranked entity is a problematic aspect of Centering, for three reasons. First, the difficulty of identifying a unique C_b or C_p in all cases makes the claim difficult to test. Second, non-subjects can often be more salient than subjects. For example, [Smyth & Chambers, 1996] showed that in some conditions, equal repeated-name penalties can be found for both object- and subject-referents, and that reading times for a pronoun are faster when it refers to a parallel object-referent than to a nonparallel subject-referent. Third, Centering Theory stipulates that if a pronoun is used to refer to the C_b, other pronouns are acceptable (but not necessary), but it makes no predictions about if or when a pronoun will be used to refer to a non-C_b.

Note that the C_b is not strictly synonymous with the notion of 'topic'. In terms of a graded topicality, the C_p would be the closest to the definition of an utterance's topic. And since the C_p is derived from the C_f list, which is an ordering of all the entities in an utterance, the C_f list is thus the Centering measure which is identical to the notion of a graded topicality of all the entities within an utterance. Hence, the C_b cannot be viewed as the topic of an utterance but, as mentioned before, an *intersection of the topicality* between two utterances. The point is subtle, yet important in mapping the framework of one theory to another.

The transitions that occur between utterances are used to characterize the degree of coherence. Grosz et al. state that "the coherence of a segment is affected by the kinds of centering transitions engendered by a speaker's choices of linguistic realizations in the utterances constituting the segment" [1995, pg. 210]. In terms of preferred types of transitions, the transitions that maintain a single entity as most salient across utterances are desired

over transitions that have a different entity as being most salient. As mentioned before, the transitions originally proposed by Centering Theory (*continue*, *retain*, and *shift*) are not strictly between two adjacent utterances, since they use the $C_b(U_{n-1})$ in their calculation, which is dependent upon U_{n-2}. Centering Theory's transitions will be investigated in detail in Section 3.2.

Centering Theory and anaphora resolution are coupled fairly tightly, rightfully so, since it has provided a successful foundation for the area of coreference resolution. One of the objectives of this thesis is to loosen the ties between Centering Theory and anaphora resolution so as to generalize the entity model so that the concepts can be applied to other experiments, such as determining the quality of discourse coherence. The coupling can be found evidenced in one of Centering Theory's tenets, known as **Rule 1**, which describes the realization and salience of pronouns. Rule 1 states [Grosz et al., 1995, pg. 214]:

> If any element of $C_f(U_n)$ is realized by a pronoun in U_{n+1}, then the $C_b(U_{n+1})$ must be realized by a pronoun also.

In other words,

> Rule 1 represents one function of pronominal reference: the use of a pronoun to realize the C_b signals the hearer that the speaker is continuing to talk about the same thing... Psychological research [Gordon et al., 1993, Hudson-D'Zmura, 1988] and cross-linguistic research [Di Eugenio, 1998, Kameyama, 1985, Kameyama, 1986, Kameyama, 1988, Walker et al., 1994, Walker et al., 1990] have validated that the C_b is preferentially realized by a pronoun in English [Grosz et al., 1995, pg. 214].

Again, due to the success of Centering Theory in coreference resolution, the strong association between the two is justified. The main problem with restricting the model to having a unique C_b is that because of the strong coupling to anaphora resolution, where for the most part a referent *by definition* refers to a single previously mentioned entity[7], the model may not translate to domains where tracking multiple (or all) entities is necessary. Hence, the purpose of this thesis is to extend Centering Theory to be able to track all entities between utterances and show the benefits of such a model.

---

[7]The exception being *they*, which can refer to multiple entities.

Which entities exist in the C_f list is open to discussion as well. Does the C_f list consist of only those entities which are directly realized within an utterance? This is where lexical cohesion can be used to add more entities to the C_f list in order to provide better coverage of the possible entity realizations. Fais relates lexical cohesion and inferential cohesion through the notion of **inferable entities**, which are described as "entities that are not expressed at the surface level of the utterance or immediately recoverable from the subcategorization properties of the verb (as, for example, zero pronouns are) to constitute centers of an utterance" [2004, pg. 121]. She continues:

> However, if we process discourse incrementally, this leads to the conclusion that since we do not know which inferable entity from U_{i-1} will be evoked in Ui, we need to list every inferable entity in the C_f list of U_{i-1}. This is both computationally untenable and, in view of the lack of any parameters for determining what constitutes an allowable inferable C_b, impossible. Even if it were possible and desirable, how could we define for inferable entities the type of grammatical-role information essential to determining the placement of these entities on a C_f list?

Fais also mentions the import of lexical cohesion on measuring coherence, which our experiments later validate. An essential component to accurately determining coherence is a model that incorporates lexical cohesion. Fais performs a corpus study which reveals that [Fais, 2004, pg. 119]:

> Because this corpus contains a high number of discourse elements that are inferable from the discourse context, we have an opportunity to examine the interplay between standard centering transition definitions and the presence of inferable discourse entities. We claim on the basis of intuitions of native speakers that the actual level of coherence in the corpus is much higher than the Centering account implies, primarily by virtue of the fact that transitions involving inferable entities are often difficult to specify. We conclude that the standard Centering account cannot accurately model the coherence in this corpus. Detailed analysis reveals that one major problem lies in the requirement of identity of discourse elements in adjacent utterances in order for those elements to contribute to coherence.

The model proposed in this thesis does not incorporate lexical cohesion, and is left for future study.

2.3.3 Treatment of Pronominalization

This section provides a brief overview of various approaches to the treatment of entities that are realized in pronominal form and its effect on salience. An example of a text containing a pronominal is the following:

(2.4)   Steve stole a computer from work.
        The police arrested him this morning.

The ranking of entities for the second sentence, according to either a linear order or based on grammatical role, from most salient to least is: *police > him (Steve) > morning. Steve* is realized as *him* in the second sentence. Thus, if an entity is realized as a pronoun, then it can be a candidate for receiving a boost in its salience, which may effect the ordering of the entities. A new ordering of entities in terms of a salience that incorporates boosting the entity salience due to pronominalization is: *him (Steve) > police > morning.* The issue then becomes when to boost an entity's salience and by how much. We leave for future work the implementation of such a method for handling entities that are pronominalized, and briefly mention some of the different theories of pronominalization and their effect on entity ranking.

In Givon's Scale of Topicality [Givon, 1983], topicality is treated as a continuum. Our approach is motivated by the notion of a graded topicality. Givon also does not restrict the topicality to one topic per utterance, which we also incorporate in our model. The following linguistic attributes listed by Givon are ordered from most continuous/accessible topic to most discontinuous/inaccessible topic (excerpted from [Miltsakaki, 2003]):

- zero anaphor

- unstressed/bound pronouns or grammatical agreement

- stressed/independent pronouns

- right dislocated definite NPs

- neutrally ordered definite NPs

- left dislocated definite NPs

- Y-moved NPs ('contrastive topicalization')

- cleft/focus constructions

- referential indefinite NPs

A simpler and more general form of the Scale of Topicality is in Givon's grammar of topic identification is in the following:

- zero anaphora

- unstressed/bound pronouns ('agreement')

- stressed/independent pronouns

- full NPs

In Accessibility Theory, which is useful for pronoun resolution, Ariel [1988, pg. 65] identifies four factors relevant in processing anaphoric expressions, known as the Accessibility Hierarchy:

- **distance** between antecedent and anaphor

- **number of competitors** for the role of antecedent

- importance of **topicality** in antecedent assignments

- **role of frames** in identifying antecedents

Even though anaphora resolution is not the focus of this thesis, there are many relevant factors introduced by Ariel that can be applied to coherence. Ariel's Accessibility Hierarchy is another graded scale, even though the four factors are not necessarily graded themselves.

What is important here is the notion of referential distance, which is discussed further in Section 3.3.1. This is used to show or illustrate the idea of extending the idea of local coherence to be in a space that is most suitable for the entities (e.g., within a paragraph, or a chapter) rather than restricting the notion of local coherence to only be contained between adjacent utterances. 'Topicality' is essentially another word for salience, which is further discussed in Section 3.1.3. 'Topicality' is proposed as a binary measure, whether something is 'on-topic' or 'off-topic'. The 'number of competitors' is exclusively the domain of anaphora resolution, and is not discussed here further.

The **Givenness Hierarchy** is defined by Prince [1981] in three ways: *predictability/recoverability*, *saliency*, and *shared-knowledge*. A brief description of each concept is provided as follows:

predictability/recoverability (givenness$_p$) [Prince, 1981, pg. 226] "The speaker assumes that the hearer can predict or could have predicted that a particular linguistic item will or would occur in a particular position within a sentence."

saliency (givenness$_s$) [1981, pg. 228], givenness$_s$ is something which "the speaker assumes that the hearer has or could appropriately have some particular thing/entity/... in his/her consciousness at the time of hearing the utterance".

shared-knowledge (givenness$_k$) "The speaker assumes that the hearer 'knows', assumes, or can infer a particular thing (but is not necessarily thinking about it)" [Prince, 1981, pg. 230]. What is given is what the speaker believes the listener already believes, what is assumed to be knowledge common to both participants.

Givenness is sometimes known as given-new, old-new, known-new, presupposition-focus. It is presented in terms of **pronoun resolution**, where the objective is to identify the entity which the pronouns ('he', 'she', 'they', etc.) refer to in a text.

The Givenness Hierarchy can be written as such (Gundel et al. [1993]):

in focus > activated > familiar > uniquely identifiable > referential > type identifiable

In English, the hierarchy is realized as such:

it > that, this, this *entity* > that *entity* > the *entity* > indefinite this *entity* > a *entity*

Kameyama [1998] further reduces this hierarchy as:

Zero Pronominal > Pronoun > Definite NP > Indefinite NP

Notice that this extends the notion of how salience of an entity is affected not only by its realization as a pronoun, but also by its realization as a referent. A study of the effects of referential realization is beyond the scope of this thesis.

One of the main objectives in this overview is to establish a hard distinction between coherence and coreference resolution. Even though there is significant overlap between coherence and coreference resolution (e.g., choosing the most appropriate referent is based on the one that makes the 'most sense', i.e., is the most coherent), coreference resolution can be considered a problem separate from assessing coherence[8]. As discussed above, the

---

[8]It would be easiest to consider this a *coherence after coreference*. In other words, after all the referents have been resolved, how coherent is the text? Thus, coreference resolution is treated as a module that is applied to the text *prior* to the evaluation of the text's coherence. This is the approach that is taken in all subsequent experiments.

effects on salience due to an entity's manifestation as a pronoun can be used as a feature in coherence evaluation[9]. The important contribution that the above approaches to coreference resolution provide is the notion of entity salience/focus and discourse prominence as factors contributing to their resolution. Hence, the overlap in entity salience between coreference resolution and coherence allows one to utilize coreference resolution's principles towards the evaluation of coherence. This thesis illustrates that these mechanisms provide an excellent starting point, yet still require a generalizability in order to be applicable and robust toward the evaluation of coherence.

2.3.4 Coh-Metrix

Another type of coherence metric, which is separate from the lineage of Centering, is known as Coh-Metrix [Graesser et al., 2004]. Coh-Metrix measures coherence by performing an empirical analysis of the linguistic aspects of the text. The approach that Graesser et al. take is commendable in that they leverage as much linguistic information about the text that they possibly can.

The difficulty with the Coh-Metrix approach of making as many measurements of the text as possible is that there is (as of yet) no practical way to reduce all such measures to a single score. What has been gathered is a set of empirical data, with the objective being to determine whether the data represents a text that is coherent or not. As an example, it is as if one has made many measurements about a person's features, such as their height, weight, eye colour, favourite food, etc., but the objective is to determine whether the sum of these features actually describe an actual human being (coherent data) or a non-existent person (incoherent data).

Regardless of the shortcomings associated with a purely empirical analyses of discourse, the contributions of Coh-Metrix are nonetheless important. The attributes of a text that Graesser et al. measure are all valid in terms of assessing coherence, with nearly every single measure motivated as the result of a psychological study of coherence in text.

The features that Coh-Metrix uses are described below:

---

[9]Although sometimes an entity being realized as a pronoun is more a matter of convention or to avoid a stiltedness in the text, such as can be found in the **repeated-name penalty**, which considers a text that repeats entities as being less coherent (*'James went to the store and James bought some food for James' dog'*, which is not as fluent as the more conventional *'James went to the store and he bought some food for his dog'*).

genre two types of genre considered were informational and narrative (other types of genre that were not used by Coh-Metrix are description, expository, and persuasion)

word frequency the idea is that frequent words are read more quickly and better understood

word information considers frequency of a word in print; measures the concreteness or abstractness of a word; *meaningfulness* of words (using Colorado meaningfulness [Toglia & Battig, 1978] and Paivio meaningfulness [Paivio et al., 1968] measures); age of acquisition of a word

part of speech over 50 part-of-speech (POS) tags (words tagged as either adjectives, verbs, nouns, adverbs, etc.) from the Penn Treebank and Brill tagger were used in Coh-Metrix; incidence score for each POS category is the number of occurrences per 1000 words; Coh-Metrix makes the distinction between content words and function words; this measure used to see how much substantive content there is in the text

density scores incidence, ratio, or proportion of particular word classes (per 1000); the idea is that the density of pronouns are important for potential comprehension difficulty

logical operators such as *or, and, not, if-then*; high density of logical operators correspond to an analytically dense text

connectives (density of connectives) clarifying (*in other words, that is*); additive *also, moreover*; temporal *after, before, when*; causal *because, so, consequently*; positive vs negative connectives *however, in contrast, although*

type:token ratio - each unique word is a type, each instance is a token; if 'dog' appears 7 times, its type value is 1 and its token value is 7; the measure is a ratio of number of unique words per number of tokens; if the ratio is 1, the comprehension is thus difficult because many unique words need encoding; computed for content words but not function words

polysemy # senses of a word; *hypernym* counts the number of levels from root to concept in an ontology; many hypernym levels represent a more concrete object, whereas less levels represents a more abstract concept

concept clarity measure of ambiguity, vagueness, and abstractness

syntactic complexity how difficult to analyze syntactic composition of sentences; NP density is the number of modifiers per NP; mean number of high-level constituents per word; incidence of word classes that signal logical or analytical difficulty

readability Flesch Reading Ease and Flesch-Kincaid Grade Level scores [Klare, 1974] measure the readability of a text

co-reference cohesion when a noun, pronoun, or NP refers to another constituent in the text; *When water is heated, it boils and evaporates. When the heat is reduced, it turns back into a liquid*; noun overlap between sentences; two nouns with common stem is called argument overlap

causal cohesion extent which sentences are related by causal cohesion relations; many causal verbs may mean causal content; causal signals *so that, because, cause of, as a consequence, cause, enable, make*; metric is the ratio of causal particles to causal verbs

latent semantic analysis measures similarity between text units using Latent Semantic Analysis [Landauer and Dumais, 1997]

2.3.5   Summary of Related Work

The different kinds of coherence metrics span from tracking the salience of a single entity within a text (Centering Theory) to one that classifies a text according to its linguistic attributes (Coh-Metrix). The majority of the measures discussed have been traditionally used for the purpose of anaphora resolution. Also discussed were various ways entities can be realized and their potential effect on salience.

2.4   Summary

This chapter outlined the terminology, ideas, and motivations used in the rest of this thesis. A discussion of entity coherence and discourse, how they are related, and the different approaches were investigated in order to provide a basis for the ideas that follow. A computationally implementable approach of measuring entity coherence, Centering Theory,

leads us to the model presented in this thesis in the next chapter. The proposed model extends Centering Theory by tracking all the entities realized in an utterance, rather than only tracking the most salient entity, as well as allowing for the tracking of all previous utterances through the notion of *recency*.

Chapter 3

Measuring Entity Coherence

3.1 Vector Representation of Sentence Entities

Sentences and utterances can be represented as vectors, with the elements of the vector corresponding to the words of the sentence. Consider the example text "Steve likes Mary. Mary is a Republican." This text is represented as a matrix (a series of vectors) as illustrated in Table 3.1. Matrix representations have also been referred to as *entity grids* [Barzilay & Lapata, 2008] or *text grids* [Fillipova & Strube, 2004]. In the first few examples we represent all words in the matrix, independent of whether the word is an entity or non-entity.

In Table 3.1, sentence 1 (**S1**) has a vector representation of $(X, X, X, -, -, -)$, where (X) corresponds to the entity being realized in the sentence, and (-) corresponds to the entity not being realized. Thus, the words 'Mary', 'Steve', and 'likes' all appear in sentence

	Words In Sentence					
	Steve	likes	Mary	is	a	Republican
S1	X	X	X	-	-	-
S2	-	-	X	X	X	X

Table 3.1: Basic representation of a text as a series of vectors that does not distinguish between entities and non-entities. All words of the text are considered in the vector representation. Each row corresponds to a sentence. The columns correspond to whether the word was realized in the sentence (X) or not (-).

	Words In Sentence					
	Steve	likes	Mary	is	a	Republican
S1	3	2	1	0	0	0
S2	0	0	4	3	2	1

Table 3.2: A representation of the text incorporating linear order. The text is represented as a series of vectors that includes the salience/prominence of the word in the sentence. Larger values correspond to a higher prominence. Each row corresponds to a sentence (S1, S2). The columns correspond to the salience of that word in the sentence.

	Words In Sentence		
	Steve	Mary	Republican
S1	2	1	0
S2	0	2	1

Table 3.3: Matrix of the text "Steve likes Mary. Mary is a Republican." with non-entity words removed from the representation.

1. Sentence 2 has the vector representation of $(-, -, X, X, X, X)$, which reflects the words that are realized in the sentence.

From Table 3.2, one can see that non-entity words (adjectives, verbs, and stopwords) effect the salience values of the actual entities. In Table 3.3, where the non-entity words have been removed from the representation, it can be observed that the entities have salience values that are similar to their salience in the text. Thus, only words that are entities are considered for representation within the entity coherence model.

The text "Steve likes Mary. But he doesn't think she is attracted to him." creates a difficulty in trying to represent the second sentence as a vector. The difficulty is that 'Steve' is referenced twice in sentence 2 (realized as 'he' and 'him'), yet the vector model can only capture a single instance of each entity. To overcome this problem, the less salient mention is ignored, and only the most salient realization is taken into account, as in Table 3.4. Barzilay and Lapata [2008] also mention the issue of multiple realizations within an utterance and handle the problem identically.

Vectors have typically been used as a representation of words [Deerwester et al., 1990]. Other applications in computational linguistics that have employed vector representations

	Words In Sentence	
	Steve	Mary
S1	2	1
S2	2	1

Table 3.4: Vector representations of the text "Steve likes Mary. But he doesn't think she is attracted to him." illustrating how repeated entities are handled. 'Steve' is repeated in the second utterance through the coreferring words 'he' and 'him'.

are: word sense discrimination [Schutze, 1998] and ranking [McCarthy et al., 2004]; text segmentation [Choi et al., 2001]; contextual spelling correction [Jones and Martin, 1997]; automatic thesaurus extraction [Gref, 1994, Lin, 1998]; and notably, information retrieval [Salton et al., 1975]. Mitchell and Lapata note that "vector-based models are typically directed at representing words in isolation and methods for constructing representations for phrases or sentences have received little attention in the literature" [2008, pg. 236]. The model proposed here formulates a representation of utterances and their entities using vectors. The vector representation is a straightforward way to track the salience and realization of all the entities in a text, and how their values change over the text. The main benefit of a vector representation is that the mathematical structure is well-studied and common. Vectors and their operations are used in a variety of disciplines that create simulations and models of real-world processes.

3.1.1   Tracking Salience Change Between Sentences

Utterances will be represented using a vector, where the elements of the vector correspond to entities in the text. Since the objective is to compare salience transitions between utterance pairs, we thus compute similarity between utterance vectors by comparing entity salience values. The premise is that sentences having common entities with near-identical salience features are sentences that should be considered highly similar. Hence, the structure of the vector (which entries receive non-zero values, and the actual values themselves) is intended to mimic the structure of the utterance (which entities are realized, and the salience value associated with the entity).

For simplicity, the example within this section will consider the sentence to be the unit of utterance. All aspects of the following example can be translated to other units considered

as candidates for the utterance unit. The sample text will consist of the following five sentences:

$$
\begin{aligned}
&\text{Cinderella was a student in computer science.}_{S_1} \\
&\text{Russia is a large country.}_{S_2} \\
&\text{Cinderella went to a private school in New York.}_{S_3} \quad\quad (3.1) \\
&\text{Moscow is the capital of Russia.}_{S_4} \\
&\text{She likes computer science.}_{S_5}
\end{aligned}
$$

The vector representation for each sentence is located below. The values in the vector represent the salience of each entity within the sentence. For simplicity in these examples, salience is computed by linear order without taking grammatical role into account. Thus, the entity mentioned first is considered to be the most salient. The entity mentioned last in the sentence is considered least salient. The order in which the entities appear determines their salience values. Entities that are not mentioned (unrealized) have zero salience.

$$
\text{Sent} = \begin{pmatrix} \text{Cinderella} \\ \text{student} \\ \text{compsci} \\ \text{Russia} \\ \text{country} \\ \text{school} \\ \text{New York} \\ \text{Moscow} \\ \text{capital} \end{pmatrix}, \; S_1 = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \; S_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 3 \\ 2 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \; S_3 = \begin{pmatrix} 3 \\ 0 \\ 0 \\ 0 \\ 0 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \; S_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 3 \\ 2 \end{pmatrix}, \; S_5 = \begin{pmatrix} 3 \\ 0 \\ 2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}
$$

The above illustrates how sentences are represented as vectors. If an entity is not mentioned in a sentence, then it receives a zero value. If it is realized in a sentence, then it receives a non-zero value. If the entity in that sentence is the most salient/prominent, then it receives the highest value, with less salient entities receiving a lesser value. Our use of the vector can either be as a column or as a row, depending on which is more suitable for illustrative purposes.

3.1.2   Realization

When an entity is mentioned within an utterance, the entity is said to be **realized** in the utterance[1]. If the entity is not mentioned in the utterance, then it is referred to as *not realized*, or *unrealized*. There are three different types of realizations: *direct realization*, *indirect realization*, and *empty realization*.

Poesio et al. [2004a] test an instantiation of realization which allows for **lexical bridging** references (specifically, they use the term **'associative reference'**). They define a realization that allows lexical bridging as an **indirect realization**. The example used by Poesio et al. to illustrate lexical bridging can be seen in the two utterances "John walked to the house. The door was open." The entities that are directly realized are *John*, *house*, and *door*. Lexical bridging occurs due to the associative reference created by the mereological relationship between *house* and *door*. Hence, *house* is indirectly realized in the second sentence due to bridge provided by *door*. Lexical bridging is subsumed under lexical cohesion, which we will not explore further in this thesis. See Poesio et al. [2004c] and Fais [2004] for further work on resolving bridging references.

A **direct realization** does not include lexical bridging. A direct realization consists of the case where the entity is explicitly mentioned by name (or through a pronominal reference). A direct realization does not involve synonyms or mereological associations. Thus, '8888 University Dr.' and 'Simon Fraser University' would be treated as two distinct entities, even though they possibly refer to the same thing in certain contexts.

Traces, known as **empty realizations** [Poesio et al., 2004a], are also not considered a part of directly realized entities. Empty realizations occur in reduced relative clauses and coordinated VPs where an explicit realization has been made implied. In this case, the instantiation of utterance one decides to use would determine whether or not empty realizations are an issue for the coherence model. If the unit of utterance is a sentence, empty realizations do not cause much of an issue compared to the case where the unit of utterance were the clause. The impact on utterance granularity and empty realizations was not part of this study. Empty realizations can be classified under grammatical coherence, rather than entity coherence or lexical coherence.

Since our implementation uses coreference resolvers as a front-end (OpenNLP and Ng

---

[1]When an entity appears multiple times in an utterance (e.g., 'John likes himself.'), it's most prominent realization is tracked, with all other realizations ignored.

and Cardie [2002]), we are not being strict in our entity definition, but taking the output of the resolver as our instantiation of what defines entities.

3.1.3 Salience

Salience describes a way to arrange entities realized in an utterance so that they can be ranked according to their topicality or prominence at any given point in the text. Ranking allows one to distinguish between entities that are most *important* or *topical* in an utterance. There are two types of **entity salience** ranking that will be considered here: a salience based on grammatical order, and a salience based on linear order. Grammatical salience refers to ranking entities according to their grammatical role. The entities here take on the form of subject (S), object (O), other (X), or unrealized (-). Salience based on linear order ranks entities according to the order they are mentioned within an utterance. In a salience based on linear-order, entities mentioned first receive higher salience than entities mentioned later in the utterance. The linear order thus captures preposed subordinate clauses (when they occur) in a very basic way. As Miltsakaki says, it may be that "discourse entities in adjunct subordinate clauses are assigned lower salience than main clauses entities, especially subjects" [2003][pg. vi].

Miltsakaki and Kukich also discuss how complex noun phrases (NPs) and possessive relationships are handled with linear order [2004, pg. 36]:

> In the case of complex NPs, which have the property of evoking multiple discourse entities (e.g., his mother, software industry), the working hypothesis commonly assumed [Walker & Prince, 1996] is ordering from left to right. With respect to complex NPs containing possession relationships the following clarification is in order. English has two types of possessive constructions. The first construction is the genitive construction realized with an apostrophe plus the letter *s* at the end of the noun. In this construction, the possessor is to the left of the possessee, for example *Mary's father*. The second construction contains the preposition *of*. In this case, the possessor is to the right of the possessee. To maintain uniformity for the ranking of the complex NP, we assume linearization of the complex NP according to the genitive construction and then rank from left to right. In (b), for example, TLP ranks higher than both success and the secret. The ranking is easy to see if we linearize The secret of TLP's success to

> TLP's success's secret:
>
> a. Trade Leisure Publications is a successful publishing house in Russia, with two market-leading monthly consumer magazines.
>
> b. The secret of TLP's success, however, is not based on developing or exploiting some new technology or business strategy.
>
> c. Rather, TLP follows a business strategy that has been known since business began.

In the case of Miltsakaki and Kukich's example (b), we currently have no way of rearranging the NPs, and thus follow a strict left to right ordering for the salience of the complex NP. The final ranking in our model for *'The secret of TLP's success'* would be *secret*, *TLP*, and finally with lowest salience, *success*.

Miltsakaki proposes an augmented entity ranking, one which distinguishes between entities appearing in the main clause from those entities appearing in subordinates [2004, pg. 33]

> To construct the ranking of the C_f list under the assumption that the 'utterance' contains both a main clause and its subordinate clauses, we assume the augmented C_f ranking rule proposed in [Miltsakaki, 2002]. The 'M' prefix stands for main clause and the 'S_n' prefix stands for the nth subordinate clause.

The following is her proposed augmented ranking rule that ranks entities less according to whether they appear in the main or subordinate clause [Miltsakaki & Kukich, 2004, pg. 34]:

Augmented ranking rule M-Subject > M-indirect object > M-direct object > M-other > M-QIS,Pro-ARB > S1-subject > S1-indirect object > S1-direct object > S1-other > S1-QIS,Pro-ARB > S2-subject > ...

We extend Miltsakaki's notion of distinguishing between realized main and subordinate entities, while also extending the context window beyond adjacent utterances through the notion of recency, as described fully in Section 3.3.1.

Alternative rankings for entity salience could be based on models such as the Givenness Hierarchy [Gundel et al., 1993], which would form the basis for different additional weighting schemes based on the form an entity is realized. Kameyama [1998] defines the following salience ordering, which abbreviates the Givenness Hierarchy:

Zero Pronominal > Pronoun > Definite NP > Indefinite NP

These rankings are intended for referring expressions, used in resolving coreferent entities. A modification could be that this hierarchy would provide an additional feature to the computation of salience. For example, if an object is realized as a pronoun, then it could receive a higher (or equivalent) salience with respect to the subject of the sentence (if the subject was not realized as a referring expression). Hence, referring expressions could be used as an additional feature in the computation of entity salience. This study did not involve combining the Givenness Hierarchy (or any other referential hierarchy ordering) with the grammatical and linear order. It is left for future work to study grammatical order combined with the pronominalized form of entities.

3.1.4 Utterance Transition Measures

The utterance transition types mentioned previously in Section 2.3.2, and explained fully in Section 3.2, are intended to quantify the degree of coherence between two utterances through the tracking of how the salience of the utterance's entities change. Centering Theory achieves this by tracking a single entity (i.e., the most salient entity) between utterances. The vector model proposed in this thesis allows one to track all the entities in an utterance. The problem with tracking all the entities thus becomes how to measure the utterance transitions similarly to the transition types of Centering Theory. We first describe an entity coherence measure, then describe the transitions of Centering Theory, and finally show how our model captures and extends Centering Theory's coarse transitions.

The utterance transition metric ranks and orders the different types of entity transitions based on certain types of transitions being preferred over others. The idea is that an entity transitioning from the role of not being realized in one utterance to the role of 'subject' in the next utterance is viewed as being less desirable as an entity moving from 'indirect object' to 'subject'. The "jump" in grammatical role in the former is much too abrupt compared to the latter. Smaller changes in grammatical role are given preference over larger ones. The overall objective of the sentence transition measure is to prefer sentence pairs that are "about" (*realize*) the same entities and that assign their realized entities a similar salience (e.g., entities maintain similar grammatical role).

Computing transitions between utterance pairs first involves computing the individual transitions between entities, and then averaging these individual transitions to a value that

is representative of the overall entity transitions that occur between the utterances. Hence, the utterance transition value is directly determined by the individual transitions of its realized entities.

Formula 3.2 computes the transition value of an individual entity (e_i) between two adjacent utterances $U1$ and $U2$, as follows:

$$\text{difference}(e_{iU1}, e_{iU2}) \cdot \frac{avg(e_{iU1}, e_{iU2})}{\text{Salience}_{max}} \tag{3.2}$$

Equation 3.2 ranks and weights the different kinds of entity transitions. The function difference(x,y) corresponds to

$$\text{difference}(e_{iU1}, e_{iU2}) = |e_{iU1} - e_{iU2}| \tag{3.3}$$

The function avg(x,y) corresponds to:

$$avg(e_{iU1}, e_{iU2}) = |e_{iU1} + e_{iU2}|/2 \tag{3.4}$$

Salience$_{max}$ refers to the maximum salience value an entity can achieve within any utterance in a text. Salience$_{max}$ is determined by the utterance that realizes the most number of entities and using that value as the ceiling for salience. Hence, if a sentence with the most number of entities contains 10 entities, then Salience$_{max} = 10$.

Formula 3.2 is next normalized by the normalization term given in Formula 3.5:

$$\text{Salience}_{max}/2 \tag{3.5}$$

Formula 3.5 is a normalizing factor, so that the entity transition of Equation 3.2 takes on a value in the range between 0 to 1, with 0 being a preferred transition and 1 being not preferred. A 0 value represents that there is no change in salience for an entity between adjacent utterances. A 1 value corresponds to a maximally abrupt salience change, i.e., an entity that is unrealized in one utterance and is then realized in the other utterance. The different transitions that fall within those extremes would thus have an entity transition value that lies somewhere between 0 and 1.

Equation 3.5 is identical to Equation 3.2 with the difference being that $avg(e_{iU1}, e_{iU2})$ is replaced with avg(Salience$_{max}$, 0), which becomes Salience$_{max}/2$, and difference(e_{iU1}, e_{iU2}) is replaced by difference(Salience$_{max}$, 0).

The final normalized expression, Equation 3.6, divides Formula 3.2 by Formula 3.5 to get:

$$\text{Entity Transition Value}(e_i) = 2 \cdot \text{difference}(e_{iU1}, e_{iU2}) \cdot \frac{avg(e_{iU1}, e_{iU2})}{\text{Salience}_{max}^2} \tag{3.6}$$

For example, if an entity in utterance U_n has a salience value of 4, and its salience in the previous utterance U_{n-1} was 3, this would represent an Entity Transition Value of .0625 (assuming $Salience_{max}$ of 4). For an entity salience transition from 1 to 4, this would represent an Entity Transition Value of .5625. Thus, the transition from salience of 3 to 4 (.0625) is preferred over the transition from 1 to 4 (.5625) since it involves a smaller 'jump' in transition.

The purpose behind the entity transitions being weighted as they are in Equation 3.6 is to prefer utterances which have their most salient entities in common over utterances that have their lowest salient entities in common. The idea is to add a 'mass', similar to the concept of *inertia*, to each entity according to its salience (e.g., grammatical role). To illustrate with an example of inertia, a large boulder would be more difficult to push than a small pebble over the same distance. Similarly, an entity having a grammatical role of *subject* in a main clause should be considered unfavorable to transition to an unrealized position in comparison to an entity that is in the grammatical role of direct object in a subordinate clause that transitions to unrealized. This is the main basis behind the way transition values are computed for individual entities.

The construction of the sentence transition is similar to the individual entity transition given above. The sentence transition value takes into account the number of unrealized entities common to both sentences by not including them in the individual entity computation. If they were to be included, they would artificially inflate the preferred "***no change***" transition (i.e., an entity transitioning from no salience (unrealized in the previous utterance) to no salience (unrealized in the current utterance)). Since there are typically many entities in a text, and relatively few entities realized in a sentence, then it would be safe to assume the majority of entity transitions would be of the type "no change". Thus, if the entity 'house' is unrealized in two adjacent utterances, then the utterance transition function does not reward them for being similar in what they do not realize. This provides the basis for the *transition weight*.

The above captures transitions amongst entities, ranking the different types of transitions to being either preferred or not preferred. A similar type of ranking will be used to incorporate the number of entities in common between utterances. Most utterances realize very few entities (with their corresponding vectors being sparse), resulting in many entities having the preferred *no change* transition. To avoid favoring such utterance transitions, utterances with many entities in common are weighted over utterances with few entities

in common. Equation 3.7 calculates the transition weight (W_{trans}) for favoring utterance transitions having more entities in common. The numerator is the sum of the number of entities in U_i that are realized in U_j with the number of entities in U_j that are realized in U_i. This is designed to capture the intuition that two utterances that realize the same number of entities should be considered more similar.

$$W_{\text{trans}} = \frac{\text{\# of entities in } U_i \text{ realized in } U_j}{\text{\# of distinct entities realized in } U_i \text{ and } U_j} \tag{3.7}$$

The final transition metric is found in Equation 3.8, which computes the average transition between all realized entities multiplied by W_{trans}, which is the transition weight that corrects for the sparsity of the vector in the number of realized entities in an utterance.

$$\text{Transition}(U_i, U_j) = 1 - W_{\text{trans}} \cdot (1 - \frac{\sum_{k=1}^{n} \text{Entity Transition Value}(e_k)}{\text{total \# of distinct entities realized in } U_i \text{ and } U_j}) \tag{3.8}$$

The function $\text{Transition}(U_i, U_j)$ will be the transition metric[2] used to compute all subsequent transitions between all utterances used in the rest of this thesis. The next section will outline how the common transitions from Centering are captured using the proposed vector representation.

Using the example text *"Steve likes Mary. Mary is a Republican."*, we will illustrate how to compute the transition from sentence 1 to sentence 2. Sentence 1 has the vector representation $(2, 1, 0)$, and sentence 2 has the representation $(0, 2, 1)$, where the elements of the vectors correspond to (*Steve, Mary, Republican*). $\text{Salience}_{\text{max}} = 2$ since there are 2 entities in the sentence with the maximum number of entities (in this case, both sentences have the same number of entities). Next, we compute the Entity Transition Value for each individual entity using Equation 3.6. Thus:

$$\text{Entity Transition Value}(e_1) = 2 \cdot \text{diff}(2, 0) \cdot \frac{avg(2, 0)}{2^2} = 2 \cdot 2 \cdot \frac{1}{4} = 1 \tag{3.9}$$

$$\text{Entity Transition Value}(e_2) = 2 \cdot \text{diff}(1, 2) \cdot \frac{avg(1, 2)}{2^2} = 2 \cdot 1 \cdot \frac{1.5}{4} = 0.75 \tag{3.10}$$

$$\text{Entity Transition Value}(e_3) = 2 \cdot \text{diff}(0, 2) \cdot \frac{avg(0, 2)}{2^2} = 2 \cdot 2 \cdot \frac{1}{4} = 1 \tag{3.11}$$

---

[2]Note that the Transition function is actually a pseudo-metric since there can be two distinct sentences that can have their entities represented using an identical vector. For example, *"John took Mary to the store"* and *"John saw Mary at the store"* both have the same vector representation.

From these individual entity transitions, we can see that when an entity that is realized in one sentence is not realized in the next, then it has a high transition cost (e.g., entities e_1 and e_3). We next compute the transition weight using Equation 3.6:

$$W_{\text{trans}} = \frac{1}{3} = 0.\overline{3} \qquad (3.12)$$

Next we compute the overall transition between sentences 1 and 2. The overall transition value is:

$$\text{Transition}(S_1, S_2) = 1 - 0.\overline{3} \cdot \left[1 - \frac{1 + 0.75 + 1}{3} \right] = 1 - 0.\overline{3} \cdot (0.91\overline{6}) = 0.695 \qquad (3.13)$$

Thus, this value is closer to 1 (which represents an incoherent entity transition) than it is to 0 (which represents a coherent entity transition). It is less coherent when compared to a text such as *"Steve likes Mary. He will ask her out."*, which has an overall transition value of 0 (and is thus maximally coherent in terms of its entity transitions).

Another benefit of using a fine-grained model of utterance transitions is that the model is less sensitive to such factors as having to determine an exact ranking, and it is less sensitive to whichever definition of utterance one wishes to use. The ranking and utterance definition (when used with recency, as in Section 3.3.1) are not as critical in this model. For example, in the case of conjoined subjects, such as in the sentence *'Steve and Mary went to the park'*, the model can use either grammatical role, linear order, or a combination of the two in order to rank *Steve* and *Mary* without a sharp difference in its transition from one utterance to the next, no matter what the next utterance may be (possible next utterances could be *'Steve played tennis'*, *'Mary played tennis'*, or *'They played tennis'*).

3.2 Entity Transitions From Centering Theory

Centering Theory proposes a few different ways of capturing entity coherence between utterances. This section will first describe the transitions, and then explain how they are captured using an utterance vector representation. The common Centering transitions are described as such:

Continue - this occurs when an entity has been the preferred center for three consecutive utterances (including the current utterance). $C_b(U_i) = C_b(U_{i-1})$ and $C_b(U_i) = C_p(U_i)$

An example of a sequence of utterances exhibiting a *continue* transition is the following:

$$\text{John went to the store.} \tag{3.14}$$

$$(C_f = \{\text{John, store}\}, C_b = \text{none}, C_p = \text{John})$$

$$\text{He needed to buy wine.} \tag{3.15}$$

$$(C_f = \{\text{John, wine}\}, C_b = \text{John}, C_p = \text{John})$$

$$\text{He was preparing a special dinner.} \tag{3.16}$$

$$(C_f = \{\text{John, dinner}\}, C_b = \text{John}, C_p = \text{John})$$

Retain - occurs when there is a new preferred center yet the previous preferred center occurs in the current utterance. $C_b(U_i) = C_b(U_{i-1})$ and $C_b(U_i) \neq C_p(U_i)$

An example of a sequence of utterances exhibiting a *retain* transition is the following:

$$\text{John went to the store.} \tag{3.17}$$

$$(C_f = \{\text{John, store}\}, C_b = \text{none}, C_p = \text{John})$$

$$\text{He needed to buy wine.} \tag{3.18}$$

$$(C_f = \{\text{John, wine}\}, C_b = \text{John}, C_p = \text{John})$$

$$\text{Mary was preparing a special dinner for him.} \tag{3.19}$$

$$(C_f = \{\text{Mary, dinner, John}\}, C_b = \text{John}, C_p = \text{Mary})$$

Smooth-shift - occurs when the preferred center of the current utterance was the highest ranked entity in the previous utterance that is realized in the current utterance the backward-looking centers of the previous two utterances are not the same entity. $C_b(U_i) \neq C_b(U_{i-1})$ and $C_b(U_i) = C_p(U_i)$

An example of a sequence of utterances exhibiting a *smooth-shift* transition is the

following:

$$\text{The store didn't have what John was looking for.} \qquad (3.20)$$
$$(C_f = \{\text{store, John}\},\ C_b = \text{none},\ C_p = \text{John})$$
$$\text{He needed wine from this store for later tonight.} \qquad (3.21)$$
$$(C_f = \{\text{John, wine, store, tonight}\},\ C_b = \text{store},\ C_p = \text{John})$$
$$\text{John was preparing a special dinner.} \qquad (3.22)$$
$$(C_f = \{\text{John, dinner}\},\ C_b = \text{John},\ C_p = \text{John})$$

Rough-shift - the preferred center was not the highest ranked entity in the previous utterance that also appears in the current utterance, as well as not having the same backward-looking centers. $C_b(U_i) \neq C_b(U_{i-1})$ and $C_b(U_i) \neq C_p(U_i)$

NOCB - occurs when there are no entities in common between the two utterances. This measure is unique in the sense that it only looks at the current and previous utterance since it does not incorporate the use of $C_b(U_{n-1})$. [Karamanis and Manurung, 2002, Althaus et al., 2004]

Cheapness - occurs when the most salient entity of the current utterance is also the most salient entity of the previous utterance. $C_b(U_n) = C_p(U_{n-1})$ [Strube and Hahn, 1999]

Expensive - occurs when the most salient entity of the current utterance is not the most salient entity of the previous utterance. $C_b(U_n) \neq C_p(U_{n-1})$ [Strube and Hahn, 1999]

Establishment - occurs when a center is established after a transition where there is NOCB, e.g., occurs at the beginning of the text since there are no previous utterances that can be used to compute a transition. $C_b(U_{n-1}) = \emptyset$ and $C_b(U_n) \neq \emptyset$ [Kameyama, 1986]

ZERO - intended to be the inverse of Establishment. Occurs when there is a backward-looking center in the previous utterance, but the current utterance has a NOCB transition. $C_b(U_{n-1}) \neq \emptyset$ and $C_b(U_n) = \emptyset$ [Poesio et al., 2004a]

NULL - this type of transition occurs when there are two consecutive NOCB transitions. $C_b(U_{n-1}) = \emptyset$ and $C_b(U_n) = \emptyset$ [Poesio et al., 2004a]

	$C_b(U_n) = C_b(U_{n-1})$ or $C_b(U_n) = \emptyset$	$C_b(U_n) \neq C_b(U_{n-1})$
$C_b(U_n) = C_p(U_n)$	Continue	Smooth-shift
$C_b(U_n) \neq C_p(U_n)$	Retain	Rough-shift

Table 3.5: Table of Standard Centering Theory transitions.

where $C_b(U_i)$ is the highest ranked entity in utterance U_{i-1} that is realized in U_i and $C_p(U_i)$ is the most highly ranked entity in U_i. A table of the traditional Centering Theory transitions is shown in Table 3.5.

The difficulty involved in attempting to capture the above transitions with our vector representation relates to how the backward-looking center (C_b) is determined using two utterances, and not only a single utterance as in our model. The transition that occurs between two consecutive C_b's actually needs to take into account three utterances since the computation of $C_b(U_{n-1})$ relies on U_{n-2} which is two utterances prior. Thus there is an added context whenever the computation of $C_b(U_{n-1})$ is involved in a transition. The original entity coherence measures proposed by Centering Theory (*continue, retain, shift*), along with establishment, ZERO, and NULL, are *not* actually a measure of utterance-to-utterance transitions[3]; Centering Theory actually measures *sequences* of transitions, specifically adjacent transition pairs. The difficulty thus lies in trying to capture transitions sequences with a single transition. Instead, even though we will define Centering Theory's transitions within the vector model, the main focus is to define the spectrum of values that single transitions can take on and classify them to their Centering Theory counterparts (transitions such as *continue, retain, shift*).

The following subsections illustrate how the vector model captures the transitions commonly associated with Centering Theory. $C_f(U_n)$ is used to denote a list of all the entities of utterance U_n (essentially a shorthand notation for the sentence vector associated with sentence U_n). To help illustrate the following mappings of the vector model and Centering

---

[3]The extended Centering Theory transitions that were not part of the original formulation (*NOCB, cheapness, expensive*) *are* actual utterance-to-utterance transition measures, and are therefore trivial to capture within the vector model.

Theory, we will use the following example text:

$$U_1: \qquad \text{Hans likes Mary.} \qquad\qquad (3.23)$$

$$U_2: \qquad \text{Mary is a Republican.} \qquad\qquad (3.24)$$

$$U_3: \qquad \text{He will still ask her out.} \qquad\qquad (3.25)$$

The above utterances have the following forward-looking centers lists:

$$C_f(U_1) \quad = \quad \{Hans, Mary\} \qquad\qquad (3.26)$$

$$C_f(U_2) \quad = \quad \{Mary, Republican\} \qquad\qquad (3.27)$$

$$C_f(U_3) \quad = \quad \{Hans, Mary\} \qquad\qquad (3.28)$$

The following preferred centers for each utterance are as follows:

$$C_p(U_1) \quad = \quad Hans \qquad\qquad (3.29)$$

$$C_p(U_2) \quad = \quad Mary \qquad\qquad (3.30)$$

$$C_p(U_3) \quad = \quad Hans \qquad\qquad (3.31)$$

Hence, we have $C_b(U_2)$ = Mary and $C_b(U_3)$ = Mary. Thus, the type of transition we have for this text according to Centering Theory, with $C_b(U_3) \neq C_p(U_3)$ and $C_b(U_3) = C_b(U_2)$ is a Retain transition.

The main difficulty in the following mappings is that the traditional Centering transitions due not map to a single value in our coherence model, but to a range of values. Hence they can be considered as *coarse* measures. Thus, the mappings are intended to loosely specify in which range Centering's transitions fall under.

3.2.1 Continue

Since the computation of the Continue requires the context of three utterances, it is thus equivalent to summing the three sentence vectors and observing if one of the elements in the vector sum equates to $3 \cdot \text{Salience}_{max}$. This is equivalent to an entity being the most salient for three consecutive utterances, corresponding to an entity that is ranked as C_p for three consecutive utterances.

3.2.2 Retain

Again, three utterances are required. The entity salience transitions from Salience_{max} in U_{n-2} to Salience_{max} in U_{n-1} to x in U_n, where $0 < x < \text{Salience}_{max}$. This corresponds to an entity being ranked C_p for the first two utterances, and then being realized (but not as the C_p) in the final utterance. This mapping illustrates how the Retain transition can take on a wide range of values due to the only condition being that it be realized in the current utterance without specifying exactly where it is realized with respect to the other entities.

3.2.3 Smooth-shift

Three utterances are required to compute a smooth-shift. If the most salient entity (Salience_{max}) of U_{n-1} retains being the most salient, and that entity does not have a sum equal to $3 \cdot \text{Salience}_{max}$ over the three utterances. Again, this covers a wide range of possibilities for the entity to occupy in the salience list. In a model that incorporates all entities in its transition measure, it could end up scoring more coherently than a Retain transition (which is considered more favourable than a smooth-shift in the Centering literature).

3.2.4 Rough-shift

Three utterances again are required. If the most salient entity (Salience_{max}) of U_{n-1} is not the most salient entity of the current utterance, U_n. Similar to smooth-shift, but the most salient entity of the current utterance is not realized in the previous. Formally:

$\text{C}_p(\text{U}_{n-2}) \in \text{C}_f(\text{U}_{n-1})$ and $\text{C}_p(\text{U}_{n-2}) \notin \text{C}_f(\text{U}_n)$

Again, this corresponds to a wide range of values in a model that incorporates all entities.

3.2.5 NOCB

Simply, this occurs when $\text{C}_f(U_n) \cap \text{C}_f(U_{n-1}) = \emptyset$. This means the two utterances have no entities in common.

3.2.6 Cheapness

This occurs when the most salient element in the previous utterance is realized in the current utterance. This occurs in our model whenever an entity transitions from having maximum salience (Salience_{max}) to any salience value other than 0.

3.2.7 Expensive

This occurs when the most salient element in the previous utterance is not realized in the current utterance. This occurs in our model when an entity transitions in salience with a value of $-\text{Salience}_{max}$. This corresponds to the case where an entity transitions from most salient to least salient.

3.2.8 Establishment

Three utterances again are required. This is a NOCB transition between U_{n-1} and U_{n-2}, with $C_f(U_n) \cap C_f(U_{n-1}) \neq \emptyset$.

3.2.9 ZERO

Three utterances again are required. This is a NOCB transition between U_{n-1} and U_n, with $C_f(U_{n-2}) \cap C_f(U_{n-1}) \neq \emptyset$.

3.2.10 NULL

Three utterances again are required. This occurs when we have two consecutive NOCBs occurring. Thus, $C_f(U_n) \cap C_f(U_{n-1}) = \emptyset$ and $C_f(U_{n-1}) \cap C_f(U_{n-2}) = \emptyset$.

3.3 Extended Centering Theory

This section will outline the range of transitions in a vector model, using the transitions associated with Centering Theory as a reference point for the discussion. The main difference between the two notions of *transition* will be that the vector model only requires two utterances to compute a transition, whereas Centering transitions (Continue, Retain, and Shifts) require three utterances due to the computation of the backward-looking center of the previous utterance. Using only two utterances, the purpose here will be to describe the values associated with the vector transitions in a way that can describe Centering Theory's transitions.

The main intent of this section is to show that the utterance vector model captures the transitions of Centering Theory, while at the same time extending the coarse transitions proposed by Centering Theory by allowing the capture of all transitions beyond the four standard ones.

3.3.1   Recency

The purpose of **recency** is twofold: to minimize the compromises associated with having to define utterances exclusively as either (complex) sentence or clause, and to capture the stack/focus model [Grosz & Sidner, 1986]. The intuition behind the concept of recency is that the *"most recent"* utterance mentioned is more salient than an utterance mentioned some utterances before. Miltsakaki uses the term *referential distance*. She describes referential distance as something that "is counted with respect to the number of clauses intervening between the current reference to an entity and the most recent reference to the same entity in the previous discourse" [2003, pg. 26]. Hence, the utterance most recently mentioned is considered more *accessible* than utterances mentioned prior. Recency also allows one to weigh main clauses with greater import than subordinate clauses.

Recency refers to the salience of a previous utterance with respect to the current utterance. The intuition is that the immediately preceding utterance is more "recent" (in comparison to utterances further back in the discourse) in terms of its focus and relevance to the current utterance. Hence, the entities of utterance U_{n-1} are higher on the current stack or focus space (e.g., the current utterance U_n is the current focus) than are the entities of utterance U_{n-2}.

Recency allows for distinguishing between the two different types of sentences, compound and complex, and provides a way for handling each one uniquely. A **compound** sentence is the *conjunction* of two main clauses that are *coordinated* by *coordinators* such as *but, and, or*, etc. In compound sentences, the two (or more) main clauses are treated as having equal status and prominence within a sentence (i.e., one clause is not subordinate to the other). An example of a compound sentence is *'I went to the store and I bought a new shirt'*. The relationship between the two clauses being on the same status in salience is called **parataxis**. A **complex** sentence has one main clause and one (or more) subordinate clause. In a complex relationship, the main clause has higher salience than its subordinate, and the relationship is referred to as **hypotaxis**. Subordinate clauses need not be marked with explicit subordinators such as *although, because, while, when, which*, etc. An example of a complex sentence is *'I went to the store because I needed a new shirt'*. Kameyama fleshes out the different types of clausal and sentential relationships that need to be considered [Kameyama, 1998, pg. 12]:

A number of dimensions need to be considered for a full account: nominal versus

> clausal units, matrix versus subordinate clauses, the linear ordering of clauses, tensed versus untensed clauses, conjuncts versus adjuncts, adjuncts versus complements, direct versus indirect reports, and restrictive versus nonrestrictive relative clauses. For each dimension, we would like to know when and how the center is updated.

This thesis addresses such issues through the use of the vector representation (to determine how centers are updated) and the notion of recency (to account for the various clause and sentence types)[4].

Taboada & Hadic Zabala [2008] and Poesio et al. [2004a] consider different units of analysis within Centering Theory. The entity vector model alleviates the problems associated with choosing the granularity of the utterance. Because of recency's weighting of a sentence's constituents, the choice of what is an utterance is not as critical. In an experimental evaluation of the type of utterance transitions in a corpus, Poesio et al. found that the most common transition was one where two adjacent utterances had no realized entities in common [Poesio et al., 2004a, pg. 328].

> If the vanilla instantiation were the "right" way of setting the parameters, we would have to conclude that in the genres contained in our corpus, utterances are very likely to have a unique C_b, but entity coherence does not play a major role in ensuring that a text is coherent: only 35.4% of utterances in our corpus would be "entity-coherent", that is, would contain an explicit mention to an entity realized in the previous finite clause.

One of the possible reasons for so many NOCBs is that the transitions can be very abrupt at the local level (between utterances), even though the text is coherent. What recency allows, in combination with tracking multiple entities, is a type of *smoothing* over an expanded window of utterances. Thus, one of the questions that results from Poesio et al.'s findings is what effect extending the **utterance window** would have on the NOCB transitions. The *utterance window* is the number of previous utterances that are to be included in the computation of coherence (Centering Theory uses an utterance window of three, where the two previous utterances are used to compute the coherence of the current utterance

---

[4]The one aspect we do not explore is weighing the different types of clauses, e.g., what salience weighting should an adjunct receive versus a relative clause. Primarily due to there not existing a clause segmenter that identifies the clause type, such a study is left for future work.

through an entity transition measure). Since the text is known to be coherent, is using only the immediate predecessor a restriction in assessing coherence quality in a text? Would the number of NOCBs be reduced in coherent texts if the number of previous utterances were increased in determining coherence? What is the "magic number" in terms of how many utterances should be included in the utterance window? We speculate that the "magic number" for the size of the utterance window cannot be determined, let alone by entity coherence alone. Other types of coherence, such as relational coherence, would be critical in determining how large the window is. For example, if there is an *elaboration* relationship between one or more utterances in a text, then one can choose the context window to consist of solely those utterances that are subsumed under that common relation. Similarly, the process can work in reverse. If it is found that there exists a contiguous chain (lexical chain) of entities that exists for a short span in the text, then the discourse span becomes a possible candidate for having an elaboration relationship that elaborates upon the topic which forms the lexical chain. The size of the utterance window (and its relationship to relational coherence) is not explored further in this thesis.

In Miltsakaki and Kukich's previous work [2004, pg. 31], they argue that:

> ...for the purposes of discourse organization and textual coherence, subordinate clauses are not processed as independent processing units. ...treating subordinate clauses as independent units ('utterances') yields counter-intuitive topic transitions. This can be seen for English in the constructed example:
> Sequence: Main-subordinate-main

> a. John had a terrible headache.
> C_b= ?, C_f= John > headache, Transition=none
> b. When the meeting was over,
> C_b=none, C_f = meeting, Transition=Rough-Shift
> c. he rushed to the pharmacy store.
> C_b=none, C_f=John, Transition=Rough-Shift

> Allowing the subordinate clause to function as a single update unit yields a sequence of two Rough-Shifts, which is diagnostic of a highly discontinuous discourse. Further, if indeed there are two Rough-Shift transitions in this discourse the use of the pronominal in the third unit is puzzling. A sequence

of two Rough-Shift transitions in this short discourse is counterintuitive and unexpected given that of all Centering transitions, Rough-Shifts in particular have been shown to (a) disfavor pronominal reference [Walker et al., 1994, Di Eugenio, 1998, Miltsakaki, 1999] (among others), and (b) be rare in corpora, to the extent that the transition has been ignored by some researchers [Di Eugenio, 1998, Hurewitz, 1998] (among others).

By subordinate clause, Miltsakaki describes it as such [2004, pg. 32]:

...the "utterance", a single Centering update unit, consists of one main clause and all its associated tensed dependent clauses, including sentential complements of verbs, relative clauses and subordinate clauses. In Miltsakaki [2002], the term 'subordinate clause' is used to describe tensed, adverbial clauses introduced by a subordinate conjunction (e.g., *when, because, as soon as, although*, etc.). To identify subordinate clauses the 'reversibility test' is applied [Quirk et al, 1985]; subordinate clauses can be preposed with respect to the main clause. For example, in (6), *although* is classified as a subordinate conjunction and the although-clause is classified as a subordinate clause because placing the although-clause before the main clause retains grammaticality. Conversely, however in (8) is not classified as a subordinate conjunction because preposing the clause it is associated with yields ungrammaticality.

(6) John traveled by air although he is afraid of flying.

(7) Although he is afraid of flying, John traveled by air.

(8) John traveled by air. However, he is afraid of flying.

(9) ∗ However, he is afraid of flying. John traveled by air.

There is no automatic way to perform the *reversibility test* to determine whether a clause is subordinate. We leave the automatic identification of subordinates for future work, and present the theoretical basis for recency in this section.

Recency allows one to be able to capture the salience of entities at any point in the text with any size of utterance window. This is achieved by observing the current utterance and taking into consideration all previous utterances that were uttered before it. For example, in Equation 3.32, the salience of entities at the nth utterance is determined by looking at the saliences at utterances U_n, U_{n-1}, and U_{n-2}. Here the recent utterances, U_{n-1} and U_{n-2}, are given equal weighting. The sum of each individual entity's salience across the window

of the three utterances determines the *current focus*, which is a snapshot of entity salience at utterance n in the text.

$$\text{Uniform Current Focus} = U_n + U_{n-1} + U_{n-2} \tag{3.32}$$

In Equation 3.33, the previous utterances are weighted to model how previous utterances are less salient than the current utterance due to their distance to the current utterance. Implicit in Equation 3.33 (and also Equation 3.32) is that the weight assigned to all utterances occurring prior to U_{n-2} are given a weight of 0. In 3.33, the weight is simply calculated by dividing by the number of utterances previous to the current utterance. The current utterance is divided by 1, the previous utterance is divided by 2, and the utterance previous to that is divided by 3, and so on. Hence, the most recent utterances have a higher recency weight than ones further back in the utterance window. The general equation (Equation 3.34) includes calculating a recency weight for all previous utterances up until the current utterance.

$$\text{Current Focus with Basic Weights} = 1 \cdot U_n + \frac{1}{2}U_{n-1} + \frac{1}{3}U_{n-2} \tag{3.33}$$

$$\begin{aligned} \text{Current Focus} &= Recency(U_n) \cdot U_n + Recency(U_{n-1}) \cdot U_{n-1} + ... \\ &= \sum_{i=1}^{n} Recency(U_i) \cdot U_i \end{aligned} \tag{3.34}$$

The model proposed here captures recency by associating a weight to all previous utterances. This weight can be proportional to how recent the utterance is to the current utterance. Also, if the definition of utterance is the clause rather than a sentence, then all utterances that originate from the same sentence can be given a similar recency weighting. This would further alleviate having to make a compromise in defining the unit of utterance. Further, if one could identify certain clauses (such as infinitival adjunct clauses) as being satellite to the discourse, one could assign their recency weight a value of 0 (or a relatively smaller non-zero weight with respect to the other utterances within the same sentence), which would minimize the salience of entities in subordinated clauses from the discourse. Hence, the variables to consider have been shifted from a concern over whether the unit of utterance is a clause or a sentence to:

- what the recency weights are for sentences

- what the recency weights are for the different clauses within a sentence (e.g., main clauses would receive a higher weight than subordinate clauses)

The model is intended to be as flexible as possible in capturing whichever definition of utterance that is used, as well as using a complete history of utterances to determine the focus at any point in the text.

An example of using recency to compute coherence is as follows. Assume the following abstracted text:

$$\text{Sentence 1: (Clause 1) (Clause 2) (Clause 3)}$$
$$\text{Sentence 2: (Clause 1) (Clause 2)} \tag{3.35}$$
$$\text{Sentence 3: (Clause 1) (Clause 2) (Clause 3)}$$

In this example, there are three sentences. Sentences 1 and 3 are made up of 3 clauses, whereas sentence 2 has only 2 clauses. Also assume that the main/subordinate relationships for the clauses within each sentence are as follows:

$$\text{Sentence 1: (Main) (Subordinate) (Subordinate)}$$
$$\text{Sentence 2: (Subordinate) (Main)} \tag{3.36}$$
$$\text{Sentence 3: (Main) (Subordinate) (Subordinate)}$$

If the weighting scheme was such that the main clauses would receive a higher weight than the subordinate clauses, and that the most recent sentences would receive more weight, then the following weights would be associated with each clause:

$$\text{Sentence 1: (.333) (.166) (.166)}$$
$$\text{Sentence 2: (.25) (.5)} \tag{3.37}$$
$$\text{Sentence 3: (1) (.5) (.5)}$$

In this simple example, main clauses are given a weight of 1, subordinates are given a weight of 0.5, and sentences are divided by their distance to the current sentence (assume sentence 3 is the current sentence). The values in each clause correspond to the weight which the entities within the clause are multiplied by. Thus according to the final weights, entities in clauses 2 and 3 in sentence 1 are considered less salient than entities in clause 2 of sentence 2. This is so as to capture what Miltsakaki observes about subjects in subordinate clauses [2003, pg. 2]:

> Subject pronouns in main clauses are associated with the highest ranked entity in the previous sentence, the entity that is the topic of that sentence and often the most likely topic of the next sentence... often the main clause subject, with the exception of certain subjects that are marked by the language for low or zero salience in the computation of topic continuity (e.g., impersonal pronouns). Entities evoked in the subordinate clauses of the sentence rank lower than their main clause counterparts, even if they have prominent grammatical roles (e.g., subjects) or are linearly closer to the pronoun.

Hence, even if an entity in a subordinate clause has its grammatical role as the most salient (i.e., subject) or its linear distance to a pronoun is closest, it may not be the topic of the discourse at that point in the text. As well, Arnold [1998, pg. 29] says:

> ... if an entity hasn't been mentioned for awhile, and the representation of that entity has not been re-activated, the activation for that representation will slowly decrease and eventually disappear. On the other hand, the reason that entities become less accessible over time may be the result of interference from other referents which are mentioned in the intervening discourse. These referents may compete with the previous referents and lead to their suppression [Gernsbacher, 1990, pg. 139]. Clark and Sengul [1979] also suggested that recency effects are not linear. They found that referents mentioned in the previous clause were significantly easier to access than referents from two or three clauses back. However, they found no difference between referents from two clauses back and referents from three clauses back. They concluded that referents from the preceding clause enjoy a privileged position, in comparison with other referents.

The above illustrates the interaction involved between multiple entities and their effect of pronoun resolution under competition and suppression. It also shows a justification for a decreased weighting with utterances mentioned further back in the discourse, at least for the case of pronoun resolution through the notion of *activation*. Although Clark and Sengul's conclusion to the number of utterances involved in pronoun resolution is closer to Centering Theory (no more than two utterances back), it cannot be extrapolated that what works for anaphora resolution would translate to a coherence measure, as evidenced by the number of NOCBs found by Poesio et al. in a coherent text. Thus, if a NOCB is to be used as a

metric for the measure of incoherence, the computation of it must be modified so that it better reflects actual incoherence in a text.

Also note in sentence 2, if there are preposed adjuncts then we would like to have the option for them to be considered more salient than adjuncts in postposition [Kameyama, 1998]. This model can handle Kameyama's notion of **raising** the clauses to different levels and embeddings of the various clauses (e.g., "tensed" adjunct clauses may be raised to positions dominating the main clause). Adjustments of whether a clause is "raised" or "lowered" can be captured via the recency weights as mentioned in the example above. Similarly, Renkema [2000] evaluates the **topicality** of a sentence versus the sentence's **linking**. Topicality hypothesizes that the information in the main clause is more salient than the information in the subordinate clause [5]. Linking hypothesizes that the position of the clause, regardless whether main or subordinate, maintains a fluid *information stream*, which can be considered as a smoother *textual flow* whose goal is to "keep the link between cohesive elements as short as possible" [Renkema, 2000, pg. 68].

The following example illustrates utterance position and preposed subordinates (for simplicity and compactness in the following figures, only the sentence with the preposed subordinate is split into clauses):

$$\text{Dave went to the store to buy food.} \tag{3.38}$$
While he was at the store, Joan took the kids to the park.

which is illustrated in Figure 3.1[6].

As well as one that illustrates clearer an idea of discourse segments:

Dave went to the store to buy food.

He bought eggs for breakfast and some steaks for dinner.

While Dave was at the store, Joan took the kids to the park.   (3.39)

They love going to the park.

There are usually other children and dogs there for them to play with.

---

[5]Renkema refers to *topicality* instead of salience. For our purposes, the terms can be considered equivalent.

[6]The assumption behind this and the following examples is that if the subordinate clause were in sentence-final position, such as in *'Dave went to the store to buy food. Joan took the kids to the park while he was at the store.'*, then the text becomes less coherent. Thus, these examples are to illustrate the role of sentence-initial subordinates in discourse continuity and how to best model them.
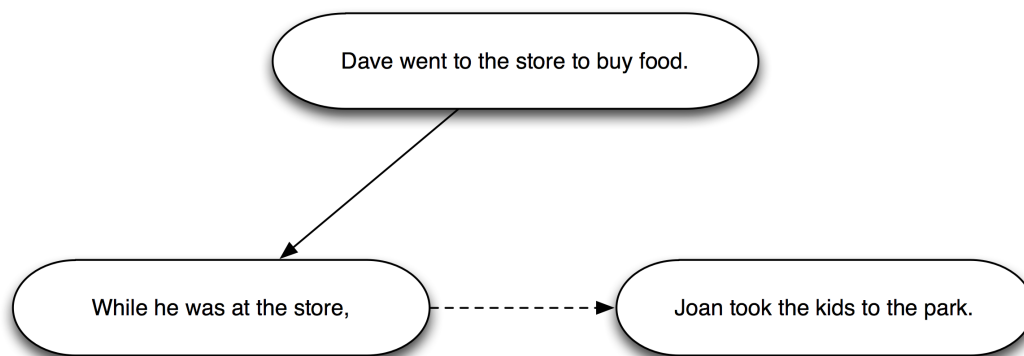
Figure 3.1: Example of linking between utterances. The solid line indicates linking between two adjacent sentences. The dotted line represents linking within a sentence between two clauses.

which is illustrated in Figure 3.2. If this example were to treat the utterances as a *sequence*, then it would be found that there would be a gap existing between the subordinate and main, as in Figure 3.3. Miltsakaki [2004, pg. 31] mentions the problem and says that "treating subordinate clauses as independent units (utterances) yields counter-intuitive topic transitions". Hence the sentential link, where utterances belonging to the same sentence are implicitly linked (as shown by the dotted line in Figure 3.2), is part of what is subsumed as *conjunctive* coherence, as described in Section 1.3.2.

Similar to the above two examples is the domain of attributive complement clauses, as mentioned by Miltsakaki [2003, pg. 8]:

> An important class of tensed subordinate clauses... is complement clauses. Un-
> like adjunct subordinate clauses, complement clauses serve as arguments of cer-
> tain verb classes (e.g., verbs of saying, knowing, belief, doubt, etc.) and occupy
> the relevant argument position in the syntactic structure... complement clauses
> seem to open up new discourses that may span across multiple clauses in the
> subsequent discourse:
> a. John said that Mary could not come to the party.
> b. He decided to hold the party anyway.
> c. He invited a lot of people
> d. who accepted his invitation.
> e. However, he was still unhappy

f. because his guest of honor would not be there.

Therefore, when processing discourse following a complement clause, one has to identify first whether the subsequent discourse is associated with the complement or the higher clause. This ambiguity is also present when the complement clause contains another subordinate clause. In (8) for example, the because clause is associated with the higher verb in the main clause while in (9) it is associated with the complement clause.

(8) John said that Mary could not come to the party because he was embarrassed to admit that he hadn't invited her.

(9) John said that Mary could not come to the party because she was sick.

The example illustrates what Miltsakaki refers to as an *unfolding* of a "new discourse". This type of discourse structure is best captured by the illustration of Figure 3.2, where one part of the sentence (e.g., the main clause) links to one discourse segment, and another part of the sentence (e.g., the subordinate clause) creates a link to an entirely different discourse segment. Thus a sentence of this type would make an ideal candidate not only as a bridge between two discourse segments, but as a point of demarcation for discourse segments. But with respect to such a demarcation, Miltsakaki [2003, pg. 206] points out that:

> it is not clear what strategies are used to help the reader recognize when the discourse opened under the complement clause is completed, the conditions under which it is possible to return to a topical entity introduced in the higher clause, and the kinds of referring expressions that can be used when doing so.

With respect to the relative clause, she says [Miltsakaki, 2003, pg. 201]:

> Configuring the relative clause as an atomic unit for updating topics yields a Centering topic transition of relatively low coherence. In contrast, when the same relative clause is processed together with its main clause, topic continuity is recovered and the use of a pronoun for reference to the topic of that unit (established in the main clause) is to be expected as a signal of a discourse continuing on the same topic.

This is another example of how an incoherent value would be computed for a coherence measure whose unit of utterance is a clause. As mentioned earlier, discourse segments and their identification are outside the scope of this thesis. We have pointed out the problems

with various instantiations of 'utterance' and propose a model that is able to handle all instantiations simultaneously.

The premise behind the utterance weighting of clauses and sentences is to capture the idea that even if two sentences that form a non-sequitur are conjoined into one sentence (e.g., "John likes movies and the earth is round.") that treating them as a single sentence instead of two separate sentences with the recency weighting preserves the speaker's intent of combining them into a single sentence. Thus, breaking "John likes movies and the earth is round" into two separate clauses has its coherence preserved over the form "John likes movies. The earth is round." which is two separate sentences. "John likes movies and the earth is round." would have the weighted representation of (using the same values for main/subordinate clauses and utterance recency):

$$\text{Sentence1: (1) (1)} \tag{3.40}$$

whereas the text "John likes movies. The earth is round." has the weighted representation of:

$$\text{Sentence1: (.5)}$$
$$\text{Sentence2: (1)} \tag{3.41}$$

Hence, two independent clauses conjoined into one sentence implies a stronger tie between the clauses over a similar representation of two independent clauses treated as two independent sentences, with this conjoining being preserved using recency weighting.

Another problem that recency can absorb is the case of empty C_f lists for an utterance. An empty C_f list means that there are no entities realized in an utterance. An example taken from Miltsakaki [2004, pg. 46]:

> It is also possible that an utterance has no C_p. For example, it is hard to identify the C_p in the following example taken from one of the students' essays:
> *Does "not possible" mean "not likely"?*

If recency were used with an utterance window that extends beyond the previous utterance, then it would be possible to handle cases such as the one just mentioned. Thus, a NOCB transition could be avoided in a model that looks back beyond the previous utterance.

This recency model can be made dynamic in handling special types of inputs, such as when NOCBs or empty C_f lists are encountered. If a NOCB gap (or empty C_f list) is

detected, then one can make both utterances of equal weight to continue the next utterance. This is done with a **modified recency weight** so that:

$$\text{Recency}(U_{n-1}) = \text{Recency}(U_{n-2}) \qquad \text{if Transition}(U_{n-1}, U_{n-2}) = \text{NOCB} \qquad (3.42)$$

Thus the weights of the two previous utterances are treated equally, or in other words, the utterance that does not contribute to the discourse coherence is *looked past* as if it did not contribute what Arnold [1998] refers to as *suppression*.

As well, if one assumes that a NOCB is a strong indicator of incoherence and it is also assumed that even a single NOCB were troublesome, then the *modified recency weight* could be used as a 'look-back' mechanism to detect whether the current utterance's entities were realized in *any* previous utterances. If not, then one could explore the different types of coherence that could be used to bridge such a gap (e.g., if a NOCB is detected and the entities of the current utterance have not been realized in the text prior, then there must be a type of relational coherence that exists that typically bridges utterances who have no entities in common, such as *cause*, as in 'The weather was very hot today and they let us off work early').

3.4 Summary

In this chapter we introduce a model that extends Centering Theory through a vector representation of utterances. The representation incorporates entity realization and entity salience, while also allowing to model how all entities transition from utterance to utterance. We show how the extended Centering Theory model can capture the basic Centering Theory model and its various transition classifications. Using recency, the extended model allows us to take into account any segment of the text in the computation of entity salience or topicality.
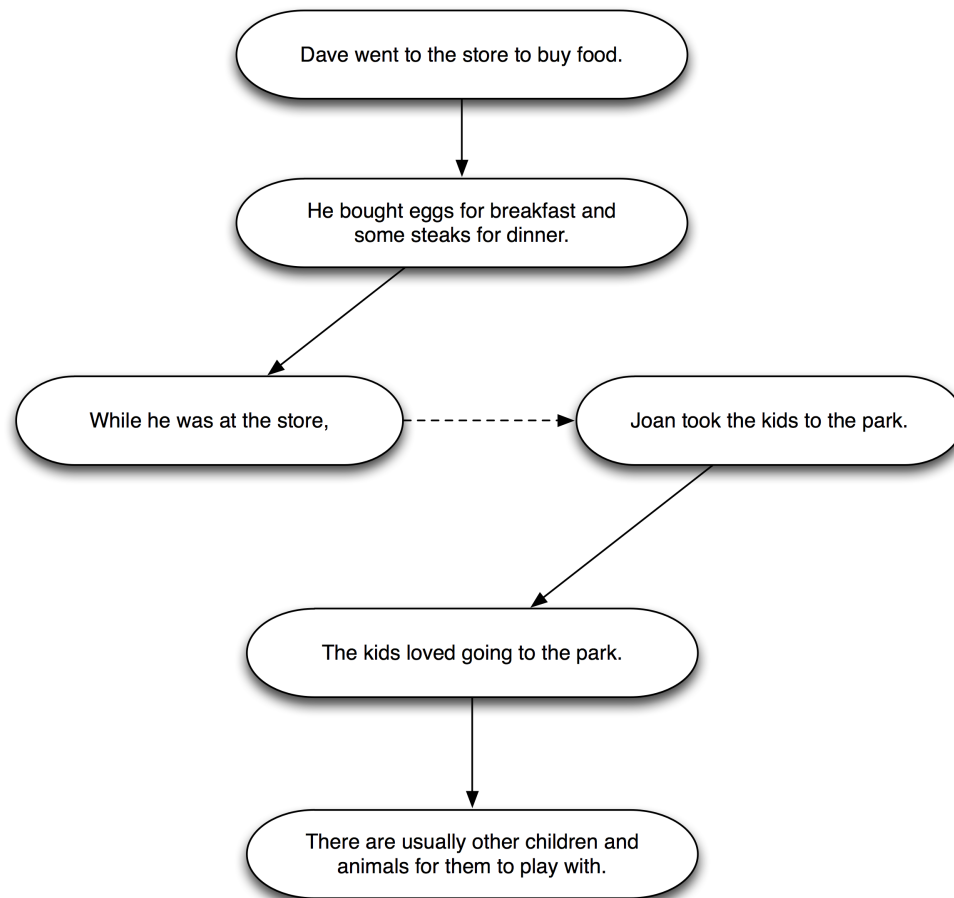
Figure 3.2: Illustration of how the two clauses of a sentence contain links to both the previous utterance and the next utterance. The preposed subordinate contains a link to the previous utterance, whereas the main clause provides a link to the next utterance. This is an example illustrating the need to treat the clause as the unit of utterance, with the clauses of the same sentence inherently being coherent due to their realization as a complex sentence.
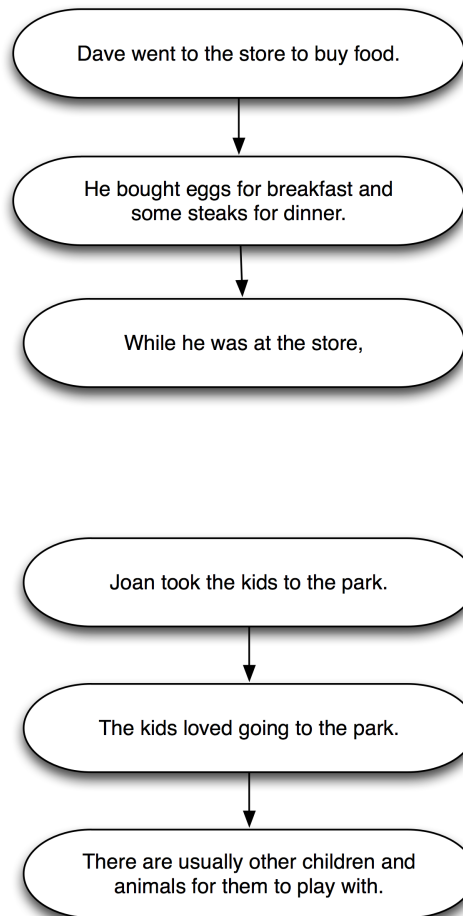
Figure 3.3: In Centering Theory, if the clause were to be used as the unit of utterance then there could result many NOCB transitions (i.e., no entities are common to the two adjacent utterances). Example of an utterance link where no shared entities exist between the subordinate and main clause exists. Thus, if the unit of utterance were the clause one could expect an increase in the number of NOCBs in a coherent text.
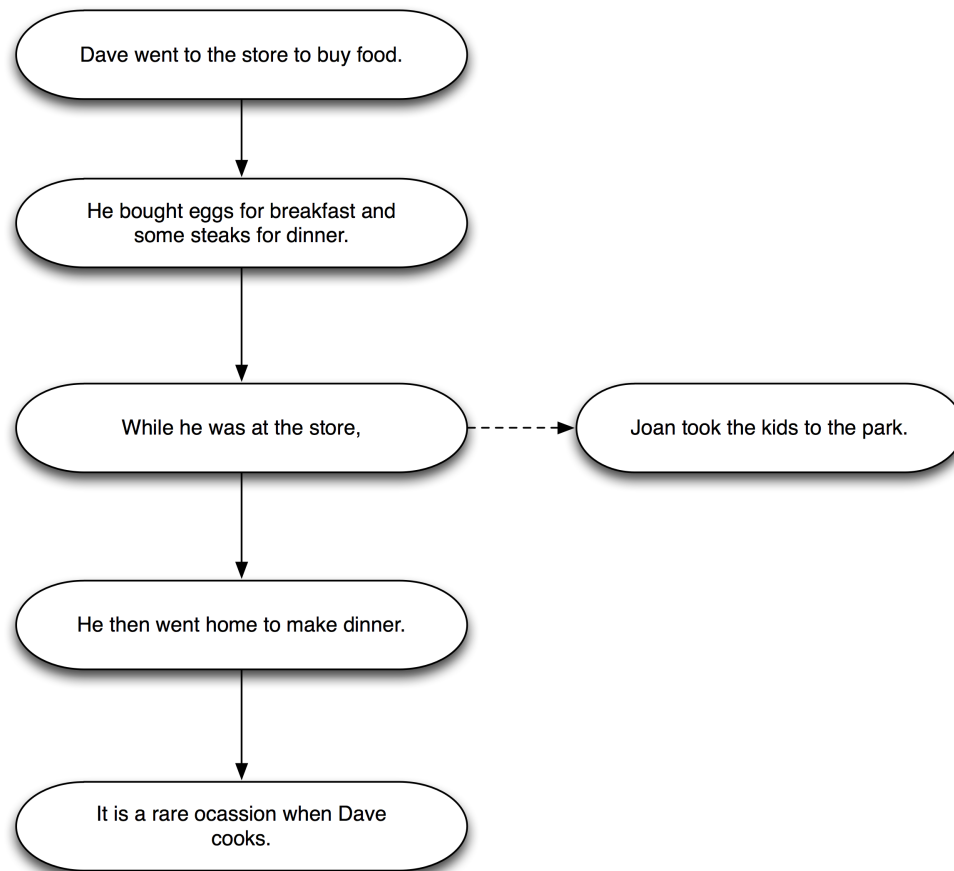
Figure 3.4: In this example, the flow of the discourse continues through the subordinate clause, with the main clause acting as an aside to the main text.

Chapter 4

Entity Coherence Trees

In this chapter, the idea of entity-based **coherence trees** are introduced. Entity-based coherence trees are similar to discourse relation trees [Mann and Thompson, 1988, Kamp and Reyle, 1993] but instead of modelling the relationships and purposes between parts of the text the entity-based tree captures the structure of the entity relationships of the sentences within the text. Such coherence trees allow the representation of a text as a tree based on each sentence's entity realization structure. The intuition and motivation behind coherence trees is to illustrate how sentences are related through their realized entities. Sentences sharing multiple entities in common should have a stronger relationship than sentences that share few or no entities. For example, The sentence pair *"Nick likes bicycles. Nick went to the store."* should be considered more coherent than *"Nick likes bicycles. Sally works late nights."*, even though both texts contain an inferential gap such that they appear as non-sequiturs.

Coherence trees are designed to solve two problems:

- measure entity coherence gaps between two sentences. Given two sentences, how coherent is the transition between them?

- identify candidates for bridging entity coherence gaps. Given a coherence gap between two sentences, how can the gap be bridged?

The coherence measure tracks all entities through their salience and realization. The vector representation of sentences discussed in Section 3.1 is used to compute *entity transitions* between sentences. Another unique aspect of the proposed entity coherence trees representation is in identifying sentence candidates for bridging such gaps. The idea of *bridging*

coherence gaps can be traced back to Clark [1977], where the types of gaps requiring bridging are inferential. The types of gaps investigated here are entity gaps.

The coherence tree structure that we outline here is similar to the discourse trees used in Discourse Representation Theory (DRT) [Kamp and Reyle, 1993] and Rhetorical Structure Theory (RST) [Mann and Thompson, 1988]. The main difference is the hierarchy is based on the number of entities realized in sentences and the branching is based on whether sentences share any entities in common, whereas DRT and RST use the relationships between utterances to organize the text into trees. A model that unifies the different types of trees into a single tree is not investigated.

4.1   Data Structure

The data structure used to represent the sentences of a text is a rooted tree. The purpose of the tree representation is to be able to efficiently compute the entity relatedness between sentences when trying to determine candidates for bridging coherence gaps. The nodes of the tree correspond to sentences, with the edges between the nodes representing that there is some entity relatedness between the sentences (i.e., the two sentences share at least one entity in common) between the connected sentences.

The nodes of a tree can contain multiple sentences, which is to handle the cases where a child has multiple parents. These parents are thus collapsed into a single node. The sentences are represented by vectors and do not necessarily all have the same vector representation. The first few examples will discuss nodes containing a single sentence, with later examples illustrating the handling of nodes with multiple sentences.

A general example of a tree with sentences as nodes is illustrated in Figure 4.1. The *Root* node is an empty placeholder holding together the entire text. The tree is not restricted to being a binary tree. The tree could have as many branches as there are nodes. Branches in a tree correspond to **orthogonality**. Orthogonality occurs when two sentences have no entities in common (i.e., in Centering Theory terms their C_f lists have an empty intersection). An example of orthogonality occurs between the two sentences "The food that remains is rotten" and "The fridge needs to be filled". The C_f list (a list of all the entities that are realized/mentioned explicitly in the sentence) for the first sentence is $\{food\}$, whereas for the second sentence it is $\{fridge\}$. Thus, these two sentences are *orthogonal* due to not having any entities in common (i.e., an empty intersection of their entity sets), even though there

is a lexical relationship that can be inferred between the entities *bridge* and *food*.
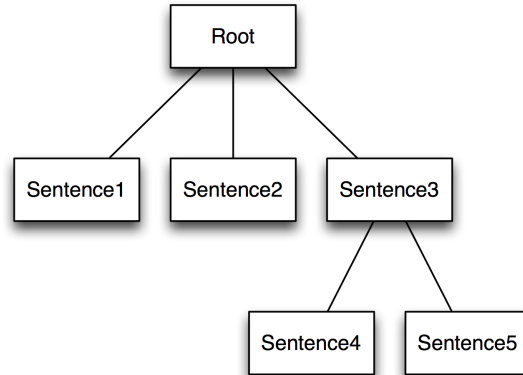
Figure 4.1: Example of a coherence tree. The nodes of the tree correspond to sentences from the source text. In this example, the text consists of five sentences. The root is an empty placeholder whose only purpose is to anchor all sentence branches.

The maximum depth a tree can have is at most the number of sentences in a text. A tree having a depth that is the same as the number of sentences would represent a text where every child node contains one less entity mentioned than its parent. An example of such a text is shown in Figure 4.4, which would consist of a text such as *"John made Mary bring the box to the store. Mary was tired of John complaining about the box. John stopped complaining when Mary returned."*. A text with all entities mentioned in every sentence would consist of a single node having all sentences contained within it, as in the text *"Dave bought Steve new golf clubs. They went to try the golf clubs immediately."*.

The maximum number of branches (maximum width) a tree can have is, again, the total number of sentences in the text. A tree with maximum width would represent a maximally incoherent text, characterized by a continuous sequence of NOCB transitions in the terminology Centering Theory. An incoherent text in this case occurs when every sentence contains a distinct set of entities. Thus, no sentences share *any* entities in common. An example of an incoherent text represented as a tree is in Figure 4.5, where the example text is *"The streets were empty. Mice like to eat cheese. I bought a new phone today."*.
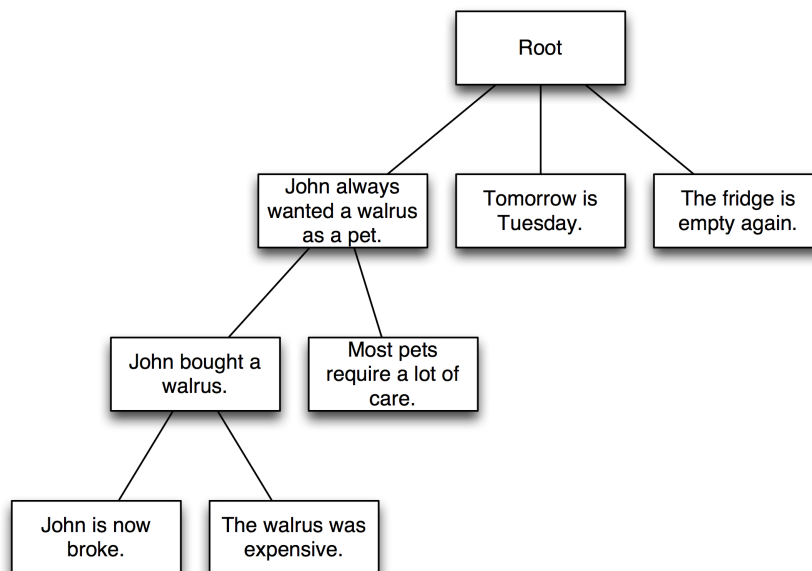
Figure 4.2: Example of a tree with actual sentences in its nodes. Sentences in separate branches reflect that they have no entities in common.

4.2 Nodes

The **nodes** of the tree represent the sentences of the text. A node can consist of multiple sentences, which would correspond to sentences with the same number of entities that realize at least one entity in common, i.e., that do not have an empty C_f list intersection. The sentences "John looked at the painting" and "The painting spoke to John" both have the same realized entities in their vector representations, $(2,1)$ and $(1,2)$. Hence, they will occupy the same node in the tree. Thus, nodes consist of sentences, with every sentence having its own (possibly distinct) vector representation. A node is not identified with a unique vector, but with all the sentences it contains.

4.3 Edges

Edges between nodes represent the fact that the nodes share at least one entity in common. Since the nodes consist of sentences that are represented as vectors, we can associate distances between the sentences in the node using the utterance transition measure proposed in Formula 3.8. Actual values are only determined along a single branch and
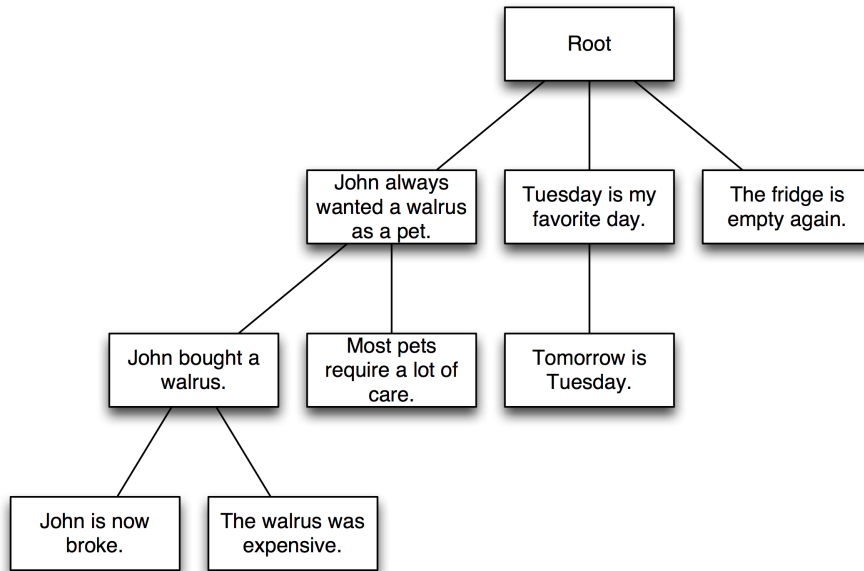
Figure 4.3: The insertion of a sentence "Tuesday is my favorite day." The new sentence gets placed in the branch which it shares entities. The level the new sentence gets placed in corresponds to the number of entities it contains. The new sentence realizes two entities, "Tuesday" and "day".

not across orthogonal branches since orthogonal branches correspond to orthogonal vectors which in turn mean there is zero similarity. Orthogonal branches automatically receive a score of 1, which represents maximum incoherence. A measure cannot be made across these two orthogonal paths, which would be similar to trying to provide a value to the question "What is the difference between person A's height and person B's age?" Also note that there are multiple sentences within a node, with each possibly having a different vector representation. Hence, the entire node cannot be identified with a single vector, which results in the edge between nodes not having a unique distance associated with it.

As an example of measuring distances between nodes, in Figure 4.2, there is a distance associated between the sentences "The walrus was expensive" and "John bought a walrus", but there is an infinite distance associated between "The walrus was expensive" and "John is now broke" since the nodes lie in separate branches and thus have no elements that are shared.
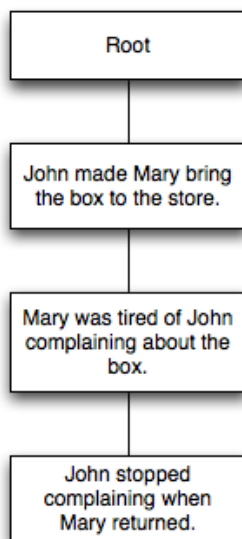
Figure 4.4: Example of a tree exhibiting maximum depth. All the sentences share at least one entity in common. The top=most sentence contains the most number of entities (4), with the bottom sentence realizing the least (2).

4.4 Relations for the Data Structure

A child/parent relationship between two nodes represents that the number of entities realized in the child node is less than the number of entities realized in the parent and that their intersection of realized entities is non-empty. An example is the two sentences "Steve went to the store to buy some groceries" and "The store was closed due to snow". Both sentences mention *store*, and the first sentence is a parent to the second due to it containing 3 distinct entities and the second containing 2. The relationship between parent/child nodes is *not* one of being a proper subset of their realized entities. The hierarchy is primarily based on the *number* of entities realized in a sentence.

Child nodes can only have a single parent node. Parent nodes can have as many children as they have realized entities. For example, the parent node can consist of the sentence *"The dog ate its food by the garage"* and its child nodes can be *"The dog was hungry"*, *"The food was warm"*, and *"The garage was locked"*, thus having 3 orthogonal child nodes.

In Figure 4.3, the node "John always wanted a walrus as a pet" contains all the sentences that contain the entities "John", "walrus", "pet" in its subtrees. Any entities not mentioned
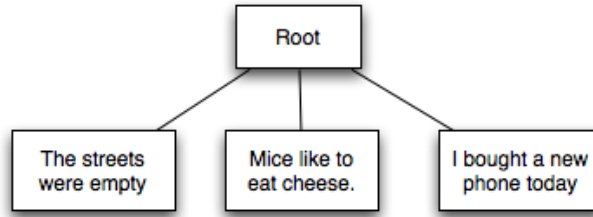
Figure 4.5: An example of a maximally incoherent text. All the branches (e.g., sentences) are orthogonal to each other (no sentences have any entities in common).

in "John always wanted a walrus as a pet", such as "Tuesday" or "fridge", will be contained in nodes above it or in other orthogonal branches.

The root node essentially is a node that implicitly contains mention of all entities within the text. Its purpose is to denote that all sentences are part of the same text.

4.5 Example of Inserting Sentences Into a Tree

The example text that will be used for illustrating how coherence trees are constructed consists of the following sentences:

$$
\begin{aligned}
&\text{The school is very far away.}_{Sent_1}\\
&\text{The fire kept growing.}_{Sent_2}\\
&\text{John drove his car to school where his mom worked.}_{Sent_3} \qquad (4.1)\\
&\text{John went to school to see his mom.}_{Sent_4}\\
&\text{John's mom was happy to see him.}_{Sent_5}
\end{aligned}
$$

The sentences have the following vectorial representation, which will be used in the figures illustrating the construction of the tree:

$$
\text{Sent} = \begin{pmatrix} \text{school} \\ \text{fire} \\ \text{John} \\ \text{car} \\ \text{mom} \end{pmatrix}, \text{Sent}_1 = \begin{pmatrix} 4 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{Sent}_2 = \begin{pmatrix} 0 \\ 4 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{Sent}_3 = \begin{pmatrix} 2 \\ 0 \\ 4 \\ 3 \\ 1 \end{pmatrix}, \text{Sent}_4 = \begin{pmatrix} 3 \\ 0 \\ 4 \\ 0 \\ 2 \end{pmatrix}, \text{Sent}_5 = \begin{pmatrix} 0 \\ 0 \\ 3 \\ 0 \\ 4 \end{pmatrix}
$$

The values (0 to 4) represent entity salience/topicality within the sentence. The highest value (4 in this example) represents the most salient entity, whereas the value of 1 represents

the least salient entity. Entities not appearing the sentence (i.e., are not realized in the sentence) are given the value 0. Salience can be determined by a number of factors (e.g., grammatical role or the order in which the entities appear). For simplicity, in the examples the entity which appears first in the sentence is considered most salient, while the entity appearing last is the least salient.

To construct a coherence tree, we begin with the leaf nodes (nodes at the bottom of the tree) and build to the root. We also start with a Root node that holds together all the orthogonal nodes. Inserting sentences into branches is determined by whether the incoming sentence has entities in common with any of the nodes in the tree. If the incoming sentence does not have any entities in common with the existing sentences, then it is orthogonal and occupies its own separate branch. If it has any nodes in common, then it becomes a parent of sentences having less entities, or it is combined in the same node if the sentences being compared have the same number of entities. In Figure 4.6 the tree consists of a single sentence from the text (e.g., "The school is very far away."). The rest of the figures for this example represent the state of the tree after inserting each sentence one at a time. We will also see that the order of sentences being inserted into the tree does not matter, provided we adhere to inserting sentences with the least number of entities first.

The sentences and the number of entities they realize are ordered as such:

one entity: The fire kept growing. The school is very far away.

two entities: John's mom was happy to see him.

three entities: John went to school to see his mom.

four entities: John drove his car to school where his mom worked.

five entities: *None*

To build a tree, begin inserting the sentences that realize the least number of entities one at a time. In this example "The fire kept growing" and "The school is very far away" each realize only one entity. Figure 4.7 represents the tree consisting of the sentences "The school is very far away" and "The fire kept growing". Since the two sentences share no entities in common, they are placed in an orthogonal relation on separate branches. Also, contrast Figure 4.7 with Figure 4.8, which represents the text "The fire kept growing" and "The school is very far away". This example illustrates how the order of orthogonal
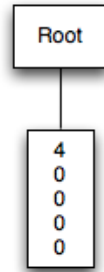
Figure 4.6: Example of a single sentence inserted into a tree. The sentences are shown using their vector representation.

branches does not matter. Coherence only becomes an issue in this model when a sentence is introduced that has entities in common with the existing nodes.
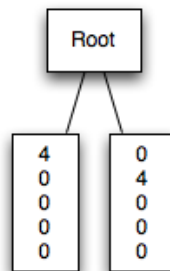
Figure 4.7: Example of a tree comprised of 2 sentences. "The school is very far away. The fire kept growing."

We next insert one at a time the sentences that realize two entities. In our example text, only one sentence satisfies this criteria. Figure 4.9 illustrates the sentences "The fire kept growing", "John's mom was happy to see him" and "The school is very far away." "John's mom was happy to see him" is orthogonal to both sentences existing in the tree, "The fire kept growing" and "The school is very far away". It is thus placed on its own orthogonal branch.

We next insert all sentences realizing three entities. Again, we only have a single candidate: "John went to school to see his mom." Figure 4.10 illustrates the tree after we insert this sentence. In this case the sentence becomes a parent to two existing orthogonal nodes. The orthogonal nodes become children to the parent, while still retaining their orthogonal
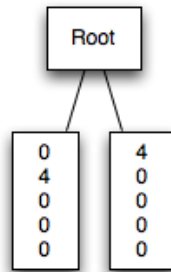
Figure 4.8: Example of 2 sentences with order reversed to illustrate the coherence of exactly 2 sentences is not dependent on order. "The fire kept growing. The school is very far away."
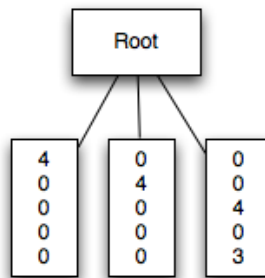
Figure 4.9: The tree after a 3rd sentence is introduced. "The fire kept growing. John drove his car to school where his mom worked. The school is very far away."

relationship. Hence, this can be read as the parent having an entity in common with its children, but the children having no entities in common with each other. This relationship is important, as it provides the basis for finding a candidate sentence that bridges the gap between two orthogonal sentences.

Similarly, the final sentence is inserted, which has something in common with the previously inserted sentence. Figure 4.11 illustrates the final tree after all sentences have been inserted.
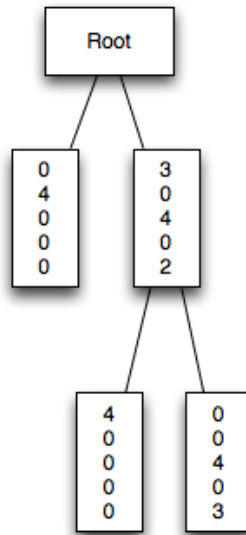
Figure 4.10: The tree after a 4th sentence is introduced. "The fire kept growing. John drove his car to school where his mom worked. John went to school to see his mom. The school is very far away."

4.6   Special Cases of Coherence Trees

Even though the above example illustrates the core idea behind constructing coherence trees, it covers a fairly straightforward example. The example does not detail how special cases are handled. This section covers the different cases involving parent and child relationships in coherence trees.

4.6.1   Non-Orthogonal Children Where Neither Node is a Subset of the Other

The example in the previous section illustrates the simplest case where each child's sentence entities are a proper subset of its parent, in other words all the entities of a child sentence are contained in the parent. A difficulty exists in the vector representation shown in Figure 4.12, whose vectors are as follows:

$$\text{Sent} = \begin{pmatrix} \text{John} \\ \text{cat} \\ \text{dog} \end{pmatrix}, \text{Sent}_1 = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}, \text{Sent}_2 = \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix}, \text{Sent}_3 = \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix}$$
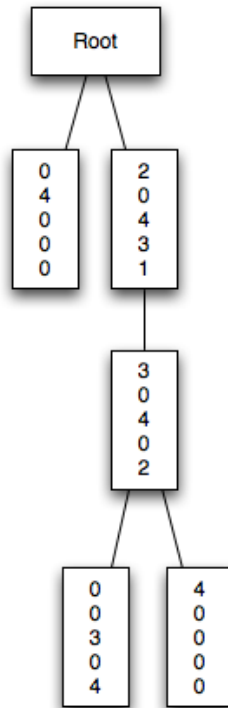
Figure 4.11: The final tree after the 5th sentence is inserted. "The fire kept growing. John drove his car to school. John went to school to see his mom. The school is very far away. John's mom was happy to see him."

The sentences "John likes cats" and "John likes dogs" are not orthogonal, but they are also not subsets of each other. We handles this case by collapsing the two nodes into a single node since they both contain the same number of entities. Otherwise, if they contained different numbers of entities, then the node with more entities would be the parent to the node having less entities. The resultant tree after collapsing nodes is shown in Figure 4.13.

If we generalize the vectors in 3-space and construct a tree from them, we are presented with the following representation that consists of all possible permutations for vectors having 3 elements (i.e., all different combinations of sentences with 3 entities):

$$\text{Sent}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \ \text{Sent}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \ \text{Sent}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \ \text{Sent}_4 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \ \text{Sent}_5 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \ \text{Sent}_6 =$$
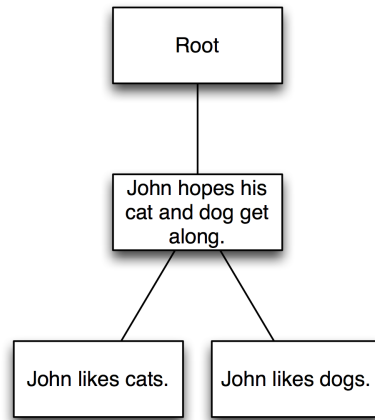
Figure 4.12: A text with a partial number of entities in common ("John") and other entities not being in common ("cat", "dog"). Orthogonality does not exist and therefore cannot be placed along separate branches as they are shown in the figure. Because "John" is common to both of the sentences at the leaves, the two nodes must be collapsed into a single node containing both sentences.

$$
\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \text{Sent}_7 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.
$$

The question becomes how to order Sent_1 to Sent_7, which is concisely illustrated in the previous example of Figure 4.12. The idea is that the parent sentences cannot be considered strictly orthogonal since they both share a common child. The reason the roles of the child and parents is not reversed and the child does not become a parent of its parents and the parents do not become the children is that the number of entities takes precedence over the number of entities in common. The parents have more entities than its children, and the assumption is that they capture more information than sentences with less entities, hence taking precedence in the tree hierarchy.

Figure 4.14 shows a generalized form of a text with 7 sentences having 3 distinct entities. The sentences with 2 entities each are collapsed into a single node because they have at least one entity in common with each other.
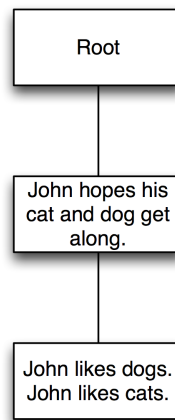
Figure 4.13: A text with some entities in common ("John") and other entities not being in common ("cat", "dog"). The two sentences are combined so that they become a single node.

4.6.2 Child Node With Multiple Orthogonal Parents

This section describes how child nodes having multiple parents (with the parents being orthogonal to each other) are handled. The parent nodes are combined into a single node in order for the child to have a single parent.

Figure 4.15 shows two sentences having no entities in common and are thus orthogonal ("John likes cats from Persia" and "Susan's dog is barking at the mailman"). Figure 4.16 introduces a third sentence ("The dog is barking at a cat"). This third sentence has one entity in common with each of the parent nodes ("cat" and "dog"), and thus has two distinct parents. To handle this situation, the two sentences of the orthogonal nodes are combined into a single node, as illustrated in Figure 4.17.

4.7 Algorithm for Constructing a Coherence Tree

The algorithm constructs a coherence tree from the bottom up. In other words, the tree is built from its leaves to its root. This allows orthogonal parents with common children to be merged into a single node when encountered. This procedure is repeated until the tree is built to the root node.

The following steps are followed in constructing a coherence tree:

- part-of-speech tags and syntactic parsing is performed on original text
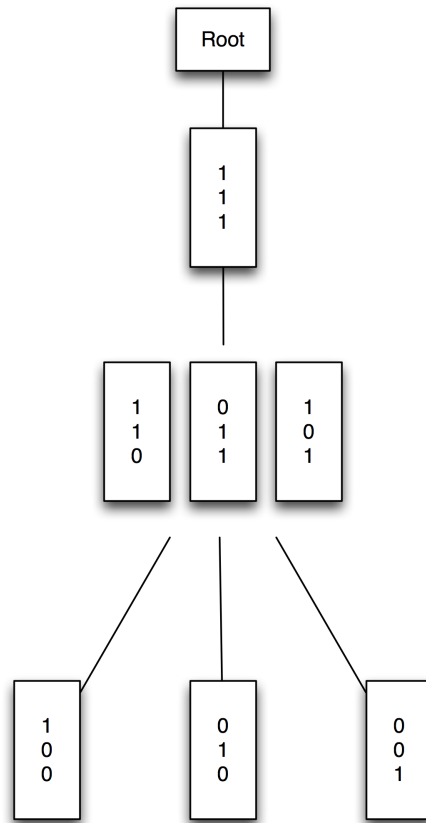
Figure 4.14: Example of a full tree showing multiple parent nodes for children in 3-space (maximum 3 entities in the text). The multiple parents are resolved by collapsing parent nodes into a single node. Thus the nodes at the middle level that contain two entities are collapsed into a single node containing the 3 sentences.

- coreference resolution is performed

- sentences have their entities identified

- sentences are sorted in ascending order according to the number of entities they contain

- sentences are inserted one at a time into the coherence tree beginning with the sentences with the lowest number of entities

Ignoring the linguistic preprocessing (part-of-speech tagging, etc.), the steps for building an entity coherence tree are shown in Algorithm 1. One of the steps involved in building a tree is inserting a sentence. Algorithm 3 describes inserting a sentence into a tree branch.
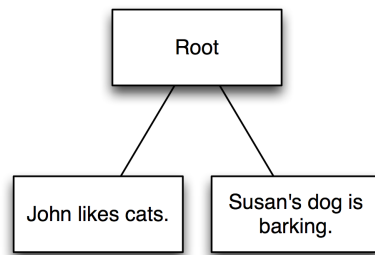
Figure 4.15: Orthogonal nodes. The two sentences have no entities in common, thus they are placed in separate orthogonal branches.
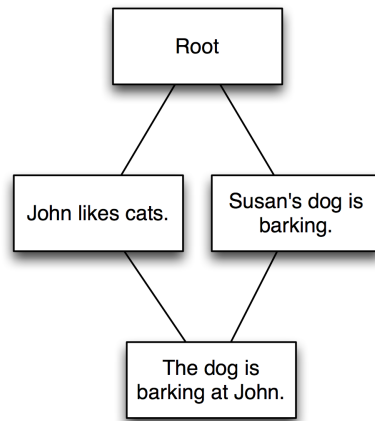
Figure 4.16: Child node having multiple parents that are orthogonal. The sentence "The dog is barking at John" has two parent sentences which have no entities in common with each other. Since they share a common child, the parents must be collapsed into a single node.

Merging of branches by combining the root nodes of two orthogonal branches may be necessary when building a tree. Algorithm 2 describes how branches are merged during the sentence insertion process.

Algorithm 1 Steps For Building An Entity Coherence Tree

1: Input: plain text document
2: split text into sentences: $\{\text{sentence}_1, ..., \text{sentence}_K\}$
3: $n \leftarrow$ the number of entities in the sentence containing the most entities
4: Partition sentences according to the number of entities they realize (Sets C_1 ... C_n contain sentences that realize n entities)
5: $i \leftarrow 1$
6: **while** $i <= n$ **do**
7:     **while** $C_i \neq \emptyset$ **do**
8:         Randomly choose sentence_j from C_i
9:         **if** sentence_j is orthogonal to all branches **then**
10:             Create new branch with sentence_j as its only node
11:         **else if** sentence_j intersects with more than one branch **then**
12:             Merge branches that are not orthogonal with sentence_j ▷ See Algorithm 2
13:         **else**
14:             Insert sentence_j into tree branch   ▷ see Algorithm 3
15:             Attach new branch back to Root node
16:         **end if**
17:         Remove sentence_j from C_i
18:     **end while**
19:     $i \leftarrow i + 1$
20: **end while**
21: Output: tree consisting of sentences as nodes

Algorithm 2 Merging Root Nodes Of Branches In An Entity Coherence Tree

---

1: Input: sentence_j to be inserted into tree, set of orthogonal tree branches that are not orthogonal to sentence_j
2: $B \leftarrow$ set of orthogonal tree branches
3: $ents \leftarrow$ number of entities contained in sentence_j
4: remove all branches b_i from B s.t. their parent node contains $ents$ number of entities
5: combine all root nodes of every b_i and sentence_j into a single node (maintaining distinct subtrees as being orthogonal) thus forming a new tree
6: **while** $B \neq \emptyset$ **do**
7:     remove one branch from B
8:     add branch to new tree as a child to the root node
9: **end while**
10: **return** new tree with merged root nodes of branches
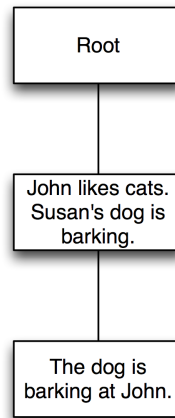
Figure 4.17: Child node having multiple parents that are orthogonal with the parent nodes combined into a single node. The two orthogonal parent nodes are collapsed into a single parent node.

Algorithm 3 Inserting A Sentence Into A Branch Of An Entity Coherence Tree

---
 1: Input: tree branch, sentence$_j$ to be inserted into tree
 2: Add all entities of the branch into sentence$_j$'s node entities list
 3: Make sentence$_j$ parent node of branch
 4: **return** branch with sentence$_j$ as parent node

4.8 Coherence Bridges and Gaps

4.8.1 Identifying Coherence Gaps Using Sentence Vectors

Using the coherence measure obtained from our sentence vector representation, coherence gaps are identified by a threshold ranging in value from 0 to 1[1]. All sentence pairs are assumed to have some type of coherence gap associated with them, except for the case where two sentences have identical entity salience rankings associated with their constituent entities (i.e., the sentences are entity-equivalent and thus have no gap). The objective here is to first quantify what constitutes a coherence gap that corresponds to incoherence.

The utterance transition measure from Formula 3.8 is used to determine the size of the

---

[1]We have found that an appropriate threshold (i.e., one that indicates a *rough-shift*) is dependent on the maximum number of entities in a sentence over the whole text. This is equivalent to coming up with a cutoff value that decides whether a text is *coherent* or *incoherent*, which is a grey area and not a black-and-white decision. This investigation for a cutoff value that demarcates the boundary between a text being coherent or incoherent, if such a boundary can exist, is left for future study.

coherence gap that results between two utterances. The transition value lies in the range 0 and 1, with 0 corresponding to maximum coherence and 1 being maximum incoherence. The question is where to draw a cutoff in terms of a coherence gap needing to be bridged (e.g., does 0.5 constitute a cutoff? 0.75?). A main impediment is due to texts with sentences realizing different numbers of entities having their cutoff shifted due to the transition metric not normalizing against such different texts. The normalization of the overall coherence score based on the size of the text is left for further investigation (e.g., one may first evaluate the coherence of individual discourse segments, and then measure the coherence of the combined discourse segments similar to how the coherence of utterances were computed).

4.9 An Application of Coherence Trees

4.9.1 Bridging Coherence Gaps Using Coherence Trees

After a coherence gap has been detected using the sentence vector model, the next step is to determine how that gap can be bridged. Using the coherence tree, we choose the lowest common subsumer in the tree for bridging the two sentences having the gap. This approach is motivated by Centering Theory, where the bridging sentence is intended to improve the type of Centering transition from one that is not preferred (e.g., a NOCB transition) to one that is preferred (i.e., a retain transition).

The following examples illustrate the spectrum of the different types of candidates for bridging gaps:

- A coherence bridge that preserves lexical and relational coherence. The original text is "John went to the store. Mary doesn't believe in unicorns." The bridged version is "John went to the store. John wanted to buy a book for Mary on the history of mythological creatures. Mary doesn't believe in unicorns."

- A coherence bridge that preserves lexical and relational coherence with both entities in subject position in the bridging sentence. The original text is "John went to the store. Mary doesn't believe in unicorns." Bridged by "John went to the store. John and Mary wanted to rent a movie on fairy tales. Mary doesn't believe in unicorns."

- A coherence bridge that preserves lexical and relational coherence, but reverses the coordinated clauses while maintaining both entities in subject position. The original

text is "John went to the store. Mary doesn't believe in unicorns." Bridged by "John went to the store. Mary wanted to read a book about mythological creatures and John was going to buy it for her. Mary doesn't believe in unicorns."

- A coherence bridge that does not preserve either lexical or relational coherence, but maintains both entities in subject position and an ordering within the coordinated sentence that transitions in topic smoothly (i.e., first and second utterances involve 'John', third and fourth utterances involve 'Mary'). The original text is "John went to the store. Mary doesn't believe in unicorns." Bridged by "John went to the store. John's favorite month is October and Mary just received her driver's license. Mary doesn't believe in unicorns."

- A coherence bridge that does not preserve either lexical or relational coherence, but maintains both entities in subject position with an ordering within the coordinated sentence that is not a smooth topic transition (i.e., first and third utterances involve 'John', second and fourth utterances involve 'Mary'). The original text is "John went to the store. Mary doesn't believe in unicorns." Bridged by "John went to the store. Mary's favorite month is October and John just received his driver's license. Mary doesn't believe in unicorns."

The bridging candidate sentence is taken from the first node that is parent to both sentences which contain a gap between them. As can be seen by the examples above, the purpose of the bridge sentence is to link two sentences which have no entities in common.

An example that illustrates the need for lexical coherence in entity coherence is the following example: "The buildings shook from the explosion. The windows and doors were all damaged." These two sentences have no entities in common. Implicit in these sentences are the lexical (i.e., mereological) relationships between a "building" and its constituent "windows" and "doors". A possible candidate bridging sentence in this case would be "Buildings consist of windows and doors", which is common knowledge that is always implied between "buildings" and "doors" and rarely expressed explicitly. Thus, incorporating lexical into the entity coherence model would identify that there is no coherence gap and also modify the structure of the tree such that the two original sentences are not actually orthogonal.

Coherence trees can be used in automatic summary systems where there may be coherence gaps between sentences. The sentences that are extracted from documents are typically not consecutive in the original text, thus an automatically generated summary may exhibit

coherence gaps no matter how perfect the ordering of the extracted sentences is. In such a case, a coherence tree of the original documents can be constructed, with the bridging sentence chosen from the tree. From our analysis of output obtained from existing summary systems (see Appendix and Section 5.2) reveal that the type of incoherence is primarily causal and inferential, and not entity incoherence. This is mainly due to the summaries focusing on single topics and thus having similar entities in common for all sentences, which is one of the factors involved in the summaries themselves. The sentences extracted for the summary are also very long, with the objective being to contain as much information relating the entities as possible. Thus, the sentences chosen as part of the summary contain many entities, which result in coherence measures that would not be classified as gaps. Further experimental studies need to be performed on automatically summarized data that does not emphasize sentences that are lengthy or realize many entities as part of the summarized output.

4.10   Summary

This chapter has looked at how entity coherence trees can be constructed and their utility toward bridging coherence gaps. Coherence trees can be used to illustrate global entity structure in a text based on the premise that sentences that realize more entities are parents to sentences that realize fewer entities (i.e., the sentences realizing more entities are more informational, thus occur higher in the tree).

The principles for constructing a coherence tree are summarized here in order of priority. Priority is determined by "sentence entity salience". The following define the criteria for what it means for a sentence to have more "entity salience" than another sentence. The coherence tree essentially captures the hierarchy associated with the the sentences' entity salience of the entire text. The criteria, in order, are:

- a node with multiple parent nodes that are orthogonal is the child of a single parent node that combines the two orthogonal parent nodes

- nodes with no entities in common are orthogonal and create separate branches

- the node having more entities is the parent node

- nodes that have entities in common (as well as having the same number of entities) are combined into a single node. This sets the preference for sentences having more

entities as parents of sentences that have less sentences. The attempt here is to state that *sentence salience* is determined by the number of entities discussed in a sentence.

Chapter 5

Experimental Evaluations

There are two experiments that will be performed in order to illustrate the effects of extending Centering Theory and to evaluate the applications of our representation: sentence ordering and evaluating text coherence. These experiments will evaluate the ability of the systems under comparison to determine the quality of coherence in a text. The system proposed in this thesis will be compared against coherence measures that adhere to Centering Theory principles, as well as best-of-class systems for their respective experiment.

5.1 Experiment: Sentence Ordering

5.1.1 Overview of Sentence Ordering

Sentence ordering is the process of automatically determining the most coherent ordering of a set of sentences. The sentences could consist of a set of facts extracted from a database or sentences extracted from multi-document summaries, where the objective is to present the sentences in an ordering which is most coherent. Applications for such a task include natural language generation and multi-document summarization. In both applications, the objective is to determine in which way to present the sentences such that there is a coherence and naturalness to their order. Other applications for sentence ordering include essay-scoring system feedback [Miltsakaki & Kukich, 2004], where the system would provide feedback to the user for a more coherent ordering of their essay structure.

In the earliest paper on discourse, the difficulties of sentence ordering and the way it relates to discourse were considered by Zellig Harris [1952, pg. 9]:

Here we have no prior linguistic knowledge to tell us which orderings of sentences (if any) are automatic and therefore not to be represented, or which orderings can be replaced by different but equivalent orderings. A closer study of sentence sequences in the language may some day give us such information in the future; for instance, to take a very simple case, it might show that sentence sequences of the form P because Q are equivalent to sequences of the form Q so P, or that P and Q is interchangeable with Q and P (whereas P but Q may not be similarly interchangeable with Q but P). Furthermore, a closer study of a particular text, or of texts of a particular type, may show that certain whole sequences of sentences are equivalent or inter-changeable; and with this information we may be able to simplify the vertical axis of the double array[1], for example by finding periodically repeated vertical patterns.

Zellig also makes explicit his concern for the role of clause order and its effect on sentence ordering, both in compound and complex sentences. He foreshadows the role of discourse segment order (*"whole sequences of sentences are equivalent or inter-changeable"*), and how the discourse segments themselves might be identified through lexical entity chains (by way of *"periodically repeated vertical patterns"*). Grosz and Sidner also mention the importance of utterance sequence within a discourse [1986, pg. 177]:

The first component of discourse structure is the structure of the sequence of utterances that comprise a discourse. Just as the words in a single sentence form constituent phrases, the utterances in a discourse are naturally aggregated into discourse segments.

Using our definition of coherence, whose objective is such that it minimizes the change in entity salience in a text, the objective is then to arrange the sentences in such a way that we are minimizing changes in the sentence's entity salience from one utterance to the next.

5.1.2 Previous Work

Althaus et al. [2006] show that the problem of optimally ordering information is equivalent to the traveling salesman problem (TSP). The objective in a traveling salesman formulation is to minimize a cost function associated with ordering sentences. The cost function

---

[1]In modern terminology, a "double-array" is a matrix. The representation used for a text is also a matrix, where the rows correspond to the entities and columns correspond to utterances.

is ideally a function that measures the amount of coherence between adjacent sentences. Hence, the order which produces the minimum cost (i.e., minimizes incoherence) between sentences is the order that maximizes coherence.

Conroy et al. [2006] follow Althaus et al. by computing a sentence similarity measure for multi-document summarization using a distance metric that is based on the number of terms the two sentences have in common (if sentences came from the same document, the distance was multiplied by a factor of 1.6). Their formula for the similarity between sentences j and k is:

$$c_{jk} = 1 - \frac{b_{jk}}{\sqrt{b_{jj}b_{kk}}}$$

where b_{jk} is the number of terms in common between sentences j and k, the denominator is a normalizing factor, and $c_{jk} \in [0, 1]$ is the final distance. A distance of 0 represents two essentially identical sentences, and 1 corresponds to what has been discussed in this thesis as NOCB.

The sentence transition measure that is proposed in this thesis (Equation 3.8) is different from the sentence similarity measures used in the TSP formulations of their cost function. By employing a vector representation, as used in this thesis, one could use a variety of available vector similarity measures that determine distance between vectors (e.g., cosine similarity, Levenshtein distance, Jaccard index, etc.). The Transition metric of Equation 3.8 is motivated by principles of Centering Theory (transitions between utterances are based on entity salience change between utterances) and also linguistic principles (utterances that realize a common set of entities are given stronger similarity than sentences that don't have as many entities in common, i.e., two sentences are 'about' the same set of entities). The other vector similarity metrics use angles, lengths, or intersection as their basis for similarity, whereas the Transition metric relates the *change*, or *differences*, of the two vector's elements.

The current best-of-class sentence ordering system that will be used for comparison in these experiments is the model found in Barzilay and Lapata [2008]. Barzilay and Lapata's model is based on an entity-grid model that utilizes machine learning (Support Vector Machines). Their training and testing used SVM^{light} [Joachims, 2002]. Their model intends to capture the movement of an entity's change in grammatical role from sentence to sentence. For example, the entities can change role between any of subject (S), object (O), other (X), or unrealized (-). The transitions between roles are assigned a probability that is computed

from what is observed in each text. For example, transitions from (-) to (-) are the most frequent, whereas transitions from (S) to (O) are less frequent. Barzilay and Lapata also consider *salience*, but their definition of salience is more global than the notion of salience referred to here. Salience in their system is based on the frequency of a word in the text. To minimize confusion, we will use the word *frequency* when referring to their salience feature. The assumption is that word frequency is based on discourse prominence. In addition to grammatical role (which they classify under the feature of *syntax*) and frequency, the other feature of a text that they considered was *coreference resolution*. Their experiments evaluate every permutation involving all three linguistic features, but for simplicity we only cite the models that include either all the features or none. See [Barzilay & Lapata, 2008] for full performance results.

Barzilay and Lapata's system is trained with the objective of yielding a ranking score that "minimizes the number of violations of pairwise rankings in the training set" [2008, pg. 11]. The features of their model correspond to entity transition probabilities. In their study, they compare their system against a number of other systems, the main ones of which are described in the next section.

5.1.3 Experiment and Results

One way of simulating the problem of determining coherent sentence orders, is to take an already coherent text and permute its sentences, then have the systems determine which is the original text and which is the permutation, with the assumption being that the original text is the most coherent. This is a much more straightforward and simpler evaluation than to have to determine an *absolute* most-coherent ordering, of which there may not be a single *unique*, best ordering. It is possible that there may be multiple permutations of a text that would be considered acceptably coherent or equivalently coherent to each other. Madnani et al. [2007] look at the variability amongst humans in terms of the sentence ordering task.

The data used is obtained from Barzilay and Lapata [2008]. Two different genres were collected: newspaper articles and accident reports written by government officials. The newspaper articles are from the North American News Corpus [Graff, 1995], with the topic of the articles being earthquakes. The accident reports consist of aviation accidents and are taken from the National Transportation Safety Board aviation accident database[2]. Rather

---

[2]http://www.ntsb.gov/avdata/

than find an optimal ordering, the objective is to rank pairs of text permutations with the assumption that the original text is more coherent than any of its permutations. Only 20 permutations are generated since the objective of this experiment is to identify the original text from its permutation, with the assumption that the original text is more coherent than any of its permutations.

The data that we use from Barzilay and Lapata treats sentences as their unit of utterance. In many cases, there are multiple subjects associated with each sentence (e.g., in the cases of complex sentences and compound sentences where subordinate clauses also have subjects). Since there can possibly be multiple entities assuming the rank of subject (one for the main clause, and one each for all subordinate clauses), linear order is also incorporated in the salience measure for our proposed system. So in various instantiations of our model, the first mentioned subject has a higher salience rank than the subjects mentioned later.

Our system employs OpenNLP[3] for its front end that determines the entities of the text. OpenNLP performs part-of-speech tagging, constructs syntactic parse trees, and coreference resolution. OpenNLP is a tool necessary for the detection of entities in our system. After identifying all NPs and its referents[4], our system constructs entity vectors based on the entities in the sentence. The vectors have an associated transition measure, upon which the sum of all the transitions is used in determining the more coherent ordering. We sum the transitions between adjacent sentences and use that value in choosing the text with the lowest score as the more coherent ordering.

In addition to Barzilay and Lapata's system, the other sentence ordering systems that will be compared in the evaluation are a HMM-based content model [Barzilay & Lee, 2004] and a model based on Latent Semantic Analysis [Foltz et al., 1998]. These systems, along with Barzilay and Lapata's system, were not reimplemented for this experiment. Instead, the results from Barzilay and Lapata [2008] are cited here for comparison. Implementations based on Centering Theory (that were not evaluated in their study), such as NOCB and

---

[3]Available at http://OpenNLP.sourceforge.net/

[4]Except for the instances where OpenNLP has resolved an entire noun phrase as a single entity, we treat all entities in compound nouns as being distinct. For example, in the case of *John's sister*, this is treated as two separate entities: *John* and *sister*. A version of our system had considered treating NPs as single entities, but this had a negative impact on performance, i.e., the entities *John's sister* and *John* would not have a relationship within the sentence vector model since they would be considered distinct. Thus, if one wanted to capture the relationship between compound nouns and their individual nouns then lexical cohesion would need to be incorporated in such a system. Hence, treating the individual components of compound nouns is a simple implementation of lexical cohesion, which improves performance of the entity coherence model.

rough-shift transitions, were implemented for this experiment.

The *HMM-based content model* is based on a hidden Markov model. In an HMM model, the states represent distinct topics, and state transitions are the probability of changing from one topic to another. The content models are domain specific e.g., capturing probabilities of topic changes between 'building construction' to 'earthquake strength' does not allow the same model to be used in a different domain, such as computer reviews. Sentences were used as the units for the topic models. The criteria for determining coherence between two text permutations is selecting the one with the higher probability based on their HMM model. Two separate topic models were constructed by Barzilay and Lapata: a model for the Accident corpus, and a model for the Earthquake corpus.

The *Latent Semantic Analysis* model involves looking at the semantic relatedness between sentences. A multidimensional space is used to capture a word's co-occurrence with neighbouring words. A matrix of frequencies is used to model the semantic space of the linguistic context, where the context can take the form of a document, paragraph, sentence, etc. Singular value decomposition reduces the dimensionality to provide latent structure in the data. The sentence's meaning is represented as a vector that is the mean of the vectors of its words. Sentence similarity is performed by taking the cosine similarity of the means of two sentences [Foltz et al., 1998]:

$$
\begin{aligned}
\mathrm{similarity}(S_1, S_2) &= \cos(\mu(\vec{S}_1), \mu(\vec{S}_2)) \\
&= \frac{\sum_{j=1}^{n} \mu_j(\vec{S}_1)\mu_j(\vec{S}_2)}{\sqrt{\sum_{j=1}^{n}(\mu_j(\vec{S}_1))^2}\sqrt{\sum_{j=1}^{n}(\mu_j(\vec{S}_2))^2}}
\end{aligned}
\tag{5.1}
$$

where n is the number of sentences, $\mu(\vec{S}_i) = \frac{1}{S_i}\sum_{\vec{u} \in S_i} \vec{u}$, with $\vec{u}$ the vector for word u. The coherence of the text T is computed by averaging the similarity measure of all adjacent sentence pairs S_i and S_{i+1}:

$$
\mathrm{coherence}(T) = \frac{\sum_{i=1}^{n-1} \cos(S_i, S_{i+1})}{n - 1}
\tag{5.2}
$$

The LSA model is mainly a lexical coherence measure, which is in contrast to Barzilay and Lapata's model and our entity vector model that consists primarily of transition sequences.

The *NOCB* model is based on Centering Theory, and counts the number of NOCB transitions that occur between utterances. This measure counts the number of adjacent utterances that have no entities in common.

$$
\begin{array}{lccccc}
\text{MONITORING} & -X & -- & -- & -X & XO \\
\text{RADIO} & -- & -- & -S & -- & -- \\
\text{ALTITUDE} & -- & SO & OO & X- & -- \\
\text{CONTACT} & -- & -- & -S- & -- & - \\
\text{WITNESSES} & -- & -- & -- & -- & -S \\
\text{CLOUDS} & -O & OO & -- & -- & -S \\
\text{EVENING} & X- & -- & -- & -- & -- \\
\end{array}
\tag{5.3}
$$

Figure 5.1: Format of the entity-grid data. The entity is followed by its realization status in each of the 10 sentences. S means the entity is the subject, O is object, X is other, - is unrealized.

Rough-shift is also based on Centering Theory, and counts the number of rough-shift transitions (including NOCB transitions) that occur following the definition of rough-shift.

The data that was used for the experiments was processed by Barzilay and Lapata [2008] and is available from their website[5]. The entity extraction and grammatical function assignments are obtained from Barzilay and Lapata's data. For entity extraction they used Ng and Cardie's [2002] coreference resolution system. For grammatical function, a statistical parser [Collins, 1997] was used to determine the structure of each sentence, which was transformed into a dependency tree where grammatical relations were extracted. Figure 5.1 is an example of the data format used for representing entities in a text. In the example text there are 10 sentences, with 7 distinct entities realized in the entire text.

5.1.4   Results and Discussion

For the notation in Table 5.1, Barzilay & Lapata (no coref, no syntax, no salience) refers to a system that only tracks whether an entity exists in a sentence or not. Our Sentence Vectors implementation (linear + grammatical role) is similar, but the difference being that our approach does not require training. All coreference is provided by Ng and Cardie [Ng and Cardie, 2002]. No other coreference system was used on the earthquake dataset due to the poor performance of the coreference resolver on text input that is in all uppercase or all lowercase.

---

[5]http://people.csail.mit.edu/regina/coherence/

Sentence Ordering		
System	**Earthquakes**	**Accidents**
Barz & Lap (coref + grammatical role + freq)	87.2	**90.4**
Barz & Lap (no coref, no grammatical role, no freq)	81.4	86.0
HMM-based Content Model	**88.0**	75.8
Latent Semantic Analysis	81.0	87.3
Our Sentence Vectors		
Sentence Vectors (linear order)	**72.6**	85.6
Sentence Vectors (linear + weighted grammatical role)	71.4	87.65
Sentence Vectors (weighted grammatical role)	71.3	**88.0**
Sentence Vectors (grammatical role)	70.8	84.15
Sentence Vectors (linear + grammatical role)	70.8	85.5
Sentence Vectors (no linear, no gramm. role)	68.4	86.2
Centering Theory Transition Measures		
Rough-Shift (linear + grammatical role)	38.2	49.8
NOCB (linear + grammatical role)	**61.0**	**75.7**
NULL (linear + grammatical role)	54.1	72.2

Table 5.1: Experimental results of identifying the original text against a permuted variation of its sentences. The objective is to identify the original text, with the assumption that the original is more coherent than its permutation. All our models along with the Centering models utilized the coreference data from Barzilay and Lapata.

5.1.5 Discussion of Results (Earthquake Reports)

There were three main issues involving the measure of coherence within the earthquake reports: a need for lexical cohesion, improved coreference resolution, and poor formatting of the input data. We only provide one example of each (see appendix for a complete analysis).

An illustration of the need for lexical cohesion is shown in document apwsE941210.0275, where the original text was determined to be less coherent than 11 of the 20 permutations. Sentences 15 to 17 from the original text are shown as (5.4) to (5.6):

(5.4)   THE QUAKE BRIEFLY KNOCKED OUT ELECTRICITY TO SOME AREAS OF THE CAPITAL.

(5.5)   WINDOWS CRACKED AND BROKE IN SOME HIGH-RISE BUILDINGS AND FIRE TRUCKS CRUISED THE STREETS IN SEARCH OF POSSIBLE GAS LEAKS.

(5.6)   LARGE SECTIONS OF DOWNTOWN MEXICO CITY WERE DEVASTATED BY A 8.1 MAGNITUDE QUAKE IN 1985.

Sentences (5.5) and (5.6) break from the topic of (5.4) and return to describing the impact of the quakes on the city's buildings and utilities. This also illustrates the mereological relationships that need to be captured (e.g., 'streets', 'capital', 'city', 'building', 'windows', etc.) for a full model of coherence.

Document apwsE941210.0243 illustrates the need for improved coreference. Sentences 2 to 4 of the original text are shown in (5.7) to (5.9):

(5.7)   MEXICO CITY (AP) A STRONG EARTHQUAKE SHOOK CENTRAL MEXICO ON SATURDAY BUT THERE WERE NO IMMEDIATE REPORTS OF CASU-ALTIES.

(5.8)   THE QUAKE WAS FELT IN MEXICO CITY AMTROPOLIS OF 21 MILLION.

(5.9)   IT STARTED AROUND 10 20 A.M. AND LASTED MORE THAN A MINUTE.

Note that again 'earthquake' is not resolved to 'quake' throughout this text, and that 'central' was identified by the software as an entity in the 1st sentence but not in the 2nd. Other parts of the text identified as entities: 'BC-MEXICO-QUAKE', '21—MILLION', and

'MEXICO—EDS'. Between sentences (5.8) and (5.9), 'it' was not resolved to 'quake', thus resulting in a maximum coherence gap.

In the case where the input data was badly formatted, the final three sentences of document apwsE950116.0114 are shown in sentences (5.10) to (5.12):

(5.10) A QUAKE OF SIX CAN SHOOK BUILDINGS VIOLENTLY CRACK ASPHALT AND BURST UNDERGROUND PIPES.

(5.11) NO WARNING WAS ISSUED FOR TSUNAMI OR TIDAL WAVES CAUSED BY EARTHQUAKES.

(5.12) (KPH).

This document set includes an extra sentence in the original text that does not occur in its permutations. The last sentence consists solely of 'KPH', whereas the other permutations have it merged within one of its other sentences. Hence, the original text receives a penalty of 1.0 in its overall coherence score. If this sentence were removed, the original text would be identified as being more coherent than all of the other permutations. This error occurred for three other documents, resulting in 45 texts that would have been identified as more coherent if not for the additional sentence penalty. Taking into consideration these texts, our sentence vectors model (using linear order) would have scored 79.0% rather than the 72.6% reported in Table 5.1. In fairness, the original penalized score is left in Table 5.1 since we are not sure if the other systems (Barzilay and Lapata's, HMM-based model, and LSA model) are negatively affected by the extra sentence as well.

5.1.6 Discussion of Results (Aviation Accident Reports)

As with the earthquake reports, the two main issues with measuring the coherence of the aviation accident reports involved the need for incorporating lexical cohesion and improved coreference. Again, we only provide one example of each (see appendix for a complete analysis).

Document NTSB-3745 illustrates the need for lexical cohesion. Sentences 3 and 4 are excerpted in (5.13) to (5.14):

(5.13) AT THAT TIME THE AIRCRAFT WAS ABOUT 9 MILES SOUTHWEST OF BIMINI BAHAMAS ON A SOUTHEASTERLY HEADING AT A SPEED OF ABOUT 200 KNOTS AND AN ALTITUDE OF ABOUT 21,000 FEET.

(5.14) RESCUE PERSONNEL LOCATED DEBRIS AND FUEL AT LATITUDE 25 DE-
GREES 47.0 MINUTES NORTH LONGITUDE 079 DEGREES 23.6 MINUTES
WEST CONSISTENT WITH THE POSITION RELAYED BY ARTCC.

Between sentences 3 and 4 there is a mereological relationship between 'aircraft' and 'debris
and fuel' that is not captured. Also, throughout the text there is temporal and spatial
information regarding times and positions of the aircraft that do not get resolved to each
other (e.g., words such as 'feet', 'knots', 'minutes', 'position', 'latitude', 'longitude', 'degrees',
'altitude', 'radar information', etc. add more entities to our utterance vector thus increasing
the dissimilarity measure between utterances).

Document NTSB-3633 illustrates the need for improved coreference. Sentences 2 to 5
are excerpted in (5.15) to (5.18):

(5.15) THE PILOT ADJUSTED THE MIXTURE SETTING AND APPLIED CARBU-
RETOR HEAT BUT WAS UNSUCCESSFUL IN RESTORING FULL POWER.

(5.16) SHE ATTEMPTED TO DIVERT TO THE LIMON COLORADO AIRPORT 14
NAUTICAL MILES TO THE EAST BUT THE AIRCRAFT WAS UNABLE TO
MAINTAIN ALTITUDE.

(5.17) THE AIRPLANE IMPACTED THE GROUND BECAME AIRBORNE THEN HIT
THE GROUND AGAIN IN A ROUGH OPEN FIELD.

(5.18) THE AIRCRAFT THEN NOSED OVER.

Between (5.15) and (5.16) the pronoun 'she' has not been resolved to 'pilot'. In sentences
(5.16) and (5.17) 'aircraft' and 'airplane' are treated as two distinct entities when in fact they
should be resolved to the same entity. Between sentences (5.17) and (5.18), again 'aircraft'
and 'airplane' are treated as two distinct entities. Resolving these entities properly would
improve the document's coherence score significantly since they appear in highly salient
positions.

5.1.7 Summary

From the aviation accident reports and the earthquake reports, it can be seen that the
next necessary step for the vector model is the incorporation of lexical cohesion. Improved

coreference resolution systems also would help performance in detecting original sentence orderings. As can be seen in the appendices, the examples above are fairly representative of the cases where our coherence model incorrectly determined the original sentence ordering. There were even document permutation sets where the original text was not very coherent, and our model properly detected the more coherent orderings (document NTSB-3633 in the aviation reports).

As Barzilay and Lapata mention, the aviation accident report data tends to repeat the entities between sentences and has fewer referential entities, whereas the earthquake corpus has many lexical and referential gaps. As Barzilay and Lapata point out, 'the same area', 'the remote region', and 'site' all refer to 'Menglian county' but end up being treated as separate distinct entities. Thus, the aviation accident reports are closer to a text which has its referents resolved.

Promising in the results for both data sets is the performance of linear order alone compared to using linguistic information such as grammatical role. Remember that linear order does not use any syntactic information, it merely assigns a salience value to an entity based on which position it occupies in the text. Using linear order alone alleviates the need for a syntactic and dependency parser, thus simplifying the model considerably. One caveat is that the sentences in the sentence ordering texts were fairly simple sentences. For the most part they did not contain complex sentences or deep embeddings. This data set can be contrasted with the summarization experiment that follows, which contained much more complex sentences that realized more entities per sentence than the data within the sentence ordering experiment.

5.2 Experiment: Identifying Coherent Texts

A second experiment that was performed identifies the more coherent automated text summary, where the data used is from the Document Understanding Conference (DUC 2003). The task involves identifying a higher ranked summary as evaluated by human judges [Barzilay & Lapata, 2008]. The judges ranked the summaries for their coherence on a scale from 1 to 7, with 7 being the most coherent and 1 being the least coherent. In order to arrive at summaries with explicit coherence rankings, 16 document clusters were randomly selected with summaries from five systems along with summaries produced by humans, and were rated by 177 unpaid volunteers. There were approximately 23 ratings

per summary, with coherence scores between 1 and 7. For the annotation quality, Barzilay and Lapata found the inter-subject agreement was .768 ($p < .01$). Unlike the sentence ordering experiment, the sentences in a summary are extracted out of context and include problems such as dangling anaphors and thematically unrelated sentences. Thus coreference information was obtained from the individual summaries since there was no multi-document coreference tool available.

The model of Barzilay and Lapata is the same as the one described above in the sentence ordering experiment. Their model is trained (similar to the above sentence permutation experiment) on summary pairs of a single document, where one summary is more coherent than the other. They supplemented their training data with other DUC summaries that were performed by humans, with these human summaries assumed maximally coherent. Their objective was to have a model that is in high agreement with the assessments made by the human judges.

The measures *NOCB*, *Rough-shift*, and *Cheapness* are all identical as to how they are defined by Centering Theory (Section 3.2). These measures will be used to evaluate the performance of Centering Theory as a coherence measure.

The *Latent Semantic Analysis* model is also the same as described in the sentence ordering experiment. Barzilay and Lee's HMM topic model was not implemented due to lack of training data [Barzilay & Lapata, 2008].

The data consists of multi-document summaries produced by automatic summarization systems taken from the Document Understanding Conference (DUC 2003). Five different systems performed summaries of 16 documents sets, to give a total of 80 summaries. The summaries were evaluated by humans using a coherence rating that ranged from 1 to 7. Details of the human evaluation ranking process is described in Barzilay and Lapata [Barzilay & Lapata, 2008].

Barzilay & Lapata's system attained accuracy of 80% for a full-featured model (coreference resolution, syntax, and salience, not reported in Table 5.2), and 73.8% for a model without any features (also not reported in Table 5.2). Comparing to the previous state-of-the-art, Latent Semantic Analysis (LSA), their model outperformed LSA, which achieved 52.5%. The best performing model (83.8%) is Barzilay and Lapata's that only takes syntax and salience into account.

We did not implement a sentence vector model that used the grammatical role of the entity. The entity grid information was not provided by Barzilay and Lapata for a direct

Identifying Coherent Texts	
System	**Accuracy**
Barzilay & Lapata	**83.8**
Sentence Vectors	60.6
NOCB	55.63
Latent Semantic Analysis	52.5
Rough-Shift	47.5
NULL	36.9
Cheapness	30.0

Table 5.2: Results of the text coherence experiment used to identify the more coherent text. The texts were given rankings by humans ranging in score from 1 to 7. The objective was to identify the text given the higher ranking by the humans in a pairwise comparison.

comparison to be performed. The sentence vector model solely used the linear ordering of the entities to compute its salience within the utterance.

5.2.1 Results and Discussion

Before the output of the summarization systems is discussed, we provide a quick overview of the systems themselves. Five summarization systems were evaluated for their coherence. The principles behind each of the summarization systems that participated in DUC 2003 are taken from the respective system's papers describing their approaches:

System 6 [Guo and Stylios, 2003] use seven modules in their summarization system, which they refer to as follows: Content Reconstruction, Syntactic Parsing, Indices Extraction, Clustering Sentences, Cluster-Filtering, Cluster-Reduction and Size Control.

System 13 [Zhou and Hovy, 2003] their paper refers only to their participation on task 1 (short headline summaries) and does not mention how it was used on task 2. It is not clear how much of the same approach is reused since the paper describes approaches tailored for headline generation (which was a separate task in DUC 2003).

System 16 [Dunlavy et al., 2003] steps taken: preprocess the data for input to the system (consisting of term identification, part-of-speech (POS) tagging, sentence boundary detection and SGML DTD processing), apply single-document and multi-document summarization algorithms (the summarization algorithm uses a hidden Markov model

(HMM) to select sentences for a single-document extract summary; a pivoted QR algorithm is added to select from those to generate multi-document extracts), post-process the data for DUC evaluation (The following eliminations were made, when appropriate: sentences that begin with an imperative; sentences that contain a personal pronoun at or near the start; gerund clauses; restricted relative-clause appositives; intra-sentential attribution; lead adverbs)

System 18 [Angheluta et al., 2003] first summarized the individual single documents by using topic segmentation, and then summarized those summaries; non-redundant sentences were chosen using k-medoid clustering; sentence importance between viewpoint and sentence based on word overlap

System 26 (no paper available describing their approach, the system was known as Saarland.2003)

The average number of sentences in the summaries was 3.9 (high of 6, low of 2). The average sentence length of the summaries was 26.0 words. In the case of summaries with 2 sentences, we were unable to compute Centering Theory transitions due to 3 sentences being required for a transition. In those cases, the summaries were scored as being maximally incoherent.

Table 5.3 provides statistics of the human rankings for each document set along with the number of pairwise comparisons our system identified the more coherent document correctly.

The problems that affected the sentence ordering experiment, namely coreference resolution and the need for lexical cohesion, also played a factor in the evaluation of system summaries. The summaries also had fewer sentences than the earthquake and aviation accident reports, but had more entities per sentence. Thus, the sentence transitions within the summaries would typically have lower coherence scores due to having less entities in common between two adjacent sentences. We provide one example to illustrate such problems, with an analysis of other summaries provided in the appendix.

For document set 31001, System 18 produced the following summary:

1. JOHANNESBURG, South Africa _ The deal that South Africa's Truth and Reconciliation Commission offered was simple enough: Confess your crimes, apply for amnesty and you will go free.

2. He concedes that his is a lonely voice.

Doc Set	Mean Of Human Rank	Variance	# Identified Correct (out of 10)
30005	4.66	0.72	9
30028	4.72	**0.63**	**6**
30003	4.57	0.93	**6**
30034	4.33	**0.45**	**6**
30042	3.88	0.93	8
30040	3.78	1.16	8
30012	3.84	1.08	**6**
31001	3.81	**0.54**	**4**
31010	4.85	**0.66**	**6**
30056	4.09	**0.55**	**6**
31041	3.67	0.78	**6**
31050	3.75	**0.48**	**4**
31028	4.05	0.71	**5**
31033	3.9	**0.61**	**4**
31009	3.84	**0.56**	8
31038	3.87	0.77	**5**

Table 5.3: Analysis of the text coherence experiment data. Variance less than 0.7 is in bold (number was chosen arbitrarily), as well as the number of texts correctly ranked less than 7 times.

3. The disagreement stems from Thursday's release of the report by the Truth and Reconciliation Commission on horrors committed during apartheid.

4. The institution exploring apartheid's horrors will issue a report that finds the African National Congress shares blame for human rights violations as it struggled to overcome white rule.

Sentences 1 and 2 have a single shared entity due to 'his' being misresolved to 'Johannesburg'. Sentence 2 causes strong incoherence due to its dangling pronouns 'he' and 'his' ('he' was left unresolved). Sentence 2 also causes a maximally incoherent transition between it and sentence 3 for the same reasons. The relationship between 'apartheid' and 'white rule' was not captured, illustrating a need for not just lexical cohesion, but entailment and paraphrase detection as well.

5.3   Summary

In this chapter we have demonstrated the benefits of tracking multiple entities using an utterance vector model in comparison to Centering Theory's tracking of only the most salient entity. We also illustrate that the kind of coherence being measured by humans in the coherence experiment is not strictly one of entity coherence, which could account for some of the poor results from our model. Depending on the type of text and its style, entity coherence itself is insufficient for determining text coherence. A text that is rich in lexical relationships or the use of inferential knowledge would result in a low entity coherence score. A more complete coherence measure that accounts for all the different types of coherence (e.g., relational, lexical, inferential, etc.) is needed. Work on extending the entity-grid model to capture lexical relatedness was investigated by Fillipova and Strube [2004]. In their preliminary study they used WikiRelate! to compute a relatedness score between entities. Their initial results were found to be promising, though they found relatedness not to be as helpful as coreference resolution. Integrating relational coherence with entity and lexical coherence would first require a study of the discourse relations and whether patterns in the entity coherence can be correlated to the various types of relations (e.g., elaboration, contrast, etc.) [Knott et al., 2001].

Chapter 6

Conclusion

The main objective of this thesis involved applying Centering Theory to coherence, where it has its traditional and conventional use in anaphora resolution. Centering Theory is successful as a technique employed in explaining coreference resolution, but it is not quite as successful when applied to the analysis of global coherence. Thus, by investigating some of the assumptions of Centering Theory (tracking a single entity; using an utterance window of three consecutive utterances) and also investigating some of the philosophical questions in discourse (determining a consensus over the various definitions of *utterance*), Centering Theory can be improved in performance when applied to the measure and analysis of coherence in a text.

Centering Theory was extended by tracking all the entities of an utterance instead of only tracking a single entity. Tracking all entities is important in constructing a finer-grained metric that identifies the complete spectrum of entity coherence gaps that can occur between utterances. It also further distinguishes amongst the individual types of the coarse-grained transitions, such as a *continue* transition between utterances not sharing many entities in common in their C_f compared to the same *continue* transition where multiple entities in their C_f lists are shared. The experiments we have performed show the benefits to tracking multiple entities over utterances versus tracking a single entity. Centering Theory is thus strengthened for usage in coherence evaluation by extending it to track all entities within an utterance.

An extension to discourse provided in this thesis is the subsuming of the various definitions of utterance through the notion of *recency*. Rather than trying to form a consensus or list the advantages and disadvantages with respect to the different instantiations of what is

meant by an *utterance*, recency allows one to incorporate both the sentence and the clause as its unit of utterance, with their distinction being maintained in the way the sentence and its constituents are weighted within the context of the discourse. Recency is a natural extension to the idea of tracking multiple entities in a text by tracking all utterances within a text in a linguistically principled way. The principles being used involve dividing a sentence into its constituent clauses, and then applying weights to each individual clause that reflects its function or purpose within the sentence (e.g., main clauses receive higher weights than do subordinates). The recency model also allows for the idea of *functional zones* [Teufel, 2005] to receive various weights based on their function within a text and whether that function may cohere with the entities discussed.

6.1 Contributions

There are four main contributions in this thesis:

- constructing a method that tracks all entities in an approach based on Centering Theory

- a definition of utterance that integrates both the sentence and the clause

- a method that identifies and quantifies coherence gaps

- propose a method of bridging entity coherence gaps

These contributions are not only useful toward the study of coherence, but also to discourse. An implicit contribution is the illustration of a closer relationship between discourse and coherence through recency, which allows for an overlap between the notion of local coherence and global coherence. The implementation of recency could only occur in a model that tracks the entire history of utterances in a discourse, which is what our model allows.

The first contribution constructs a method to track multiple entities by proposing a vector representation of utterances. This vector representation allows one to construct a partial ordering between the sentence transitions, similar to the way Centering Theory's transitions are ordered in terms of preference. We show experimentally the benefits of tracking multiple entities in a study of the coherence of automatic summarizations and a sentence ordering task. Tracking the movement of all entities significantly improves performance compared to

Centering Theory's tracking of only the most-salient entities. Thus, tracking a single entity is insufficient for measuring coherence.

Another contribution involves minimizing the difficulty involved in deciding on the various units of utterance. Discussions in the discourse literature regarding the definition of utterance involve either using the sentence as the unit of utterance or using the clause (and not all clause types are typically considered). We alleviate such a problem by allowing the use of both the sentence and the clause in our definition of utterance. This is achieved through a proposed weighting scheme that allows one to maintain the hierarchy within sentences amongst its constituents, e.g., clauses in paratactic relationships would receive equal weighting, whereas clauses in a hypotactic relationship would receive different weightings, with the main clause receiving greater weight than its subordinates. Similar to how a sentence is a composition of clauses, the weighting allows us to construct larger utterances (the sentence) from smaller ones (the clause), preserving both utterance units while simultaneously treating them as a variation of a single type of utterance.

We also formulate within this thesis a method to identify entity coherence gaps using the notion of *coherence trees*, whose structure is similar in principle to discourse trees. Here the hierarchy is between the number of entities realized in each sentence. Sentences with a greater number of entities are parents to sentences containing a lesser number of entities. Candidates for bridging are sentences that have entities in common with both sentences that contain the coherence gap. The premise is based on Centering Theory, with the objective being to replace a less-preferred transition to a more-preferred one.

The coherence trees proposed in this thesis are not only of utility for determining where entity coherence gaps exist, but are useful for proposing candidates for bridging such gaps. Candidates for such bridges are taken from the tree, where the lowest common subsumer of the two sentences containing the gap becomes a candidate for bridging.

The approach to coherence taken in this thesis minimizes Centering measures of NOCB to where they truly should occur: when one is introducing an *entirely* new set of topics, or when one has produced a non-sequitur. By definition, a NOCB is when there is *no* backward-looking center, i.e., the entities in the current utterance have never been previously mentioned. In Centering Theory, this *previousness* consists only of the previous utterance, whereas we extend the *previousness* to the beginning of the text by way of recency. Hence, a NOCB transition retains its intended usage as an indicator of a very strong violation of coherence.

6.2 Future Work

As illustrated from the sentence ordering and summary coherence experiments, the two most useful and immediate improvements to entity coherence involve improving coreference resolution and incorporating lexical cohesion within entity coherence. Whereas coreference resolution can be treated as a separate problem from entity coherence, lexical coherence is tightly related to entity coherence. The main difficulty lies in merging lexical similarity measures within Centering's C_f list (which is essentially the utterance vector model). Fais [2004] discusses the problems and difficulties involved in merging lexical cohesion and inferables within Centering.

The coherence tree model proposed here only considers nouns or noun phrases as entities. The impact verbs have on coherence should be explored in a similar way to how entities have been investigated[1]. Incorporating verbs into the entity model is not a trivial extension. The individual verbs would still maintain their orthogonality to nouns, but there would need to be a metric introduced that looks at the lexical semantic similarity between verbs and nouns. Currently there is no metric that considers verbs and nouns (as well as other parts of speech) as existing in the same lexical semantic space. Current word similarity/relatedness measures only discuss nouns within a restricted *noun-space*. Also, lexical similarity would fall under an analysis involving lexical cohesion, and not strictly entity coherence. Hence, in order to incorporate lexical relatedness, the current implementation of coherence trees would have to be extended beyond pure entity coherence and incorporate lexical cohesion, which is an extension of the problem of modelling lexical cohesion within utterance vectors since the vectors are what form the nodes of the coherence tree.

This study did not incorporate the effect of pronominalization on salience (e.g., Givenness Hierarchy). This would study the effects on salience when entities are realized as referents. The principle here would be similar to ideas proposed in Centering Theory, where results show that pronouns typically realize the most salient entities in an utterance. For the purposes of evaluating coherence, the principle is reversed such that if an entity is realized as a pronoun, then its salience within the utterance is increased by a certain factor over non-pronominalized entities.

One of the difficulties with using clauses as the unit of utterance is that one must have available a software that identifies and classifies the kind of subordination or embeddings.

---

[1]Some theories of grammar consider the verbs as the locus of information in sentences or clauses.

We are currently unaware whether such studies or software exist, although Marcu [2000] has looked at identifying which clauses in complex sentences are the main or subordinate clauses.

The current study of discourse and its aspects is heavily fragmented. Zellig [1952] recognized the importance of sentence ordering, lexical chains, coreference resolution, lexical similarity, paraphrase detection, textual entailment, and discourse segmentation for discourse analysis. Other studies could involve using coherence to assist in improving the quality of machine translations, aid in natural language generation systems, and also automatic summarization systems. Coherence (and discourse analysis) is the study of the interaction and interplay of all the listed aspects. Thus, a framework which treats these facets of a discourse in a more holistic approach, as opposed to the fragmented piecewise approach proposed in our study and also by Coh-Metrix, is necessary for a system that can completely handle coherence and discourse analysis. Thus, a framework which would incorporate all linguistic factors would be necessary for a full study of coherence. A model of coherence that *only* incorporates entities is shortsighted if there is no intention to incorporate inference, lexical semantics, and temporal coherence.

Appendix A

Sample of Earthquake Reports

In apwsE950121.0073 we only correctly detected 8 of 17 permutations as less coherent than the original text. The original text is:

1. BC-INDONESIA-QUAKE—STRONG EARTHQUAKE JOLTS SUMBA ISLAND—EDS ONE NAME CORRECT FOR SARIMAN.

2. JAKARTA INDONESIA (AP) A STRONG EARTHQUAKE WITH A PRELIMINARY MAGNITUDE OF 5.1 ROCKED PART OF SUMBA ISLAND IN EASTERN INDONESIA THE METEOROLOGY AND GEOPHYSICS AGENCY SAID SATURDAY.

3. THERE WERE NO IMMEDIATE REPORTS OF DAMAGE OR CASUALTIES.

4. THE QUAKE WHICH STRUCK AT 4 11 AM LOCAL TIME SATURDAY JOLTED THE TOWN OF WAIPANGU IN SUMBA ISLAND 1,550 KILOMETERS (970 MILES) EAST OF JAKARTA SAID SARIMAN AN AGENCY SPOKESMAN.

5. HE LOCATED THE QUAKE 'S EPICENTER IN THE SAWU SEA BETWEEN SUMBA AND TIMOR ISLAND AT A DEPTH OF 61 KILOMETERS (38 MILES).

The entity grid for the original text is:
METEOROLOGY - S - - -
SATURDAY - X - X -
KILOMETERS - - - X X
JOLTS X - - - -

AP - X - - -
JAKARTA - S - X -
SUMBA X X - X X
TIME - - - O -
NAME X - - - -
TOWN - - - O -
INDONESIA - S - - -
REPORTS - - O - -
ISLAND - X - X X
EARTHQUAKE X X - - -
AGENCY - S - X -
MILES - - - X X
QUAKE - - - S X
HE - - - - S
GEOPHYSICS - S - - -
EPICENTER - - - - O
CASUALTIES - - X - -
MAGNITUDE - X - - -
TIMOR - - - - X
WAIPANGU - - - X -
SARIMAN X - - O -
DEPTH - - - - X
SAWU - - - - X
PART - O - - -
SPOKESMAN - - - O -
ISLAND—EDS X - - - -
DAMAGE - - X - -
SEA - - - - X

Between sentences 2 and 3 there is a maximum coherence gap. Sentence 3 elaborates on the effects of the earthquake mentioned in sentence 2, thus there being an inferred mention of the earthquake (e.g., 'no immediate reports of damage or casualties due to the earthquake'). Between sentences 3 and 4 there is yet again a maximum coherence gap due to a return to the specifics about the quake. Also note that 'quake' and 'earthquake' are not resolved

to each other and thus are treated as two distinct entities. Between sentences 4 and 5, 'he' is not resolved to 'Sariman'. This last sentence pair scores fairly well. Note that any permutation that has sentence 3 either the first or last sentence would score better than the original text.

Permutation 16 was ranked most coherent in apwsE950121.0073:

1. THERE WERE NO IMMEDIATE REPORTS OF DAMAGE OR CASUALTIES.

2. JAKARTA INDONESIA (AP) A STRONG EARTHQUAKE WITH A PRELIMINARY MAGNITUDE OF 5.1 ROCKED PART OF SUMBA ISLAND IN EASTERN INDONESIA THE METEOROLOGY AND GEOPHYSICS AGENCY SAID SATURDAY.

3. BC-INDONESIA-QUAKE—STRONG EARTHQUAKE JOLTS SUMBA ISLAND—EDS ONE NAME CORRECT FOR SARIMAN.

4. THE QUAKE WHICH STRUCK AT 4 11 AM LOCAL TIME SATURDAY JOLTED THE TOWN OF WAIPANGU IN SUMBA ISLAND 1,550 KILOMETERS (970 MILES) EAST OF JAKARTA SAID SARIMAN AN AGENCY SPOKESMAN.

5. HE LOCATED THE QUAKE 'S EPICENTER IN THE SAWU SEA BETWEEN SUMBA AND TIMOR ISLAND AT A DEPTH OF 61 KILOMETERS (38 MILES).

The entity grid for permutation 16 is:

```
METEOROLOGY - S - - -
SATURDAY - X - X -
KILOMETERS - - - X X
JOLTS - - X - -
AP - X - - -
JAKARTA - S - X -
SUMBA - X X X X
TIME - - - O -
NAME - - X - -
TOWN - - - O -
INDONESIA - S - - -
REPORTS O - - - -
```

ISLAND - X - X X

EARTHQUAKE - X X - -

AGENCY - S - X -

MILES - - - X X

QUAKE - - - S X

HE - - - - S

GEOPHYSICS - S - - -

EPICENTER - - - - O

CASUALTIES X - - - -

MAGNITUDE - X - - -

TIMOR - - - - X

WAIPANGU - - - X -

SARIMAN - - X O -

DEPTH - - - - X

SAWU - - - - X

PART - O - - -

SPOKESMAN - - - O -

DAMAGE X - - - -

ISLAND—EDS - - X - -

SEA - - - - X

In apwsE941210.0243 we only correctly detected 9 of 21 permutations as less coherent than the original text. The original text is:

1. BC-MEXICO-QUAKE 2ND LD-WRITETHRU,0185—STRONG QUAKE SHAKES CENTRAL MEXICO—EDS UPDATES WITH DETAIL FIXES MEXICO CITY POPULATION TO 21—MILLION (STED 18 MILLION).

2. MEXICO CITY (AP) A STRONG EARTHQUAKE SHOOK CENTRAL MEXICO ON SATURDAY BUT THERE WERE NO IMMEDIATE REPORTS OF CASUAL-TIES.

3. THE QUAKE WAS FELT IN MEXICO CITY AMTROPOLIS OF 21 MILLION.

4. IT STARTED AROUND 10 20 A.M. AND LASTED MORE THAN A MINUTE.

5. BUILDINGS ALONG REFORMA AVENUE THE MAIN EAST-WEST THOROUGH-FARE COULD BE SEEN SWAYING WILDLY.

6. THERE WAS NO IMMEDIATE READING ON MAGNITUDE.

7. MANY TRAFFIC LIGHTS QUICK WORKING ELECTRIC POWER WENT OUT MOMENTARILY IN SOME SECTIONS OF THE CITY AND FIREFIGHTING UNITS CRUISED THE STREETS SEARCHING FOR NATURAL GAS LEAKS.

8. THE QUAKE BROKE WINDOWS IN SOME HIGH-RISE BUILDINGS.

9. JOHN MINSCH OF THE U.S. NATIONAL EARTHQUAKE CENTER IN GOLDEN COLO. SAID THE CENTER WAS AWARE OF THE QUAKE IN MEXICO BUT HAS N'T BEEN ABLE TO GET ANY FIX ON THE LOCATION OR MAGNITUDE YET.

10. LARGE SECTIONS OF MEXICO CITY 'S DOWNTOWN AREAS WERE DEVAS-TATED BY A FIERCE QUAKE IN 1985.

11. AT LEAST 9,500 PEOPLE WERE KILLED.

The entity grid for the original text is:

UNITS - - - - - - S - - - -
MEXICO O S X - - - - - X X -
SATURDAY - X - - - - - - - - -
IT - - - X - - - - - - -
JOHN - - - - - - - - S - -
CENTER - - - - - - - - X - -
TRAFFIC - - - - - - X - - - -
READING - - - - - O - - - - -
AP - X - - - - - - - - -
MINUTE - - - X - - - - - - -
BUILDINGS - - - - S - - X - - -
SECTIONS - - - - - - X - - S -
AMTROPOLIS - - X - - - - - - - -
REPORTS - O - - - - - - - - -
EARTHQUAKE - S - - - - - - X - -

EAST-WEST - - - - X - - - - - -

NATIONAL - - - - - - - - X - -

QUAKE S - S - - - - S X X -

THOROUGHFARE - - - - X - - - - - -

POWER - - - - - - S - - - -

CENTRAL O - - - - - - - - - -

MINSCH - - - - - - - - S - -

AREAS - - - - - - - - - X -

POPULATION O - - - - - - - - - -

U.S. - - - - - - - - X - -

GOLDEN - - - - - - - - X - -

MEXICO—EDS O - - - - - - - - - -

21—MILLION X - - - - - - - - - -

UPDATES O - - - - - - - - - -

CITY O S X - - - X - - X -

CASUALTIES - X - - - - - - - - -

MAGNITUDE - - - - - X - - X - -

LOCATION - - - - - - - - X - -

LIGHTS - - - - - - X - - - -

DETAIL X - - - - - - - - - -

PEOPLE - - - - - - - - - - S

COLO. - - - - - - - - - X - -

GAS - - - - - - X - - - -

FIX - - - - - - - - O - -

LEAKS - - - - - - X - - - -

AVENUE - - - - X - - - - - -

STREETS - - - - - - O - - - -

BC-MEXICO-QUAKE S - - - - - - - - - -

A.M. - - - O - - - - - - -

REFORMA - - - - X - - - - - -

WINDOWS - - - - - - - O - - -

Between sentences 1 and 2, there are only two entities in common: 'Mexico' and 'city'. Note that again 'earthquake' is not resolved to 'quake' throughout this text, and that

'central' was identified as an entity in the 1st sentence but not in the 2nd. Other parts of the text identified as entities: 'BC-MEXICO-QUAKE', '21—MILLION', and 'MEXICO—EDS'. Between sentences 3 and 4, 'it' was not resolved to 'quake', thus resulting in a maximum coherence gap. Sentence 5 discusses the impact of the earthquake on the city. Sentence 6 discusses the strength of the earthquake. Sentences 7 and 8 returns to discussing the impact of the earthquake on the city. Sentence 9 returns to discussing the strength and location of the earthquake. Sentence 10 returns to discuss the impact of the earthquake on the city. Sentence 11 should be written as a subordinate clause to sentence 10, e.g., 'a fierce quake in 1985 where at least 9500 people were killed'. One can observe that a permutation of the text can lead to stronger coherence than the original text. Also necessary is lexical cohesion (e.g., 'downtown area's' and 'sections of the city') in order to capture the relationships fully.

Permutation 7 was ranked most coherent in E941210.0243:

1. MEXICO CITY (AP) A STRONG EARTHQUAKE SHOOK CENTRAL MEXICO ON SATURDAY BUT THERE WERE NO IMMEDIATE REPORTS OF CASUAL-TIES.

2. LARGE SECTIONS OF MEXICO CITY 'S DOWNTOWN AREAS WERE DEVAS-TATED BY A FIERCE QUAKE IN 1985.

3. MANY TRAFFIC LIGHTS QUICK WORKING ELECTRIC POWER WENT OUT MOMENTARILY IN SOME SECTIONS OF THE CITY AND FIREFIGHTING UNITS CRUISED THE STREETS SEARCHING FOR NATURAL GAS LEAKS.

4. THE QUAKE WAS FELT IN MEXICO CITY AMTROPOLIS OF 21 MILLION.

5. THE QUAKE BROKE WINDOWS IN SOME HIGH-RISE BUILDINGS.

6. BUILDINGS ALONG REFORMA AVENUE THE MAIN EAST-WEST THOROUGH-FARE COULD BE SEEN SWAYING WILDLY.

7. JOHN MINSCH OF THE U.S. NATIONAL EARTHQUAKE CENTER IN GOLDEN COLO. SAID THE CENTER WAS AWARE OF THE QUAKE IN MEXICO BUT HAS N'T BEEN ABLE TO GET ANY FIX ON THE LOCATION OR MAGNITUDE YET.

8. BC-MEXICO-QUAKE 2ND LD-WRITETHRU,0185—STRONG QUAKE SHAKES CENTRAL MEXICO—EDS UPDATES WITH DETAIL FIXES MEXICO CITY POPULATION TO 21—MILLION (STED 18 MILLION).

9. AT LEAST 9,500 PEOPLE WERE KILLED.

10. IT STARTED AROUND 10 20 A.M. AND LASTED MORE THAN A MINUTE.

11. THERE WAS NO IMMEDIATE READING ON MAGNITUDE.

The entity grid for permutation 7 is:

```
UNITS - - S - - - - - - - -
MEXICO S X - X - - X O - - -
SATURDAY X - - - - - - - - - -
IT - - - - - - - - - - X -
JOHN - - - - - - S - - - -
CENTER - - - - - - X - - - -
TRAFFIC - - X - - - - - - - -
SECTIONS - S X - - - - - - - -
AP X - - - - - - - - - -
BUILDINGS - - - - X S - - - - -
MINUTE - - - - - - - - - X -
READING - - - - - - - - - - O
AMTROPOLIS - - - X - - - - - - -
REPORTS O - - - - - - - - - -
EARTHQUAKE S - - - - - X - - - -
EAST-WEST - - - - - X - - - - -
NATIONAL - - - - - - X - - - -
QUAKE - X - S S - X S - - -
POWER - - S - - - - - - - -
THOROUGHFARE - - - - - X - - - - -
MINSCH - - - - - - S - - - -
AREAS - X - - - - - - - - -
CENTRAL - - - - - - - O - - -
POPULATION - - - - - - - O - - -
```

```
U.S. - - - - - - X - - - -
GOLDEN - - - - - - X - - - -
CASUALTIES X - - - - - - - - - -
MAGNITUDE - - - - - - X - - - X
MEXICO—EDS - - - - - - - O - - -
CITY S X X X - - - O - - -
LOCATION - - - - - - X - - - -
UPDATES - - - - - - - O - - -
21—MILLION - - - - - - - X - - -
LIGHTS - - X - - - - - - - -
DETAIL - - - - - - - X - - -
PEOPLE - - - - - - - - - S - -
COLO. - - - - - - X - - - -
GAS - - X - - - - - - - -
FIX - - - - - - O - - - -
LEAKS - - X - - - - - - - -
AVENUE - - - - - X - - - - -
STREETS - - O - - - - - - - -
BC-MEXICO-QUAKE - - - - - - - S - - -
A.M. - - - - - - - - - - O -
REFORMA - - - - - X - - - - -
WINDOWS - - - - O - - - - - -
```

In apwsE941210.0275 we only correctly detected 9 of 21 permutations as less coherent than the original text. The original text is:

1. BC-MEXICO-QUAKE 4TH LD-WRITETHRU,0380—STRONG QUAKE SHAKES CENTRAL MEXICO—EDS RETRANSMITTING TO FIX WRITETHRU SEQUENCE UPDATES WITH NEW—DETAIL QUOTES AND COLOR ADDS.

2. QUAKE SHAKES CENTRAL MEXICO—BY ISAAC LEVI—ASSOCIATED PRESS WRITER.

3. MEXICO CITY (AP) A STRONG EARTHQUAKE SHOOK CENTRAL MEXICO SATURDAY SENDING PANICKED TOURISTS RUNNING FROM AN AIRPORT TERMINAL AND SHAKING BUILDINGS IN THE CAPITAL.

4. THERE WERE NO IMMEDIATE REPORTS OF SERIOUS INJURIES.

5. THE QUAKE HAD A PRELIMINARY MAGNITUDE OF 6.3 AND IT EPICENTER WAS IN GUERRERO STATE 290 KILOMETERS (165 MILES) SOUTHWEST OF MEXICO CITY SAID RUSS NEEDHAM OF THE U.S. GEOLOGICAL SURVEY 'S EARTHQUAKE INFORMATION CENTER IN GOLDEN COLO.

6. PART OF THE ROOF OF AN AIRPORT TERMINAL IN THE BEACH RESORT OF ZAHUATANEJO COLLAPSED AND ITS WINDOWS SHATTERED SENDING SCORES OF TOURISTS RUNNING OUTSIDE.

7. POWER AND TELEPHONE SERVICE WERE BRIEFLY INTERRUPTED IN THE TOWN ABOUT 340 KILOMETERS (200 MILES) SOUTHWEST OF MEXICO CITY.

8. A FENCE WAS TOPPLED IN A POOR NEIGHBORHOOD IN ZIHUATANEJO.

9. THE RED CROSS SAID AT LEAST 10 PEOPLE SUFFERED FROM NERVOUS DISORDERS CAUSED BY THE QUAKE.

10. THE QUAKE STARTED AROUND 10 20 AM AND WAS FELT FOR MORE THAN A MINUTE IN MEXICO CITY A METROPOLIS OF ABOUT 21 MILLION PEOPLE.

11. BUILDINGS ALONG REFORMA AVENUE THE MAIN EAST-WEST THOROUGHFARE SHOOK WILDLY.

12. I WAS SO SCARED.

13. EVERYTHING JUST BEGAN SHAKING SAID SONIA ARIZPE A MEXICO CITY STREET VENDOR WHOSE ALUMINUM CART STARTED ROLLING AWAY DURING THE TEMBLOR.

14. BUT FRANCISCO LOPEZ A VISITING LOS ANGELES BUSINESSMAN SAID IT COULD HAVE BEEN MUCH WORSE.

15. I 'VE BEEN THROUGH PLENTY OF QUAKES IN L.A. AND THIS WAS NO BIG DEAL.

16. THE QUAKE BRIEFLY KNOCKED OUT ELECTRICITY TO SOME AREAS OF THE CAPITAL.

17. WINDOWS CRACKED AND BROKE IN SOME HIGH-RISE BUILDINGS AND FIRE TRUCKS CRUISED THE STREETS IN SEARCH OF POSSIBLE GAS LEAKS.

18. LARGE SECTIONS OF DOWNTOWN MEXICO CITY WERE DEVASTATED BY A 8.1 MAGNITUDE QUAKE IN 1985.

19. AT LEAST 9,500 PEOPLE WERE KILLED.

20. MEXICO IS PRONE TO FREQUENT QUAKES BECAUSE OF TWO MAJOR GEOGRAPHIC FAULTS OFFSHORE IN THE PACIFIC OCEAN.

21. ALSO SATURDAY AN EARTHQUAKE OF MAGNITUDE 6 HIT NORTHEASTERN COLOMBIA.

22. THERE WERE NO REPORTED CASUALTIES.

23. THE QUAKE SHOOK BUILDINGS IN THE CITY OF BUCARAMANGA 300 KILOMETERS (185 MILES) NORTHEAST OF BOGOTA.

The entity grid for the original text is:

```
LOPEZ - - - - - - - - - - - - - - X - - - - - - - - -
CASUALTIES - - - - - - - - - - - - - - - - - - - - - O -
COLOMBIA - - - - - - - - - - - - - - - - - - - - X - -
NORTHEAST - - - - - - - - - - - - - - - - - - - - - X
PART - - - - - S - - - - - - - - - - - - - - -
UPDATES X - - - - - - - - - - - - - - - - - - - - -
QUAKES - - - - - - - - - - - - - - X - - - - X - - -
BOGOTA - - - - - - - - - - - - - - - - - - - - - X
POWER - - - - - - X - - - - - - - - - - - - - -
COLOR X - - - - - - - - - - - - - - - - - - - - -
WRITER - X - - - - - - - - - - - - - - - - - - - -
THOROUGHFARE - - - - - - - - - - X - - - - - - - - - - - -
BUSINESSMAN - - - - - - - - - - - - - X - - - - - - - - -
TOWN - - - - - - X - - - - - - - - - - - - - -
```

METROPOLIS - - - - - - - - - X - - - - - - - - - - - - -
BRIEFLY - - - - - - S - - - - - - - - - - - - - - -
RUSS - - - - O - - - - - - - - - - - - - - - -
ZAHUATANEJO - - - - - X - - - - - - - - - - - - - - -
BUCARAMANGA - X
DEAL - - - - - - - - - - - - - - O - - - - - - - -
ADDS O -
LEVI—ASSOCIATED - X - - - - - - - - - - - - - - - - - - -
RESORT - - - - - X - - - - - - - - - - - - - -
ALUMINUM - - - - - - - - - - - - X - - - - - - - - -
SURVEY - - - - X - - - - - - - - - - - - - - -
ZIHUATANEJO - - - - - - - X - - - - - - - - - - - -
BC-MEXICO-QUAKE S -
QUAKE X X - - S - - - X X - - - - - S - X - - - - S
MEXICO—BY - X - - - - - - - - - - - - - - - - - - -
L.A. - - - - - - - - - - - - - - X - - - - - - - -
ELECTRICITY - - - - - - - - - - - - - - - O - - - - - - -
MINUTE - - - - - - - - - - X - - - - - - - - - -
EPICENTER - - - - S - - - - - - - - - - - - - - - -
STREET - - - - - - - - - - - - - X - - - - - - - -
SECTIONS - - - - - - - - - - - - - - - - - - S - - - - -
AIRPORT - - X - - X - - - - - - - - - - - - - - -
BUILDINGS - - O - - - - - - - S - - - - - X - - - - - O
CENTRAL O X -
IT - - - - S - - - - - - - - - S - - - - - - - - -
BEACH - - - - - X - - - - - - - - - - - - - - -
GEOLOGICAL - - - - X - - - - - - - - - - - - - - - -
RED - - - - - - - - S - - - - - - - - - - - -
CITY - - S - X - X - - X - - X - - - - X - - - - X
PLENTY - - - - - - - - - - - - - - X - - - - - - - -
SEQUENCE O -
KILOMETERS - - - - X - O - - - - - - - - - - - - - - O
EAST-WEST - - - - - - - - - - X - - - - - - - - - - -

```
NEIGHBORHOOD - - - - - - - X - - - - - - - - - - - - - -
TEMBLOR - - - - - - - - - - - - X - - - - - - - - - -
HIT - - - - - - - - - - - - - - - - - - - X - -
SATURDAY - - X - - - - - - - - - - - - - - - - - X - -
AREAS - - - - - - - - - - - - - - X - - - - - - -
U.S. - - - - X - - - - - - - - - - - - - - - -
SCORES - - - - - S - - - - - - - - - - - - - -
LOS - - - - - - - - - - - - - X - - - - - - - -
CROSS - - - - - - - - S - - - - - - - - - - - -
PRESS - X - - - - - - - - - - - - - - - - - - -
SEARCH - - - - - - - - - - - - - - - X - - - - - -
TERMINAL - - X - - X - - - - - - - - - - - - - - -
SERVICE - - - - - - X - - - - - - - - - - - - -
GOLDEN - - - - X - - - - - - - - - - - - - - -
DOWNTOWN - - - - - - - - - - - - - - - - X - - - - -
SONIA - - - - - - - - - - - - O - - - - - - - - -
TOURISTS - - S - - X - - - - - - - - - - - - - - -
TRUCKS - - - - - - - - - - - - - - - - S - - - - - -
FIRE - - - - - - - - - - - - - - - - S - - - - - -
DISORDERS - - - - - - - - X - - - - - - - - - - - -
AVENUE - - - - - - - - - - X - - - - - - - - - - -
CART - - - - - - - - - - - - X - - - - - - - - -
PACIFIC - - - - - - - - - - - - - - - - - - X - - -
MEXICO - - S - X - X - - X - - X - - - - X - S - - -
COLO - - - - X - - - - - - - - - - - - - - - -
ISAAC - X - - - - - - - - - - - - - - - - - - - -
FRANCISCO - - - - - - - - - - - - - - X - - - - - - - -
AP - - X - - - - - - - - - - - - - - - - - - - -
REPORTS - - - O - - - - - - - - - - - - - - - - -
ROOF - - - - - X - - - - - - - - - - - - - - -
GAS - - - - - - - - - - - - - - - - X - - - - - -
MAGNITUDE - - - - O - - - - - - - - - - - X - - X - -
EVERYTHING - - - - - - - - - - - - S - - - - - - - - -
```

```
SHAKES - X - - - - - - - - - - - - - - - - - - - -
WINDOWS - - - - - S - - - - - - - - - - - X - - - - - -
I - - - - - - - - - - - S - - S - - - - - - - - -
INJURIES - - - X - - - - - - - - - - - - - - - -
ARIZPE - - - - - - - - - - - - O - - - - - - - - - -
MILES - - - - X - X - - - - - - - - - - - - - X
ANGELES - - - - - - - - - - - - - X - - - - - - - - -
FAULTS - - - - - - - - - - - - - - - - - - X - - -
MEXICO—EDS O - - - - - - - - - - - - - - - - - - - - -
FENCE - - - - - - - S - - - - - - - - - - - - - -
LEAKS - - - - - - - - - - - - - - - - X - - - - - -
OCEAN - - - - - - - - - - - - - - - - - - - X - - -
CAPITAL - - X - - - - - - - - - - - X - - - - - - -
INFORMATION - - - - X - - - - - - - - - - - - - - - -
VENDOR - - - - - - - - - - - - O - - - - - - - - -
GUERRERO - - - - X - - - - - - - - - - - - - - -
TELEPHONE - - - - - - X - - - - - - - - - - - - -
QUOTES X - - - - - - - - - - - - - - - - - - - -
EARTHQUAKE - - S - X - - - - - - - - - - - - - X - -
WRITETHRU O - - - - - - - - - - - - - - - - - - - - -
STATE - - - - X - - - - - - - - - - - - - - -
STREETS - - - - - - - - - - - - - - - - O - - - - - -
PEOPLE - - - - - - - - S X - - - - - - - - S - - - -
CENTER - - - - X - - - - - - - - - - - - - - -
NEEDHAM - - - - O - - - - - - - - - - - - - -
REFORMA - - - - - - - - - X - - - - - - - - - - - -
```

This was one of the longer texts used in the sentence ordering evaluation, with many distinct entities being introduced. Sentence 1 seems to be an artifact of the newspaper byline, with sentence 2 being the main headline. The only entity in common between the two is 'quake', with the other 16 entities not shared. Sentences 2 and 3 do not have any entities in common, with 'quake' and 'earthquake' not being resolved to each other. Sentence 4 discusses reports of injuries, with sentence 5 returning to a description of the earthquake. Sentences 6 through 8 discuss the impact of the earthquake on the city of

'Zahuatanejo', with the phrase 'the town about 340 kilometers southwest of Mexico City' in sentence 7 not being resolved to 'Zahuatanejo'. Sentence 9 discusses the impact of the quake on the civilians. Sentence 11 returns to discussing the impact of the quake on the city. Sentence 12 begins a quote, where 'I' is not resolved to 'Sonia' in sentence 13. Sentence 14 introduces another person, with them being quoted in sentence 13, where 'I' is not resolved to 'Fransisco', nor is 'L.A.' resolved to 'Los Angeles'. Between sentences 14 and 15, 'it' is not resolved to 'quakes'. Between sentences 15 and 16, 'quake' is not resolved to 'quakes'. Sentences 16 and 17 return to describing the impact of the quakes on the city's buildings and utilities. Sentence 18 discusses an earthquake from 1985, where again sentence 19 should be written as a subordinate to 18. Sentence 20 discusses the cause of Mexico's earthquakes, with sentences 21 through 23 discussing an earthquake in Colombia. This document shows the import of discourse purposes between sentences (e.g., comparing the current earthquake to earthquakes in the past or in other countries), as well as journalistic conventions (e.g., breaks in the text to introduce points of view from citizens and how they are impacted). There were also a few mereological relationships that needed to be captured as well (e.g., 'streets', 'capital', 'city', 'building', 'windows', etc.).

Permutation 3 was ranked most coherent in E941210.0275:

1. PART OF THE ROOF OF AN AIRPORT TERMINAL IN THE BEACH RESORT OF ZAHUATANEJO COLLAPSED AND ITS WINDOWS SHATTERED SENDING SCORES OF TOURISTS RUNNING OUTSIDE.

2. I 'VE BEEN THROUGH PLENTY OF QUAKES IN L.A. AND THIS WAS NO BIG DEAL.

3. THERE WERE NO REPORTED CASUALTIES.

4. THE RED CROSS SAID AT LEAST 10 PEOPLE SUFFERED FROM NERVOUS DISORDERS CAUSED BY THE QUAKE.

5. LARGE SECTIONS OF DOWNTOWN MEXICO CITY WERE DEVASTATED BY A 8.1 MAGNITUDE QUAKE IN 1985.

6. EVERYTHING JUST BEGAN SHAKING SAID SONIA ARIZPE A MEXICO CITY STREET VENDOR WHOSE ALUMINUM CART STARTED ROLLING AWAY DURING THE TEMBLOR.

7. MEXICO CITY (AP) A STRONG EARTHQUAKE SHOOK CENTRAL MEXICO SATURDAY SENDING PANICKED TOURISTS RUNNING FROM AN AIRPORT TERMINAL AND SHAKING BUILDINGS IN THE CAPITAL.

8. AT LEAST 9,500 PEOPLE WERE KILLED.

9. WINDOWS CRACKED AND BROKE IN SOME HIGH-RISE BUILDINGS AND FIRE TRUCKS CRUISED THE STREETS IN SEARCH OF POSSIBLE GAS LEAKS.

10. POWER AND TELEPHONE SERVICE WERE BRIEFLY INTERRUPTED IN THE TOWN ABOUT 340 KILOMETERS (200 MILES) SOUTHWEST OF MEXICO CITY.

11. THE QUAKE SHOOK BUILDINGS IN THE CITY OF BUCARAMANGA 300 KILOMETERS (185 MILES) NORTHEAST OF BOGOTA.

12. BUILDINGS ALONG REFORMA AVENUE THE MAIN EAST-WEST THOROUGHFARE SHOOK WILDLY.

13. THERE WERE NO IMMEDIATE REPORTS OF SERIOUS INJURIES.

14. BUT FRANCISCO LOPEZ A VISITING LOS ANGELES BUSINESSMAN SAID IT COULD HAVE BEEN MUCH WORSE.

15. ALSO SATURDAY AN EARTHQUAKE OF MAGNITUDE 6 HIT NORTHEASTERN COLOMBIA.

16. BC-MEXICO-QUAKE 4TH LD-WRITETHRU,0380—STRONG QUAKE SHAKES CENTRAL MEXICO—EDS RETRANSMITTING TO FIX WRITETHRU SEQUENCE UPDATES WITH NEW—DETAIL QUOTES AND COLOR ADDS.

17. MEXICO IS PRONE TO FREQUENT QUAKES BECAUSE OF TWO MAJOR GEOGRAPHIC FAULTS OFFSHORE IN THE PACIFIC OCEAN.

18. THE QUAKE HAD A PRELIMINARY MAGNITUDE OF 6.3 AND IT EPICENTER WAS IN GUERRERO STATE 290 KILOMETERS (165 MILES) SOUTHWEST OF MEXICO CITY SAID RUSS NEEDHAM OF THE U.S. GEOLOGICAL SURVEY 'S EARTHQUAKE INFORMATION CENTER IN GOLDEN COLO.

19. I WAS SO SCARED.

20. THE QUAKE BRIEFLY KNOCKED OUT ELECTRICITY TO SOME AREAS OF THE CAPITAL.

21. THE QUAKE STARTED AROUND 10 20 AM AND WAS FELT FOR MORE THAN A MINUTE IN MEXICO CITY A METROPOLIS OF ABOUT 21 MILLION PEOPLE.

22. QUAKE SHAKES CENTRAL MEXICO—BY ISAAC LEVI—ASSOCIATED PRESS WRITER.

23. A FENCE WAS TOPPLED IN A POOR NEIGHBORHOOD IN ZIHUATANEJO.

The entity grid for permutation 3 is:

```
LOPEZ - - - - - - - - - - - - - - X - - - - - - - - -
CASUALTIES - - O - - - - - - - - - - - - - - - - - - -
NORTHEAST - - - - - - - - - - X - - - - - - - - - - -
COLOMBIA - - - - - - - - - - - - - - X - - - - - - - -
PART S - - - - - - - - - - - - - - - - - - - - -
UPDATES - - - - - - - - - - - - - - - X - - - - - - -
BOGOTA - - - - - - - - - - X - - - - - - - - - - -
QUAKES - X - - - - - - - - - - - - - - X - - - - - -
POWER - - - - - - - - - X - - - - - - - - - - - -
COLOR - - - - - - - - - - - - - - X - - - - - - -
WRITER - - - - - - - - - - - - - - - - - - - - - X -
THOROUGHFARE - - - - - - - - - - - X - - - - - - - - - - -
BUSINESSMAN - - - - - - - - - - - - - - X - - - - - - - - -
TOWN - - - - - - - - - - X - - - - - - - - - - - -
DEAL - O - - - - - - - - - - - - - - - - - - - -
BRIEFLY - - - - - - - - - S - - - - - - - - - - - - -
RUSS - - - - - - - - - - - - - - - - - O - - - - -
ZAHUATANEJO X - - - - - - - - - - - - - - - - - - - - -
BUCARAMANGA - - - - - - - - - - X - - - - - - - - - - - -
METROPOLIS - - - - - - - - - - - - - - - - - - - X - -
ADDS - - - - - - - - - - - - - - - O - - - - - - -
```

```
RESORT X - - - - - - - - - - - - - - - - - - - - -
LEVI—ASSOCIATED - - - - - - - - - - - - - - - - - - - - X -
ALUMINUM - - - - - X - - - - - - - - - - - - - - - -
SURVEY - - - - - - - - - - - - - - - - X - - - - -
BC-MEXICO-QUAKE - - - - - - - - - - - - - - S - - - - - - -
QUAKE - - - X X - - - - - S - - - - X - S - S X X -
MEXICO—BY - - - - - - - - - - - - - - - - - - - - X -
ZIHUATANEJO - - - - - - - - - - - - - - - - - - - - X
L.A. - X - - - - - - - - - - - - - - - - - - - -
ELECTRICITY - - - - - - - - - - - - - - - - - - O - - -
MINUTE - - - - - - - - - - - - - - - - - - - X - -
EPICENTER - - - - - - - - - - - - - - - - S - - - - -
SECTIONS - - - - S - - - - - - - - - - - - - - -
STREET - - - - - X - - - - - - - - - - - - - -
AIRPORT X - - - - - X - - - - - - - - - - - - -
BUILDINGS - - - - - - O - X - O S - - - - - - - - - - -
CENTRAL - - - - - - - - - - - - - - - O - - - - - X -
IT - - - - - - - - - - - - - S - - - S - - - - -
BEACH X - - - - - - - - - - - - - - - - - - - - -
GEOLOGICAL - - - - - - - - - - - - - - - - - X - - - - -
RED - - - S - - - - - - - - - - - - - - - - -
CITY - - - - X X S - - X X - - - - - - X - - X - -
PLENTY - X - - - - - - - - - - - - - - - - - - -
SEQUENCE - - - - - - - - - - - - - - - O - - - - - - -
KILOMETERS - - - - - - - - - O O - - - - - - X - - - - -
EAST-WEST - - - - - - - - - - - X - - - - - - - - - -
TEMBLOR - - - - - X - - - - - - - - - - - - - - -
NEIGHBORHOOD - - - - - - - - - - - - - - - - - - - - X
HIT - - - - - - - - - - - - - - X - - - - - - - -
SATURDAY - - - - - - X - - - - - - - X - - - - - - - -
AREAS - - - - - - - - - - - - - - - - - - X - - -
U.S. - - - - - - - - - - - - - - - - X - - - - -
SCORES S - - - - - - - - - - - - - - - - - - - - - -
```

```
LOS - - - - - - - - - - - - - X - - - - - - - - -
CROSS - - - S - - - - - - - - - - - - - - - - - -
PRESS - - - - - - - - - - - - - - - - - - - - X -
SEARCH - - - - - - - - X - - - - - - - - - - - - -
TERMINAL X - - - - - X - - - - - - - - - - - - - -
SERVICE - - - - - - - - - X - - - - - - - - - - - -
GOLDEN - - - - - - - - - - - - - - - - X - - - - -
DOWNTOWN - - - - X - - - - - - - - - - - - - - - -
SONIA - - - - - O - - - - - - - - - - - - - -
TOURISTS X - - - - - S - - - - - - - - - - - - - -
TRUCKS - - - - - - - - S - - - - - - - - - - - -
FIRE - - - - - - - - S - - - - - - - - - - - -
DISORDERS - - - X - - - - - - - - - - - - - - - - -
AVENUE - - - - - - - - - - - X - - - - - - - - - -
CART - - - - - X - - - - - - - - - - - - - -
PACIFIC - - - - - - - - - - - - - - - X - - - - - -
MEXICO - - - - X X S - - X - - - - - - S X - - X - -
COLO - - - - - - - - - - - - - - - - X - - - - -
ISAAC - - - - - - - - - - - - - - - - - - - X -
FRANCISCO - - - - - - - - - - - - - X - - - - - - - -
AP - - - - - - X - - - - - - - - - - - - -
REPORTS - - - - - - - - - - - - O - - - - - - - - -
ROOF X - - - - - - - - - - - - - - - - - - - - -
GAS - - - - - - - - X - - - - - - - - - - - -
MAGNITUDE - - - - X - - - - - - - - - X - - O - - - - -
EVERYTHING - - - - - S - - - - - - - - - - - - - - -
SHAKES - - - - - - - - - - - - - - - - - - - - X -
WINDOWS S - - - - - - - X - - - - - - - - - - - -
I - S - - - - - - - - - - - - - - - S - - - -
INJURIES - - - - - - - - - - - - X - - - - - - - - -
ARIZPE - - - - - O - - - - - - - - - - - - - - -
MILES - - - - - - - - - X X - - - - - - X - - - - -
ANGELES - - - - - - - - - - - - - X - - - - - - - - -
```

```
FENCE - - - - - - - - - - - - - - - - - - - - - S
FAULTS - - - - - - - - - - - - - - - - X - - - - - -
MEXICO—EDS - - - - - - - - - - - - - - - O - - - - - - -
LEAKS - - - - - - - - X - - - - - - - - - - - - -
OCEAN - - - - - - - - - - - - - - - X - - - - - -
CAPITAL - - - - - - X - - - - - - - - - - - X - - -
INFORMATION - - - - - - - - - - - - - - - - X - - - - -
TELEPHONE - - - - - - - - - X - - - - - - - - - - - -
GUERRERO - - - - - - - - - - - - - - - - X - - - - -
EARTHQUAKE - - - - - - S - - - - - - - X - - X - - - - -
QUOTES - - - - - - - - - - - - - - - X - - - - - - -
VENDOR - - - - - O - - - - - - - - - - - - - -
WRITETHRU - - - - - - - - - - - - - - - O - - - - - - -
STATE - - - - - - - - - - - - - - - - X - - - - -
STREETS - - - - - - - - O - - - - - - - - - - - -
PEOPLE - - - S - - - S - - - - - - - - - - X - -
CENTER - - - - - - - - - - - - - - - - X - - - - -
NEEDHAM - - - - - - - - - - - - - - - - O - - - - -
REFORMA - - - - - - - - - - - X - - - - - - - - - -
```

In apwsE941220.0311 we only correctly detected 1 of 21 permutations as less coherent than the original text. The original text is:

1. BC-JAPAN-EARTHQUAKE—STRONG EARTHQUAKE SHAKES NORTHEAST-ERN JAPAN.

2. TOKYO (AP) A STRONG EARTHQUAKE WITH A PRELIMINARY MAGNI-TUDE READING OF 5.4 SHOOK NORTHEASTERN JAPAN EARLY WEDNES-DAY THE METEOROLOGICAL AGENCY SAID.

3. THERE WERE NO IMMEDIATE REPORTS OF DAMAGE OR INJURIES.

4. THE AGENCY SAID THE EARTHQUAKE STRUCK AT 12 45 AM (1545 GMT TUESDAY) AND WAS CENTERED 90 KILOMETERS (55 MILES) BELOW THE GROUND IN IWATE PREFECTURE (STATE) ABOUT 480 KILOMETERS (300 MILES) NORTH OF TOKYO.

5. NEAR THE EPICENTER THE QUAKE WAS STRONG ENOUGH TO VISIBLY SHAKE BUILDINGS AND SPILL A GLASS OF WATER THE AGENCY SAID.

6. A QUAKE WITH THE SAME MAGNITUDE READING STRUCK BENEATH THE PACIFIC ABOUT 300 KILOMETERS (190 MILES) NORTHEAST OF TOKYO DE-CEMBER 10 ALSO WITHOUT CAUSING INJURY OR MAJOR DAMAGE. (BKS).

The entity grid for the original text is:

GROUND - - - X - - -
INJURIES - - X - - - -
JAPAN O O - - - - -
KILOMETERS - - - X - X -
READING - X - - - - -
AP - X - - - - -
NORTHEASTERN O - - - - - -
METEOROLOGICAL - S - - - - -
BUILDINGS - - - - O - -
WATER - - - - X - -
NORTHEAST - - - - - X -
DECEMBER - - - - - O -
REPORTS - - O - - - -
INJURY - - - - - O -
AGENCY - S - S S - -
EARTHQUAKE S S - X - - -
MILES - - - X - X -
GMT - - - X - - -
QUAKE - - - - S S -
WEDNESDAY - X - - - - -
STATE - - - X - - -
BKS - - - - - - X
EPICENTER - - - - X - -
GLASS - - - - O - -
MAGNITUDE - X - - - S -
IWATE - - - X - - -

TOKYO - S - X - X -
TUESDAY - - - X - - -
PREFECTURE - - - X - - -
DAMAGE - - X - - O -

This was another text that caused our system problems, even though this text consisted of only 6 sentences. Sentence 3 elaborates on the impact of the earthquake without explicitly mentioning the quake, thus causing maximum coherence gaps between the sentences adjacent to it. Between sentences 4 and 5, 'quake' and 'earthquake' are not resolved to the same entity, resulting in a maximum coherence gap. Sentence 6 parallels sentences 2 and 3 in describing the earthquake in a neighbouring city. We are unsure what sentence 7 is supposed to represent, although in the permutations one can see that it is considered part of other sentences, which results in the other texts having only 6 sentences. Since our transition score is not normalized to the number of sentences in a text, the original text is penalized for having an extra sentence (one which contributes a maximum incoherence transition). The coherence score of the original text was 5.48, with the best permutation (text 5) being 4.08. The maximum incoherence score for a text with 6 sentences is 5.0, corresponding to 5 maximum incoherent transitions (1.0). If this last sentence were removed, our system would still only have detected 8 out 21 permutations as less coherent than the original 1 out of 21.

Permutation 5 was ranked most coherent in E941220.0311:

1. THERE WERE NO IMMEDIATE REPORTS OF DAMAGE OR INJURIES.

2. (BKS) A QUAKE WITH THE SAME MAGNITUDE READING STRUCK BENEATH THE PACIFIC ABOUT 300 KILOMETERS (190 MILES) NORTHEAST OF TOKYO DECEMBER 10 ALSO WITHOUT CAUSING INJURY OR MAJOR DAMAGE.

3. THE AGENCY SAID THE EARTHQUAKE STRUCK AT 12 45 AM (1545 GMT TUESDAY) AND WAS CENTERED 90 KILOMETERS (55 MILES) BELOW THE GROUND IN IWATE PREFECTURE (STATE) ABOUT 480 KILOMETERS (300 MILES) NORTH OF TOKYO.

4. BC-JAPAN-EARTHQUAKE—STRONG EARTHQUAKE SHAKES NORTHEASTERN JAPAN.

5. TOKYO (AP) A STRONG EARTHQUAKE WITH A PRELIMINARY MAGNI-
   TUDE READING OF 5.4 SHOOK NORTHEASTERN JAPAN EARLY WEDNES-
   DAY THE METEOROLOGICAL AGENCY SAID.

6. NEAR THE EPICENTER THE QUAKE WAS STRONG ENOUGH TO VISIBLY
   SHAKE BUILDINGS AND SPILL A GLASS OF WATER THE AGENCY SAID.

The entity grid for permutation 5 is:

GROUND - - X - - -
INJURIES X - - - - -
JAPAN - - - O O -
KILOMETERS - X X - - -
READING - X - - X -
AP - - - - X -
NORTHEASTERN - - - O - -
METEOROLOGICAL - - - - S -
BUILDINGS - - - - - O
WATER - - - - - X
NORTHEAST - X - - - -
DECEMBER - O - - - -
REPORTS O - - - - -
INJURY - O - - - -
AGENCY - - S - S S
EARTHQUAKE - - X S S -
MILES - X X - - -
GMT - - X - - -
QUAKE - S - - - S
BKS - X - - - -
STATE - - X - - -
WEDNESDAY - - - - X -
EPICENTER - - - - - X
GLASS - - - - - O
MAGNITUDE - X - - X -
IWATE - - X - - -

TOKYO - X X - S -
TUESDAY - - X - - -
PREFECTURE - - X - - -
DAMAGE X O - - - -

In apwsE950116.0114 we only correctly detected 6 of 21 permutations as less coherent than the original text. The original text is:

1. BC-JAPAN-EARTHQUAKE—STRONG EARTHQUAKE ROCKS WESTERN JAPAN.

2. TOKYO (AP) A STRONG EARTHQUAKE WITH A PRELIMINARY MAGNI-TUDE OF 7.2 ROCKED WESTERN JAPAN EARLY TUESDAY.

3. THERE WERE NO IMMEDIATE REPORTS OF DAMAGE OR INJURY FROM THE 5 46 AM (2046 GMT) EARTHQUAKE CENTERED 20 KILOMETERS (12.5 MILES) UNDER THE GROUND ON THE ISLAND OF AWAJI OFF OSAKA 450 KILOMETERS WEST OF TOKYO.

4. THE QUAKE MEASURED SIX ON THE JAPANESE SCALE OF SEVEN IN THE PORT CITY OF KOBE NEAR OSAKA THE CENTRAL METEOROLOGICAL AGENCY SAID.

5. A QUAKE OF SIX CAN SHOOK BUILDINGS VIOLENTLY CRACK ASPHALT AND BURST UNDERGROUND PIPES.

6. NO WARNING WAS ISSUED FOR TSUNAMI OR TIDAL WAVES CAUSED BY EARTHQUAKES.

7. (KPH).

The entity grid for the original text is:

WAVES - - - - - X -
GROUND - - X - - - -
JAPAN O X - - - - -
KILOMETERS - - O - - - -
WESTERN O - - - - - -
AP - X - - - - -
BUILDINGS - - - - X - -

```
METEOROLOGICAL - - - S - - -
SCALE - - - X - - -
TSUNAMI - - - - - X -
REPORTS - - S - - - -
INJURY - - X - - - -
ISLAND - - X - - - -
EARTHQUAKE S X X - - - -
AGENCY - - - S - - -
MILES - - X - - - -
GMT - - X - - - -
QUAKE - - - S S - -
EARTHQUAKES - - - - - X -
CENTRAL - - - S - - -
KPH - - - - - - X
ROCKS X - - - - - -
OSAKA - - X X - - -
MAGNITUDE - X - - - - -
CITY - - - X - - -
KOBE - - - X - - -
TIDAL - - - - - X -
TOKYO - X X - - - -
TUESDAY - X - - - - -
PIPES - - - - O - -
AWAJI - - X - - - -
ASPHALT - - - - O - -
DAMAGE - - X - - - -
WARNING - - - - - S -
```

This document set again includes an extra sentence in the original text than in its permutations. The last sentence consists solely of 'KPH', whereas the other permutations have it merged within one of the sentences. Hence, the original text receives a penalty of 1.0 in its overall coherence score. If this sentence were removed, the original text would be identified as being more coherent than all of the other permutations. No further analysis of this document set is provided.

Permutation 4 was ranked most coherent in E950116.0114:

1. NO WARNING WAS ISSUED FOR TSUNAMI OR TIDAL WAVES CAUSED BY EARTHQUAKES.

2. (KPH) TOKYO (AP) A STRONG EARTHQUAKE WITH A PRELIMINARY MAG-NITUDE OF 7.2 ROCKED WESTERN JAPAN EARLY TUESDAY.

3. THERE WERE NO IMMEDIATE REPORTS OF DAMAGE OR INJURY FROM THE 5 46 AM (2046 GMT) EARTHQUAKE CENTERED 20 KILOMETERS (12.5 MILES) UNDER THE GROUND ON THE ISLAND OF AWAJI OFF OSAKA 450 KILOMETERS WEST OF TOKYO.

4. THE QUAKE MEASURED SIX ON THE JAPANESE SCALE OF SEVEN IN THE PORT CITY OF KOBE NEAR OSAKA THE CENTRAL METEOROLOGICAL AGENCY SAID.

5. A QUAKE OF SIX CAN SHOOK BUILDINGS VIOLENTLY CRACK ASPHALT AND BURST UNDERGROUND PIPES.

6. BC-JAPAN-EARTHQUAKE—STRONG EARTHQUAKE ROCKS WESTERN JAPAN.

The entity grid for permutation 4 is:

```
WAVES X - - - - -
GROUND - - X - - -
JAPAN - X - - - O
KILOMETERS - - O - - -
WESTERN - - - - - O
AP - X - - - -
BUILDINGS - - - - X -
METEOROLOGICAL - - - S - -
SCALE - - - X - -
TSUNAMI X - - - - -
REPORTS - - S - - -
INJURY - - X - - -
ISLAND - - X - - -
EARTHQUAKE - X X - - S
```

AGENCY - - - S - -

MILES - - X - - -

GMT - - X - - -

QUAKE - - - S S -

EARTHQUAKES X - - - - -

CENTRAL - - - S - -

KPH - X - - - -

OSAKA - - X X - -

ROCKS - - - - - X

MAGNITUDE - X - - - -

CITY - - - X - -

KOBE - - - X - -

TIDAL X - - - - -

TOKYO - X X - - -

TUESDAY - X - - - -

PIPES - - - - O -

AWAJI - - X - - -

ASPHALT - - - - O -

DAMAGE - - X - - -

WARNING S - - - - -

In apwsE950119.0261 we only correctly detected 6 of 21 permutations as less coherent than the original text. The original text is:

1. BC-COLOMBIA-QUAKE,0078—QUAKE HITS COLOMBIA.

2. BOGOTA COLOMBIA (AP) A STRONG EARTHQUAKE ROCKED PARTS OF COLOMBIA ON THURSDAY.

3. THERE WERE NO REPORTED INJURIES.

4. AUTHORITIES REPORTED THE 10 05 AM QUAKE HAD A PRELIMINARY MAGNITUDE OF 6.9 AT ITS EPICENTER.

5. RADIO REPORTS SAID THE TEMBLOR WAS STRONGEST IN NORTHEAST-ERN COLOMBIA.

6. THE QUAKE LASTED ABOUT 30 SECONDS IN THE CAPITAL BOGOTA SHAK-
   ING BUILDINGS AND SENDING SOME FRIGHTENED RESIDENTS RUNNING
   INTO THE STREET.

The entity grid for the original text is:

RADIO - - - - S -
INJURIES - - O - - -
STREET - - - - - X
RESIDENTS - - - - - S
AP - X - - - -
BUILDINGS - - - - - O
EARTHQUAKE - X - - - -
QUAKE - - - X - S
EPICENTER - - - X - -
MAGNITUDE - - - O - -
SECONDS - - - - - O
COLOMBIA X S - - X -
THURSDAY - X - - - -
PARTS - O - - - -
AUTHORITIES - - - S - -
BC-COLOMBIA-QUAKE,0078—QUAKE X - - - - -
05 - - - S - -
TEMBLOR - - - - S -
CAPITAL - - - - - X
BOGOTA - S - - - X
HITS X - - - - -

Between sentences 1 and 2, 'BC-COLOMBIA-QUAKE,0078—QUAKE' is treated as a
single entity, with 'quake' not being resolved to 'earthquake'. The only entity in common
is 'Colombia'. Sentence 3 has a single entity 'injuries' due to the earthquake. This causes
a maximum coherence transition between sentences 2 and 3 as well as sentences 3 and 4.
Between sentences 4 and 5 there again is a maximum coherence gap due to 'temblor' and
'quake' not being resolved to each other, which also causes the same gap in sentence 5 and
6. Also between sentence 5 and 6, the relationship between 'Colombia' and the 'capital'
of Colombia, 'Bogota', is not resolved to refer to each other. It can be seen that the low

score of this text is due to the poor coreference resolution as well as sentence 3 inducing two maximum incoherent transitions. A permutation with sentence 3 as either the first or last sentence would result in a text with a higher coherence score.

Permutation 9 was ranked most coherent in E950119.0261:

1. BOGOTA COLOMBIA (AP) A STRONG EARTHQUAKE ROCKED PARTS OF COLOMBIA ON THURSDAY.

2. RADIO REPORTS SAID THE TEMBLOR WAS STRONGEST IN NORTHEAST-ERN COLOMBIA.

3. BC-COLOMBIA-QUAKE,0078—QUAKE HITS COLOMBIA.

4. THERE WERE NO REPORTED INJURIES.

5. AUTHORITIES REPORTED THE 10 05 AM QUAKE HAD A PRELIMINARY MAGNITUDE OF 6.9 AT ITS EPICENTER.

6. THE QUAKE LASTED ABOUT 30 SECONDS IN THE CAPITAL BOGOTA SHAK-ING BUILDINGS AND SENDING SOME FRIGHTENED RESIDENTS RUNNING INTO THE STREET.

The entity grid for permutation 9 is:

RADIO - S - - - -
INJURIES - - - O - -
STREET - - - - - X
RESIDENTS - - - - - S
AP X - - - - -
BUILDINGS - - - - - O
EARTHQUAKE X - - - - -
QUAKE - - - - X S
EPICENTER - - - - X -
MAGNITUDE - - - - O -
SECONDS - - - - - O
COLOMBIA S X X - - -
THURSDAY X - - - - -
PARTS O - - - - -

AUTHORITIES - - - - S -

TEMBLOR - S - - - -

BC-COLOMBIA-QUAKE,0078—QUAKE - - X - - -

05 - - - - S -

CAPITAL - - - - - X

BOGOTA S - - - - X

HITS - - X - - -

In apwsE950122.0109 we only correctly detected 1 of 21 permutations as less coherent than the original text. The original text is:

1. BC-BRITAIN-QUAKE RESCUE—WITH JAPAN-QUAKE—BRITISH RESCUE TEAMS FLIES TO JAPAN TO SEARCH FOR SURVIVORS.

2. LONDON (AP) A 15-STRONG BRITISH RESCUE TEAM LEFT FOR JAPAN SUNDAY AFTER THE GOVERNMENT CHANGED ITS MIND AND ASKED FOR HELP IN FINDING SURVIVORS OF TUESDAY 'S DEVASTATING EARTHQUAKE.

3. THE VOLUNTEERS FROM THE INTERNATIONAL RESCUE CORPS A BRITISH RELIEF AGENCY PLAN TO USE SOUND LOCATION EQUIPMENT VIDEO PROBES AND THERMAL IMAGING CAMERAS TO SEARCH THE RUINS OF KOBE.

4. WE DID N'T GO BEFORE BECAUSE THE JAPANESE GOVERNMENT SAID NO THANKS.

5. WE OFFERED OUR HELP 30 MINUTES AFTER THE DISASTER WILLIE MCMARTIN THE TEAM COORDINATOR SAID BEFORE BOARDING A FLIGHT TO OSAKA.

6. IT WAS ONLY ON SATURDAY THAT THEY CHANGED THEIR MINDS AND THIS IS THE FIRST FLIGHT OUT HE SAID.

7. MCMARTIN SAID CHANCES OF FINDING PEOPLE ALIVE WERE LOW BUT HE NOTED THAT EARTHQUAKE SURVIVORS HAVE BEEN FOUND AFTER 11 DAYS.

8. IF WE ARE REALISTIC WE ARE LOOKING FOR ONES AND TWOS.

9. WE ARE NOT GOING TO FIND HUNDREDS.

10. BUT THESE ONE OR TWO PEOPLE DESERVE THE SAME CHANCE AS THE REST.

11. SO WE ARE GOING TO GIVE IT OUR BEST SHOT HE SAID.

12. ONE OF THE TWO INTERPRETERS ACCOMPANYING THE 13 TRAINED RES-CUE WORKERS ORIGINALLY CAME FROM KOBE.

13. I AM NOT LOOKING FORWARD TO GOING IN ONE WAY BUT I 'M GLAD THAT THE RESCUE TEAM DECIDED TO GO AND RESCUE SOME OF THE PEOPLE I 'M REALLY PLEASED SAID YOKO MARSH WHO IS MARRIED TO A BRITON.

14. (EML).

The entity grid for the original text is:

HELP - X - - O - - - - - - - - -
PROBES - - X - - - - - - - - - - -
SATURDAY - - - - - X - - - - - - - -
JAPAN X X - - - - - - - - - - - -
IT - - - - - S - - - - O - - -
HUNDREDS - - - - - - - - O - - - - -
ONES - - - - - - - X - - - - - -
INTERNATIONAL - - X - - - - - - - - - - -
AP - X - - - - - - - - - - - -
RESCUE—WITH X - - - - - - - - - - - -
COORDINATOR - - - - S - - - - - - - - -
CAMERAS - - X - - - - - - - - - - -
REST - - - - - - - - - X - - - -
MIND - O - - - - - - - - - - - -
AGENCY - - X - - - - - - - - - -
RESCUE X X X - - - - - - - - O S -
EARTHQUAKE - X - - - - S - - - - - - -
WILLIE - - - - S - - - - - - - - -

BC-BRITAIN-QUAKE X - - - - - - - - - - - - - -
MINDS - - - - - O - - - - - - - - -
DAYS - - - - - - X - - - - - - -
VOLUNTEERS - - S - - - - - - - - - - -
TEAMS X - - - - - - - - - - - - -
TEAM - S - - X - - - - - - - S -
HE - - - - - S S - - - S - - -
TWOS - - - - - - - X - - - - - -
WAY - - - - - - - - - - - - X -
EML - - - - - - - - - - - - - X
FLIES X - - - - - - - - - - - - -
RUINS - - O - - - - - - - - - - -
OSAKA - - - - X - - - - - - - - -
THEY - - - - - S - - - - - - - -
WORKERS - - - - - - - - - - - O - -
CORPS - - X - - - - - - - - - - -
LOCATION - - O - - - - - - - - - -
KOBE - - X - - - - - - - - X - -
CHANCES - - - - - - S - - - - - - -
MINUTES - - - - X - - - - - - - - -
TUESDAY - X - - - - - - - - - - - -
FLIGHT - - - - O O - - - - - - - -
SURVIVORS X O - - - - S - - - - - - -
BRITON - - - - - - - - - - - - X -
WE - - - S S - - S S - S - - -
DISASTER - - - - X - - - - - - - - -
PEOPLE - - - - - - O - - X - - X -
LONDON - S - - - - - - - - - - - -
RELIEF - - X - - - - - - - - - - -
GOVERNMENT - X - S - - - - - - - - - -
CHANCE - - - - - - - - - O - - - -
SHOT - - - - - - - - - - X - - -
YOKO - - - - - - - - - - - - O -

THANKS - - - X - - - - - - - - - -
MCMARTIN - - - - S - S - - - - - - -
SUNDAY - X - - - - - - - - - - - -
EQUIPMENT - - O - - - - - - - - - - -
INTERPRETERS - - - - - - - - - - - X - -
MARSH - - - - - - - - - - - - O -
I - - - - - - - - - - - - S -

Permutation has 13 sentences, original text has 14 sentences. Thus original has an extra maximum incoherent transition penalty due to the last sentence consisting of 'EML'. If this penalty were subtracted, then our system would correctly detect 19 out of the 20 permutations as less coherent (only permutation 9 shown above is deemed more coherent than the original). No further analysis is given.

Permutation 9 was ranked most coherent in E950122.0109:

1. BUT THESE ONE OR TWO PEOPLE DESERVE THE SAME CHANCE AS THE REST.

2. MCMARTIN SAID CHANCES OF FINDING PEOPLE ALIVE WERE LOW BUT HE NOTED THAT EARTHQUAKE SURVIVORS HAVE BEEN FOUND AFTER 11 DAYS.

3. SO WE ARE GOING TO GIVE IT OUR BEST SHOT HE SAID.

4. ONE OF THE TWO INTERPRETERS ACCOMPANYING THE 13 TRAINED RES-CUE WORKERS ORIGINALLY CAME FROM KOBE.

5. I AM NOT LOOKING FORWARD TO GOING IN ONE WAY BUT I 'M GLAD THAT THE RESCUE TEAM DECIDED TO GO AND RESCUE SOME OF THE PEOPLE I 'M REALLY PLEASED SAID YOKO MARSH WHO IS MARRIED TO A BRITON.

6. IF WE ARE REALISTIC WE ARE LOOKING FOR ONES AND TWOS.

7. WE ARE NOT GOING TO FIND HUNDREDS.

8. WE OFFERED OUR HELP 30 MINUTES AFTER THE DISASTER WILLIE MC-MARTIN THE TEAM COORDINATOR SAID BEFORE BOARDING A FLIGHT TO OSAKA.

9. WE DID N'T GO BEFORE BECAUSE THE JAPANESE GOVERNMENT SAID NO THANKS.

10. IT WAS ONLY ON SATURDAY THAT THEY CHANGED THEIR MINDS AND THIS IS THE FIRST FLIGHT OUT HE SAID.

11. THE VOLUNTEERS FROM THE INTERNATIONAL RESCUE CORPS A BRITISH RELIEF AGENCY PLAN TO USE SOUND LOCATION EQUIPMENT VIDEO PROBES AND THERMAL IMAGING CAMERAS TO SEARCH THE RUINS OF KOBE.

12. LONDON (AP) A 15-STRONG BRITISH RESCUE TEAM LEFT FOR JAPAN SUNDAY AFTER THE GOVERNMENT CHANGED ITS MIND AND ASKED FOR HELP IN FINDING SURVIVORS OF TUESDAY 'S DEVASTATING EARTHQUAKE.

13. (EML) BC-BRITAIN-QUAKE RESCUE—WITH JAPAN-QUAKE—BRITISH RESCUE TEAMS FLIES TO JAPAN TO SEARCH FOR SURVIVORS.

The entity grid for permutation 9 is:

```
HELP - - - - - - - - O - - - X -
PROBES - - - - - - - - - - X - -
SATURDAY - - - - - - - - - X - - -
JAPAN - - - - - - - - - - - X X
IT - - O - - - - - - S - - -
HUNDREDS - - - - - - O - - - - - -
ONES - - - - - X - - - - - - -
INTERNATIONAL - - - - - - - - - - X - -
AP - - - - - - - - - - - X -
RESCUE—WITH - - - - - - - - - - - - X
COORDINATOR - - - - - - - S - - - - -
REST X - - - - - - - - - - - -
CAMERAS - - - - - - - - - - X - -
MINDS - - - - - - - - - O - - -
WILLIE - - - - - - - S - - - - -
EARTHQUAKE - S - - - - - - - - - X -
RESCUE - - - O S - - - - - X X X
```

```
DAYS - X - - - - - - - - - - -
AGENCY - - - - - - - - - - X - -
MIND - - - - - - - - - - - O -
BC-BRITAIN-QUAKE - - - - - - - - - - - - X
VOLUNTEERS - - - - - - - - - - S - -
WAY - - - - X - - - - - - - -
HE - S S - - - - - - S - - -
TEAM - - - - S - - X - - - S -
TWOS - - - - - X - - - - - - -
TEAMS - - - - - - - - - - - X
EML - - - - - - - - - - - - X
OSAKA - - - - - - - X - - - - -
WORKERS - - - O - - - - - - - -
THEY - - - - - - - - - S - - -
RUINS - - - - - - - - - - O - -
FLIES - - - - - - - - - - - - X
CORPS - - - - - - - - - - X - -
KOBE - - - X - - - - - - X - -
CHANCES - S - - - - - - - - - -
LOCATION - - - - - - - - - - O - -
MINUTES - - - - - - - X - - - - -
BRITON - - - - X - - - - - - - -
FLIGHT - - - - - - - O - O - - -
SURVIVORS - S - - - - - - - - - O X
TUESDAY - - - - - - - - - - - X -
PEOPLE X O - - X - - - - - - - -
WE - - S - - S S S S - - - -
DISASTER - - - - - - - X - - - - -
LONDON - - - - - - - - - - - S -
RELIEF - - - - - - - - - - X - -
CHANCE O - - - - - - - - - - - -
YOKO - - - - O - - - - - - - -
SHOT - - X - - - - - - - - - -
```

GOVERNMENT - - - - - - - - S - - X -
THANKS - - - - - - - - X - - - -
MCMARTIN - S - - - - - S - - - - -
SUNDAY - - - - - - - - - - - X -
EQUIPMENT - - - - - - - - - - O - -
INTERPRETERS - - - X - - - - - - - - -
MARSH - - - - O - - - - - - - -
I - - - - S - - - - - - - -

Appendix B

Sample of Aviation Accident Reports

Document NTSB-3745 we only detected the original text one time, and the rest we assumed the permutations to be more coherent: original text:

1. AT 1114 17 THE PILOT OF YV385CP TOLD FAA ARTCC THAT HE WAS EX-PERIENCING AN EMERGENCY AND AT 1114 29 HE ASKED FOR IMMEDIATE DIRECTIONS TO BIMINI.

2. THERE WERE NO FURTHER COMMUNICATIONS WITH YV395CP.

3. AT THAT TIME THE AIRCRAFT WAS ABOUT 9 MILES SOUTHWEST OF BI-MINI BAHAMAS ON A SOUTHEASTERLY HEADING AT A SPEED OF ABOUT 200 KNOTS AND AN ALTITUDE OF ABOUT 21,000 FEET.

4. RESCUE PERSONNEL LOCATED DEBRIS AND FUEL AT LATITUDE 25 DE-GREES 47.0 MINUTES NORTH LONGITUDE 079 DEGREES 23.6 MINUTES WEST CONSISTENT WITH THE POSITION RELAYED BY ARTCC.

5. ACCORDING TO RADAR INFORMATION AT 1113 33 YV395CP 'S ALTITUDE WAS ABOUT 23,600 FEET AND AT 1114 58 YV385CP HAD DESCENDED TO 2,800 FEET.

6. SOUND SPECTRUM ANALYSIS OF AN FAA RE-RECORDING OF COMMUNI-
   CATIONS BETWEEN THE PILOT AND ATC INDICATED THERE WERE ELEC-
   TRONIC SIGNATURES PRESENT ATTRIBUTABLE TO PROPELLER NOISE
   AT CRUISE SETTING FOR THAT TYPE OF AIRCRAFT.

7. RECORDS OBTAINED FROM VENEZUELA INDICATED THAT EARLIER YV385CP
   HAD BEEN INVOLVED IN AN ACCIDENT AND HAD INCURRED EXTENSIVE
   STRUCTURAL DAMAGE.

8. THE AIRCRAFT HAD BEEN REPAIRED IN VENEZUELA AND WAS SUBSE-
   QUENTLY FLOWN TO THE UNITES STATES FOR ADDITIONAL REPAIRS.

9. RECORDS OBTAINED FROM THE U. S. REPAIR STATION INDICATED THAT
   THE AIRPLANE INCURRED AN EXTENSIVE MAINTENANCE HISTORY WITH
   NEEDED REPAIRS INCLUDING DAMAGED WIRING AND LEAKS IN THE
   FUEL SYSTEM AS-WELL-AS STRUCTURAL LEAKS WHICH T HAD LEAD TO
   SEVERAL PRESSURIZATION DIFFICULTIES.

original grid:

ALTITUDE - - X - O - - - -
PERSONNEL - - - S - - - - -
BIMINI X - X - - - - - -
KNOTS - - X - - - - - -
DEBRIS - - - X - - - - -
SETTING - - - - - X - - -
HISTORY - - - - - - - - O
REPAIRS - - - - - - - X X
AIRPLANE - - - - - - - - S
ARTCC O - - X - - - - -
VENEZUELA - - - - - - X X -
MAINTENANCE - - - - - - - - O
LATITUDE - - - X - - - - -
TIME - - X - - - - - -
RECORDS - - - - - - S - S
U. - - - - - - - - X

RADAR - - - - X - - - -

ATC - - - - - X - - -

S. - - - - - - - - X

EMERGENCY O - - - - - - - -

DEGREES - - - X - - - - -

STATION - - - - - - - - X

RESCUE - - - S - - - - -

REPAIR - - - - - - - - X

MILES - - O - - - - - -

POSITION - - - X - - - - -

WIRING - - - - - - - - O

BAHAMAS - - X - - - - - -

SPEED - - X - - - - - -

SPECTRUM - - - - - S - - -

DIRECTIONS X - - - - - - - -

SIGNATURES - - - - - O - - -

HE S - - - - - - - -

ACCIDENT - - - - - - X - -

LONGITUDE - - - X - - - - -

FAA O - - - - X - - -

INFORMATION - - - - X - - - -

YV385CP X - - - S - X - - -

YV395CP - X - - X - - - -

COMMUNICATIONS - O - - - X - - -

ANALYSIS - - - - - S - - -

DIFFICULTIES - - - - - - - - X

UNITES - - - - - - - X -

FEET - - X - X - - - -

MINUTES - - - X - - - - -

PILOT S - - - - X - - -

AIRCRAFT - - S - - X - X -

FUEL - - - X - - - - X

LEAKS - - - - - - - - X

```
DAMAGE - - - - - - O - -
NOISE - - - - - O - - -
TYPE - - - - - X - - -
STATES - - - - - - - X -
SYSTEM - - - - - - - - X
RE-RECORDING - - - - - X - - -
CRUISE - - - - - X - - -
```

In the original text of document NTSB-3745, there is lexical cohesion that we do not capture between sentences 1 and 2 through the words 'told' and 'communication'. Further, the document has many spelling errors (e.g., they refer to aircraft 'YV385CP' and 'YV395CP' interchangeably through the entire text, leaving one unsure whether there were two distinct aircraft involved in the accident or a single one; 'United States' is spelled as 'Unites States'). Also, 'he' and the 'pilot' are not resolved to the same entity in sentence 1. Between sentences 2 and 3 there is lexical and referential cohesion between 'YV395CP' and 'aircraft' that we do not capture. Between sentences 3 and 4 there is a mereological relationship between 'aircraft' and 'debris and fuel' from the aircraft that is not captured. Throughout the text there is temporal and spatial information regarding times and positions of the aircraft that do not get resolved to each other (e.g., words such as 'feet', 'knots', 'minutes', 'position', 'latitude', 'longitude', 'degrees', 'altitude', 'radar information', etc. add more entities to our utterance vector thus increasing the dissimilarity measure between utterances). In sentence 6 'Sound spectrum analysis of an FAA re-recording of communications' introduces 5 distinct entities in the utterance, whereas this should probably be treated as at most two entities, 'analysis' and 'communications', as in the nearly equivalent expression 'Analysis of communications'. Sentence 7 presents a fairly new topic break (the 'aircraft's records and history'), so it is fair to say there is a topic shift with this new sentence. Sentences 7 and 8 again do not have their referential and lexical cohesion captured in both 'YV385CP' and 'aircraft', as well as the relationship between 'structural damage' and 'repairs'. Sentences 8 and 9 do not have 'Unites States' (sic) and 'U.S.' resolved to the same entity. Oddly enough, 'U.' and 'S.' are treated as distinct entities, which is a failure of the entity recognizer and not the coreference resolver that Barzilay and Lapata employ. Sentences 8 and 9 also do not resolve 'repair' and its plural 'repairs' to the same entity. Also not captured is the lexically cohesive tie between 'airplane' and 'aircraft', nor the mereological relationship between 'aircraft' and its 'fuel system'.

Permutation 4 in NTSB-3745:

1. AT 1114 17 THE PILOT OF YV385CP TOLD FAA ARTCC THAT HE WAS EX-PERIENCING AN EMERGENCY AND AT 1114 29 HE ASKED FOR IMMEDIATE DIRECTIONS TO BIMINI.

2. RECORDS OBTAINED FROM THE U. S. REPAIR STATION INDICATED THAT THE AIRPLANE INCURRED AN EXTENSIVE MAINTENANCE HISTORY WITH NEEDED REPAIRS INCLUDING DAMAGED WIRING AND LEAKS IN THE FUEL SYSTEM AS-WELL-AS STRUCTURAL LEAKS WHICH T HAD LEAD TO SEVERAL PRESSURIZATION DIFFICULTIES.

3. RESCUE PERSONNEL LOCATED DEBRIS AND FUEL AT LATITUDE 25 DE-GREES 47.0 MINUTES NORTH LONGITUDE 079 DEGREES 23.6 MINUTES WEST CONSISTENT WITH THE POSITION RELAYED BY ARTCC.

4. THERE WERE NO FURTHER COMMUNICATIONS WITH YV395CP.

5. SOUND SPECTRUM ANALYSIS OF AN FAA RE-RECORDING OF COMMUNI-CATIONS BETWEEN THE PILOT AND ATC INDICATED THERE WERE ELEC-TRONIC SIGNATURES PRESENT ATTRIBUTABLE TO PROPELLER NOISE AT CRUISE SETTING FOR THAT TYPE OF AIRCRAFT.

6. THE AIRCRAFT HAD BEEN REPAIRED IN VENEZUELA AND WAS SUBSE-QUENTLY FLOWN TO THE UNITES STATES FOR ADDITIONAL REPAIRS.

7. RECORDS OBTAINED FROM VENEZUELA INDICATED THAT EARLIER YV385CP HAD BEEN INVOLVED IN AN ACCIDENT AND HAD INCURRED EXTENSIVE STRUCTURAL DAMAGE.

8. ACCORDING TO RADAR INFORMATION AT 1113 33 YV395CP 'S ALTITUDE WAS ABOUT 23,600 FEET AND AT 1114 58 YV385CP HAD DESCENDED TO 2,800 FEET.

9. AT THAT TIME THE AIRCRAFT WAS ABOUT 9 MILES SOUTHWEST OF BI-MINI BAHAMAS ON A SOUTHEASTERLY HEADING AT A SPEED OF ABOUT 200 KNOTS AND AN ALTITUDE OF ABOUT 21,000 FEET.

permutation 4 grid:

ALTITUDE - - - - - - - O X

PERSONNEL - - S - - - - - -

BIMINI X - - - - - - - X

DEBRIS - - X - - - - - -

SETTING - - - - X - - - -

KNOTS - - - - - - - - X

HISTORY - O - - - - - - -

REPAIRS - X - - - X - - -

AIRPLANE - S - - - - - - -

ARTCC O - X - - - - - -

VENEZUELA - - - - - X X - -

MAINTENANCE - O - - - - - - -

RECORDS - S - - - - S - -

U. - X - - - - - - -

LATITUDE - - X - - - - - -

TIME - - - - - - - - X

S. - X - - - - - - -

ATC - - - - X - - - -

RADAR - - - - - - - X -

EMERGENCY O - - - - - - - -

DEGREES - - X - - - - - -

STATION - X - - - - - - -

REPAIR - X - - - - - - -

RESCUE - - S - - - - - -

MILES - - - - - - - - O

POSITION - - X - - - - - -

WIRING - O - - - - - - -

SPECTRUM - - - - S - - - -

BAHAMAS - - - - - - - - X

SPEED - - - - - - - - X

DIRECTIONS X - - - - - - - -

SIGNATURES - - - - O - - - -

HE S - - - - - - - - -
ACCIDENT - - - - - - X - -
LONGITUDE - - X - - - - - -
FAA O - - - X - - - -
DIFFICULTIES - X - - - - - - -
YV385CP X - - - - - X S -
YV395CP - - - X - - - X -
COMMUNICATIONS - - - O X - - - -
ANALYSIS - - - - S - - - -
INFORMATION - - - - - - - X -
UNITES - - - - - X - - -
MINUTES - - X - - - - - -
FEET - - - - - - - X X
PILOT S - - - X - - - -
FUEL - X X - - - - - -
AIRCRAFT - - - - X X - - S
LEAKS - X - - - - - - -
DAMAGE - - - - - - O - -
NOISE - - - - O - - - -
TYPE - - - - X - - - -
STATES - - - - - X - - -
SYSTEM - X - - - - - - -
RE-RECORDING - - - - X - - - -
CRUISE - - - - X - - - -

In document NTSB-3861, only have the time were the permutations found to be less coherent than the original. The original document is:

1. THE AIRPLANE COLLIDED WITH THE HIGH DESERT TERRAIN DURING MANEUVERING FLIGHT WHILE ON A DUAL INSTRUCTIONAL FLIGHT.

2. THE ACCIDENT SITE ELEVATION WAS 4,400 FEET MEAN SEA LEVEL (MSL).

3. THE INSTRUCTOR AND DUAL COMMERCIAL STUDENT WERE SCHEDULED TO PRACTICE COMMERCIAL MANEUVERS A SIMULATED EMERGENCY APPROACH AND LANDING AND SYSTEM/EQUIPMENT MALFUNCTIONS.

4. RADAR DATA DEPICTED THE AIRPLANE WAS OPERATED IN THE PRE-SCRIBED PRACTICE AREA CONDUCTING FLIGHT MANEUVERS.

5. THE RADAR DATA ALSO INDICATED THAT THE AIRPLANE 'S MAXIMUM ALTITUDE WAS 8,200 FEET MSL AND ITS LOWEST ALTITUDE WAS 6,500 FEET MSL THE MINIMUM ALTITUDE OBSERVED BY THE RADAR.

6. THE FLIGHT WAS NOT IN RADIO CONTACT WITH ANY GROUND FACILITY.

7. AT THE ACCIDENT SITE THE LANDING GEAR WAS FOUND IN THE EX-TENDED POSITION AND THE FLAPS WERE FOUND FULLY RETRACTED.

8. DAMAGE TO THE PROPELLER INDICATED IT WAS ROTATING AT THE TIME OF IMPACT.

9. OBSERVED EVIDENCE AT THE ACCIDENT SITE INDICATED THE AIRPLANE IMPACTED WITH A HIGH DESCENT RATE AND LOW FORWARD VELOCITY IN A SLIGHT LEFT WING DOWN AND NOSE DOWN ATTITUDE.

10. NO MECHANICAL ANOMALIES WERE FOUND WITH THE AIRPLANE.

11. TRACES OF ANTIHISTAMINE EPHEDRINE AND PHENYLPROPANOLAMINE WERE FOUND IN THE BLOOD DURING THE TOXICOLOGY EXAMINATION OF THE FLIGHT INSTRUCTOR.

   Original entity grid:

RADIO - - - - - X - - - - -
ALTITUDE - - - - S - - - - - -
CONTACT - - - - - X - - - - -
GROUND - - - - - X - - - - -
FLAPS - - - - - - S - - - -
TOXICOLOGY - - - - - - - - - - X
MANEUVERS - - X O - - - - - - -
WHILE X - - - - - - - - - -
LEFT - - - - - - - - X - -
IT - - - - - - - S - - -
ATTITUDE - - - - - - - - O - -

```
TERRAIN X - - - - - - - - - -
AIRPLANE S - - O X - - - O X -
VELOCITY - - - - - - - - X - -
BLOOD - - - - - - - - - - X
SITE - S - - - - X - X - -
TRACES - - - - - - - - - - S
AREA - - - X - - - - - - -
GEAR - - - - - - S - - - -
ELEVATION - S - - - - - - - - -
TIME - - - - - - - X - - -
RADAR - - - S S - - - - - -
EMERGENCY - - X - - - - - - - -
NOSE - - - - - - - - X - -
STUDENT - - S - - - - - - - -
ANOMALIES - - - - - - - - - S -
EXAMINATION - - - - - - - - - - X
DESERT X - - - - - - - - - -
POSITION - - - - - - X - - - -
EPHEDRINE - - - - - - - - - - S
LEVEL - O - - - - - - - - -
MSL - X - - - - - - - - -
MALFUNCTIONS - - X - - - - - - - -
INSTRUCTOR - - S - - - - - - - X
EVIDENCE - - - - - - - - S - -
LANDING - - X - - - S - - - -
ACCIDENT - S - - - - X - X - -
RATE - - - - - - - - X - -
PROPELLER - - - - - - - X - - -
FEET - X - - X - - - - - -
APPROACH - - X - - - - - - - -
FLIGHT X - - O - S - - - - X
ANTIHISTAMINE - - - - - - - - - - X
DATA - - - S S - - - - - -
```

PRACTICE - - - X - - - - - - -

PHENYLPROPANOLAMINE - - - - - - - - - - S

DAMAGE - - - - - - - - S - - -

IMPACT - - - - - - - X - - -

SEA - O - - - - - - - - - -

FACILITY - - - - - X - - - - -

Between sentences 1 and 2 there is lexical cohesion that is not captured between the 'accident site' and the collision between the 'airplane' and the 'high desert terrain', hence there is a maximum coherence gap between these two utterances since they have no realized entities in common. Between sentences 2 and 3 there again is a maximum coherence gap. Sentence 2 can be considered an elaboration on sentence 1, whereas sentence 3 is continuing providing background to sentence 1. Between sentences 3 and 4 there is a lexical relationship that is not captured between the 'airplane' and the 'instructor and student' of the airplane, as well as 'practice area' with 'simulated emergency approach and landing and system/equipment malfunctions'. Also, 'commercial maneuvers' and 'flight maneuvers' are not resolved to a single entity, but are treated distinct. Between sentences 5 and 6 there is a maximum coherence gap due to 'airplane' and the airplane's 'flight' not having its lexical relationship recognized. Between sentences 6 an 7 there is again a maximum coherence gap since there is no relationship established between the airplane's 'flight' and its 'landing gear'/'flaps' or the 'accident site'. Between sentences 7 and 8 'it' is treated as a distinct entity and is not resolved to the 'airplane'. Also, this sentence continues talking about parts of the airplane (e.g., 'propellor'), which do not get lexically related to the airplane's 'landing gear' or 'flaps'. Further, the 'impact' does not get resolved to 'accident'. Between sentences 8 and 9 again does not have 'impact' resolved to 'accident', or 'airplane' to 'propellor'. Plus, the verb 'impacted' is not treated as an entity, yet the previous sentence refers to 'impact' as an entity. Between sentences 10 and 11 there is a topic shift back to the crew of the airplane, but 'airplane' and 'flight' are not given a lexical association.

The 2nd permutation (ranked most coherent) in NTSB-3861 is:

1. THE ACCIDENT SITE ELEVATION WAS 4,400 FEET MEAN SEA LEVEL (MSL).

2. OBSERVED EVIDENCE AT THE ACCIDENT SITE INDICATED THE AIRPLANE IMPACTED WITH A HIGH DESCENT RATE AND LOW FORWARD VELOCITY IN A SLIGHT LEFT WING DOWN AND NOSE DOWN ATTITUDE.

3. DAMAGE TO THE PROPELLER INDICATED IT WAS ROTATING AT THE TIME OF IMPACT.

4. NO MECHANICAL ANOMALIES WERE FOUND WITH THE AIRPLANE.

5. AT THE ACCIDENT SITE THE LANDING GEAR WAS FOUND IN THE EXTENDED POSITION AND THE FLAPS WERE FOUND FULLY RETRACTED.

6. THE INSTRUCTOR AND DUAL COMMERCIAL STUDENT WERE SCHEDULED TO PRACTICE COMMERCIAL MANEUVERS A SIMULATED EMERGENCY APPROACH AND LANDING AND SYSTEM/EQUIPMENT MALFUNCTIONS.

7. THE AIRPLANE COLLIDED WITH THE HIGH DESERT TERRAIN DURING MANEUVERING FLIGHT WHILE ON A DUAL INSTRUCTIONAL FLIGHT.

8. RADAR DATA DEPICTED THE AIRPLANE WAS OPERATED IN THE PRESCRIBED PRACTICE AREA CONDUCTING FLIGHT MANEUVERS.

9. THE RADAR DATA ALSO INDICATED THAT THE AIRPLANE 'S MAXIMUM ALTITUDE WAS 8,200 FEET MSL AND ITS LOWEST ALTITUDE WAS 6,500 FEET MSL THE MINIMUM ALTITUDE OBSERVED BY THE RADAR.

10. TRACES OF ANTIHISTAMINE EPHEDRINE AND PHENYLPROPANOLAMINE WERE FOUND IN THE BLOOD DURING THE TOXICOLOGY EXAMINATION OF THE FLIGHT INSTRUCTOR.

11. THE FLIGHT WAS NOT IN RADIO CONTACT WITH ANY GROUND FACILITY.

Permutation 2 entity grid:

```
RADIO - - - - - - - - - - X
ALTITUDE - - - - - - - - - S - -
TOXICOLOGY - - - - - - - - - X -
CONTACT - - - - - - - - - - X
FLAPS - - - - S - - - - - -
GROUND - - - - - - - - - - X
LEFT - X - - - - - - - - -
ATTITUDE - O - - - - - - - - -
MANEUVERS - - - - - X - O - - -
```

```
IT - - S - - - - - - - - -
WHILE - - - - - - X - - - -
TERRAIN - - - - - - X - - - -
AIRPLANE - O - X - - S O X - -
VELOCITY - X - - - - - - - - -
BLOOD - - - - - - - - - X -
SITE S X - - X - - - - - -
TRACES - - - - - - - - - S -
AREA - - - - - - - X - - -
GEAR - - - - S - - - - - -
ELEVATION S - - - - - - - - - -
TIME - - X - - - - - - - -
RADAR - - - - - - - S S - -
NOSE - X - - - - - - - - -
EMERGENCY - - - - - X - - - - -
STUDENT - - - - - S - - - - -
ANOMALIES - - - S - - - - - - -
EXAMINATION - - - - - - - - - X -
POSITION - - - - X - - - - - -
DESERT - - - - - - X - - - -
EPHEDRINE - - - - - - - - - S -
LEVEL O - - - - - - - - - -
MSL X - - - - - - - - - -
EVIDENCE - S - - - - - - - - -
INSTRUCTOR - - - - - S - - - X -
MALFUNCTIONS - - - - - X - - - - -
LANDING - - - - S X - - - - -
ACCIDENT S X - - X - - - - - -
RATE - X - - - - - - - - -
PROPELLER - - X - - - - - - - -
FEET X - - - - - - - X - -
APPROACH - - - - - X - - - - -
FLIGHT - - - - - - X O - X S
```

ANTIHISTAMINE - - - - - - - - - X -

DATA - - - - - - - S S - -

PRACTICE - - - - - - - X - - -

PHENYLPROPANOLAMINE - - - - - - - - - S -

DAMAGE - - S - - - - - - - -

IMPACT - - X - - - - - - - -

SEA O - - - - - - - - - - -

FACILITY - - - - - - - - - - X

Document NTSB-3633 only detected 6 out of 20 permutations less coherent than the original text. The original text is:

1. WHILE IN CRUISE FLIGHT AT 8,500 FEET IN DARK NIGHT CONDITIONS THE AIRCRAFT 'S ENGINE BEGAN EXPERIENCING A GRADUAL LOSS OF RPM.

2. THE PILOT ADJUSTED THE MIXTURE SETTING AND APPLIED CARBURETOR HEAT BUT WAS UNSUCCESSFUL IN RESTORING FULL POWER.

3. SHE ATTEMPTED TO DIVERT TO THE LIMON COLORADO AIRPORT 14 NAUTICAL MILES TO THE EAST BUT THE AIRCRAFT WAS UNABLE TO MAINTAIN ALTITUDE.

4. THE AIRPLANE IMPACTED THE GROUND BECAME AIRBORNE THEN HIT THE GROUND AGAIN IN A ROUGH OPEN FIELD.

5. THE AIRCRAFT THEN NOSED OVER.

6. TWO EXAMINATIONS ON THE ENGINE AND AIRFRAME WERE PERFORMED AND NO MECHANICAL ENGINE ABNORMALITIES OR DISCREPANCIES WERE FOUND.

7. ACCORDING TO REPORTED WEATHER CONDITIONS AT LIMON THE CLOSEST WEATHER FACILITY LOCATED 14 MILES FROM THE ACCIDENT SITE THE TEMPERATURE AND DEW POINT WERE 30 AND 25 DEGREES F. RESPECTIVELY WITH OVERCAST CEILINGS AT 2,500 FEET.

8. ACCORDING TO A CARBURETOR ICING CHART THE CONDITIONS NEAR
   THE TIME OF THE ACCIDENT WERE CONDUCIVE TO THE FORMATION
   OF MODERATE TO SEVERE ICING WITH A CRUISE POWER SETTING.

Original entity grid is:

FACILITY - - - - - - X -
FORMATION - - - - - - - X
ALTITUDE - - O - - - - -
GROUND - - - S - - - -
SETTING - O - - - - - X
CARBURETOR - O - - - - - X
CHART - - - - - - - X
FIELD - - - X - - - -
ENGINE S - - - - S - -
AIRPLANE - - - S - - - -
SITE - - - - - - X -
MIXTURE - O - - - - - -
WEATHER - - - - - - X -
TEMPERATURE - - - - - - O -
NIGHT X - - - - - - -
TIME - - - - - - - X
POINT - - - - - - O -
DEW - - - - - - O -
DEGREES - - - - - - X -
RPM X - - - - - - -
MILES - - X - - - O -
ABNORMALITIES - - - - - S - -
POWER - O - - - - - X
LOSS O - - - - - - -
AIRPORT - - X - - - - -
ACCIDENT - - - - - - X X
COLORADO - - X - - - - -
FEET X - - - - - X -
PILOT - X - - - - - -

FLIGHT X - - - - - - -
HEAT - O - - - - - -
AIRCRAFT X - S - S - - -
CONDITIONS X - - - - - X O
EXAMINATIONS - - - - - S - -
LIMON - - X - - - X -
ICING - - - - - - - X
DISCREPANCIES - - - - - S - -
SHE - - S - - - - -
CEILINGS - - - - - - X -
CRUISE X - - - - - - X
AIRFRAME - - - - - X - -
F. - - - - - - X -

Between sentences 1 and 2 there is a maximum coherence gap due to not being able to detect the lexical relations between 'aircraft' and the aircraft's 'pilot', as well as the 'engine' and the engine's 'settings', 'carburetor', and 'power'. Between sentences 2 and 3 the pronoun 'she' has not been resolved to 'pilot'. Also implicit in sentence 3 is that the pilot was diverting the aircraft that was mentioned in sentence 2. Between sentences 3 and 4 'aircraft' and 'airplane' are treated as two distinct entities when in fact they should be resolved to the same entity. Between sentences 4 and 5, again 'aircraft' and 'airplane' are treated as two distinct entities. Between sentences 5 and 6 the lexical relationship between 'aircraft' and its 'engine' and 'airframe' were not established. Between sentences 6 and 7, there is a new break in the discourse where the weather conditions are now being discussed after the investigation of what caused the airplane failure were discussed. Between sentences 7 and 8, there is a tie between the discussion of the weather conditions and their potential effect on the aircraft's engine. Even though this was the original text, one can easily observe that a more coherent text can be obtained by switching the order of sentences 7 and 8 (which supports the idea proposed in our coherence tree).

Permutation 9 was detected as the most coherent in NTSB-3633:

1. TWO EXAMINATIONS ON THE ENGINE AND AIRFRAME WERE PERFORMED AND NO MECHANICAL ENGINE ABNORMALITIES OR DISCREPANCIES WERE FOUND.

2. THE AIRPLANE IMPACTED THE GROUND BECAME AIRBORNE THEN HIT THE GROUND AGAIN IN A ROUGH OPEN FIELD.

3. WHILE IN CRUISE FLIGHT AT 8,500 FEET IN DARK NIGHT CONDITIONS THE AIRCRAFT 'S ENGINE BEGAN EXPERIENCING A GRADUAL LOSS OF RPM.

4. THE AIRCRAFT THEN NOSED OVER.

5. ACCORDING TO A CARBURETOR ICING CHART THE CONDITIONS NEAR THE TIME OF THE ACCIDENT WERE CONDUCIVE TO THE FORMATION OF MODERATE TO SEVERE ICING WITH A CRUISE POWER SETTING.

6. THE PILOT ADJUSTED THE MIXTURE SETTING AND APPLIED CARBURE-TOR HEAT BUT WAS UNSUCCESSFUL IN RESTORING FULL POWER.

7. SHE ATTEMPTED TO DIVERT TO THE LIMON COLORADO AIRPORT 14 NAUTICAL MILES TO THE EAST BUT THE AIRCRAFT WAS UNABLE TO MAINTAIN ALTITUDE.

8. ACCORDING TO REPORTED WEATHER CONDITIONS AT LIMON THE CLOS-EST WEATHER FACILITY LOCATED 14 MILES FROM THE ACCIDENT SITE THE TEMPERATURE AND DEW POINT WERE 30 AND 25 DEGREES F. RE-SPECTIVELY WITH OVERCAST CEILINGS AT 2,500 FEET.

Permutation 9's entity grid:
F. - - - - - - - X
FACILITY - - - - - - - X
GROUND - S - - - - - -
ALTITUDE - - - - - - O -
SETTING - - - - X O - -
CHART - - - - X - - -
CARBURETOR - - - - X O - -
FIELD - X - - - - - -
ENGINE S - S - - - - -
AIRPLANE - S - - - - - -
SITE - - - - - - - X

MIXTURE - - - - - O - -

WEATHER - - - - - - - X

TEMPERATURE - - - - - - - O

NIGHT - - X - - - - -

TIME - - - - X - - -

POINT - - - - - - - O

DEW - - - - - - - O

DEGREES - - - - - - - X

RPM - - X - - - - -

MILES - - - - - - X O

ABNORMALITIES S - - - - - - -

POWER - - - - X O - -

LOSS - - O - - - - -

AIRPORT - - - - - - X -

ACCIDENT - - - - X - - X

COLORADO - - - - - - X -

FEET - - X - - - - X

PILOT - - - - - X - -

FLIGHT - - X - - - - -

HEAT - - - - - O - -

AIRCRAFT - - X S - - S -

CONDITIONS - - X - O - - X

EXAMINATIONS S - - - - - - -

LIMON - - - - - - X X

ICING - - - - X - - -

DISCREPANCIES S - - - - - - -

SHE - - - - - - S -

CEILINGS - - - - - - - X

CRUISE - - X - X - - -

AIRFRAME X - - - - - - -

FORMATION - - - - X - - -

Document NTSB-3362 detected 7 of the permutations as less coherent than the original. The original document is:

1. THE AIRCRAFT LOST POWER WHILE ON APPROACH APPROXIMATELY 9 MILES SOUTH OF THE AIRPORT.

2. SUBSEQUENTLY IT NOSED OVER DURING A LANDING IN TALL GRASS.

3. THE HOBBS METER SHOWED THE DURATION OF FLIGHT TO BE A TOTAL OF 3.1 HOURS.

4. THE AIRCRAFT HELD 26 TOTAL GALLONS OF FUEL WITH 22.5 USABLE.

5. THE SECOND PILOT REPORTED TO THE MECHANIC/OWNER THAT THEY ' WERE FLYING AT ABOUT 110 MPH THEY WERE NOT AT ALTITUDE AND THEY DID LITTLE LEANING.

6. ACCORDING TO THE CESSNA PILOT OPERATING HANDBOOK (POH) FOR THE 150L THE AIRCRAFT WILL BURN APPROXIMATELY 7 GALLONS PER HOUR AT 2,500 FEET MSL WITH FULL POWER AND USING 92 PERCENT HORSEPOWER.

7. THE POH SPECIFIES THE TOTAL RANGE TO BE 3.2 HOURS FOR THOSE SPECIFIC CONDITIONS.

8. NO FUEL WAS FOUND REMAINING IN OR AROUND THE AIRCRAFT AFTER THE ACCIDENT AND THE SURROUNDING VEGETATION DID NOT SHOW SIGNS OF FUEL SPILLAGE.

9. NEITHER PILOT REPORTED HOW FUEL QUANTITY WAS VERIFIED OR HOW MUCH FUEL THEY HAD ON BOARD BEFORE DEPARTURE.

10. THE ENGINE WAS EXAMINED BY A LICENSED AIRFRAME AND POWER-PLANT MECHANIC WHO FOUND NO DISCREPANCIES.

The entity grid for the original document is:

ALTITUDE - - - - X - - - - -
DURATION - - O - - - - - - -
VEGETATION - - - - - - - S - -
IT - S - - - - - - - -
GALLONS - - - O - O - - - -

ENGINE - - - - - - - - - S
PERCENT - - - - - O - - - -
HANDBOOK - - - - - O - - - -
SPILLAGE - - - - - - - X - -
BOARD - - - - - - - - X -
MECHANIC/OWNER - - - - X - - - - -
MILES X - - - - - - - - -
HOBBS - - S - - - - - - -
METER - - S - - - - - - -
POWER O - - - - X - - - -
TOTAL - - O - - - - - - -
CESSNA - - - - - X - - - -
HOURS - - X - - - O - - -
MPH - - - - X - - - - -
AIRPORT X - - - - - - - - -
HOUR - - - - - X - - - -
LANDING - X - - - - - - - -
ACCIDENT - - - - - - - X - -
THEY - - - - S - - - S -
POWERPLANT - - - - - - - - - - X
GRASS - X - - - - - - - -
QUANTITY - - - - - - - - S -
FEET - - - - - X - - - -
APPROACH X - - - - - - - - -
PILOT - - - - S X - - S -
FLIGHT - - X - - - - - - -
DEPARTURE - - - - - - - - X -
AIRCRAFT S - - S - O - X - -
FUEL - - - X - - - S S -
CONDITIONS - - - - - - X - - -
DISCREPANCIES - - - - - - - - - O
HORSEPOWER - - - - - O - - - -
SIGNS - - - - - - - X - -

POH - - - - - X S - - -
RANGE - - - - - - O - - -
AIRFRAME - - - - - - - - - X

Between sentences 1 and 2 'it' is not resolved to 'aircraft', thus there is a maximum coherence gap (i.e., NOCB) between the two sentences. Between sentences 2 and 3 (as well as sentences 3 and 4), there is no lexical relationship identified between 'it' (the aircraft) and the aircraft's 'flight', which results in a maximum coherence gap again. Between sentences 4 and 5 there is no lexical relationship made between 'pilot' and 'aircraft'. Also, 'they' is treated as a distinct entity since it has not been resolved to 'pilot'. Between sentences 5 and 6 there are many entities introduced that relate to position and speed of the aircraft, as well as details regarding its engine (e.g., 'mph', 'altitude', 'gallons', 'hour', 'feet', 'power', 'horsepower', etc.). Between sentences 6 and 7 'hours' and 'hour' have not been treated as identical entities; they are treated as distinct. Also, 'conditions' refers to the engine and flight particulars mentioned in sentence 6. Between sentences 7 and 8 there is a shift in topic back to the accident rather than continuing an elaboration on the particulars of the aircraft engine. Between sentences 8 and 9, 'aircraft' and 'on board' the aircraft were not resolved to refer to each other. Between sentences 9 and 10 there is actually an abrupt transition from discussing the fuel available on the aircraft to discussing the inspection of the aircraft's engine. Sentence 10 would make a more coherent text if it were placed between sentences 7 and 8 instead.

Permutation 16 was detected as the most coherent in NTSB-3362:

1. SUBSEQUENTLY IT NOSED OVER DURING A LANDING IN TALL GRASS.

2. NEITHER PILOT REPORTED HOW FUEL QUANTITY WAS VERIFIED OR HOW MUCH FUEL THEY HAD ON BOARD BEFORE DEPARTURE.

3. THE ENGINE WAS EXAMINED BY A LICENSED AIRFRAME AND POWER-PLANT MECHANIC WHO FOUND NO DISCREPANCIES.

4. NO FUEL WAS FOUND REMAINING IN OR AROUND THE AIRCRAFT AFTER THE ACCIDENT AND THE SURROUNDING VEGETATION DID NOT SHOW SIGNS OF FUEL SPILLAGE.

5. THE AIRCRAFT HELD 26 TOTAL GALLONS OF FUEL WITH 22.5 USABLE.

6. ACCORDING TO THE CESSNA PILOT OPERATING HANDBOOK (POH) FOR THE 150L THE AIRCRAFT WILL BURN APPROXIMATELY 7 GALLONS PER HOUR AT 2,500 FEET MSL WITH FULL POWER AND USING 92 PERCENT HORSEPOWER.

7. THE AIRCRAFT LOST POWER WHILE ON APPROACH APPROXIMATELY 9 MILES SOUTH OF THE AIRPORT.

8. THE SECOND PILOT REPORTED TO THE MECHANIC/OWNER THAT THEY ' WERE FLYING AT ABOUT 110 MPH THEY WERE NOT AT ALTITUDE AND THEY DID LITTLE LEANING.

9. THE POH SPECIFIES THE TOTAL RANGE TO BE 3.2 HOURS FOR THOSE SPECIFIC CONDITIONS.

10. THE HOBBS METER SHOWED THE DURATION OF FLIGHT TO BE A TOTAL OF 3.1 HOURS.

The entity grid for permutation 16 document is:

ALTITUDE - - - - - - - X - -
DURATION - - - - - - - - - O
VEGETATION - - - S - - - - - -
IT S - - - - - - - - - -
GALLONS - - - - O O - - - -
ENGINE - - S - - - - - - -
PERCENT - - - - - O - - - -
SPILLAGE - - - X - - - - - -
HANDBOOK - - - - - O - - - -
BOARD - X - - - - - - - -
MECHANIC/OWNER - - - - - - - X - -
MILES - - - - - - X - - -
HOBBS - - - - - - - - - S
METER - - - - - - - - - S
POWER - - - - - X O - - -
TOTAL - - - - - - - - - O
CESSNA - - - - - X - - - -

```
HOUR - - - - - X - - - -
MPH - - - - - - - X - -
AIRPORT - - - - - - X - - -
HOURS - - - - - - - - O X
LANDING X - - - - - - - - -
ACCIDENT - - - X - - - - - -
THEY - S - - - - - S - -
POWERPLANT - - X - - - - - - -
GRASS X - - - - - - - - -
QUANTITY - S - - - - - - - -
FEET - - - - - X - - - -
APPROACH - - - - - - X - - -
PILOT - S - - - X - S - -
FLIGHT - - - - - - - - - X
DEPARTURE - X - - - - - - - -
FUEL - S - S X - - - - -
AIRCRAFT - - - X S O S - - -
CONDITIONS - - - - - - - - X -
DISCREPANCIES - - O - - - - - - -
HORSEPOWER - - - - - O - - - -
SIGNS - - - X - - - - - -
POH - - - - - X - - S -
AIRFRAME - - X - - - - - - -
RANGE - - - - - - - - O -
```

Document NTSB-3452 only had 4 permutations detected as less coherent than the original. The original document is:

1. THE NON-INSTRUMENT RATED PRIVATE HELICOPTER PILOT ENCOUNTERED FOG AND LOW CEILINGS WHILE EN ROUTE ON A 132 MILE VFR CROSS COUNTRY FLIGHT.

2. WITNESSES NEAR THE ACCIDENT SITE REPORTED OBSERVING THE HELICOPTER IN A NOSE DOWN ATTITUDE AND HEARING THE SOUND OF IMPACT WITH TREES.

3. VISIBILITY IN THE VICINITY OF THE ACCIDENT SITE WAS REPORTED BY SEVERAL WITNESSES TO BE AS LOW AS 100 FEET AND AS HIGH AS 1/4 OF A MILE.

4. THE CEILING WAS OBSCURED DUE TO FOG.

5. RESIDENTS IN THE AREA REPORTED THAT FOG HAS A TENDENCY TO FORM THAT TIME OF THE YEAR AROUND SUNRISE.

6. PHYSICAL EVIDENCE AND SIGNATURES AT THE ACCIDENT SITE INDI-CATE THAT THE HELICOPTER IMPACTED THE TREES AND TERRAIN IN A 45 DEGREE NOSE LOW ATTITUDE IN A SLIGHT RIGHT TURN.

7. THE ACCIDENT OCCURRED AT 0807 LOCAL TIME AND SUNRISE OCCURRED AT 0619.

8. THE PILOT HAD ACCUMULATED A TOTAL OF 187 FLIGHT HOURS.

9. NO EVIDENCE OF PRE-IMPACT MECHANICAL FAILURE OR MALFUNCTION WAS OBSERVED AT THE ACCIDENT SITE.

10. NO ANOMALIES WERE FOUND DURING THE INVESTIGATION THAT T COULD HAVE PREVENTED NORMAL FLIGHT OPERATION.

The original document's entity grid is:

MALFUNCTION - - - - - - - - X -
WITNESSES - S X - - - - - - -
ATTITUDE - X - - - X - - - -
OPERATION - - - - - - - - - X
TENDENCY - - - - O - - - - -
TERRAIN - - - - - O - - - -
RESIDENTS - - - - S - - - - -
NON-INSTRUMENT S - - - - - - - - -
SITE - X X - - X - - X -
AREA - - - - X - - - - -
TIME - - - - O - X - - -
SOUND - O - - - - - - - -

NOSE - X - - - X - - - -

SUNRISE - - - - X - S - - -

ANOMALIES - - - - - - - - - S

MILE X - X - - - - - - -

VISIBILITY - - S - - - - - - -

FAILURE - - - - - - - - - X -

VICINITY - - X - - - - - - -

INVESTIGATION - - - - - - - - - X

HELICOPTER S O - - - S - - - -

SIGNATURES - - - - - S - - - -

TREES - X - - - O - - - -

EVIDENCE - - - - - S - - S -

TOTAL - - - - - - - O - -

HOURS - - - - - - - X - -

ACCIDENT - X X - - X S - X -

COUNTRY O - - - - - - - - -

FOG O - - - S - - - - -

ROUTE X - - - - - - - - -

VFR S - - - - - - - - -

FEET - - X - - - - - - -

PILOT S - - - - - - S - -

FLIGHT O - - - - - - X - X

DEGREE - - - - - X - - - -

CEILING - - - S - - - - - -

IMPACT - X - - - - - - - -

CEILINGS O - - - - - - - - -

TURN - - - - - X - - - -

YEAR - - - - X - - - - -

Between sentences 1 and 2 there are many entities introduced in establishing the details of the scenario and conditions involved in the helicopter accident. 'Helicopter' is the only entity in common amongst the 17 entities introduced in the first two sentences. Between sentences 3 and 4 there is no lexical relationship established between 'visibility' and 'ceiling' or 'fog', which results in a maximum coherence gap between the sentences. Between sentences 4

and 5 'fog' has been identified in sentence 5 but not in sentence 4, which again results in a maximum coherence gap. Between sentences 5 and 6 there is a transition in focus back to the accident and the helicopter instead of continuing with a description of the weather conditions. Sentence 7 discusses when the accident occurred based on the typical weather conditions discussed in sentence 5. Hence, using our notion of a more coherent transition, one can easily observe that sentence 7 would improve the coherence of the text by placing it between sentences 5 and 6. There is yet another focus shift introduced in sentence 8, where the discussion now involves the pilot and their flight experience. Yet another focus shift occurs in sentence 9, with the new focus being the accident and its causes. Finally, a final focus shift occurs with sentence 10 discusses an investigation of the flight. This text is a fairly scrambled set of sentences exhibiting fairly little coherence, which our system correctly identifies as such. One can easily imagine how various permutations of this text could result in being more coherent than this one. An example is permutation 17, mentioned above. One can observe how sentences 9 and 10 when placed adjacent to each other both realize the entities 'time' and 'sunrise'. Also, sentence 1 in permutation 17 only realizes a single entity, 'ceiling'. Thus there is less of a penalty when such a sentence is either the first or last sentence of a text where it creates only a single maximum coherence gap, as compared to being placed between sentences where it causes two maximum coherence gaps. Thus, our system for this example properly determined that the original text was not very coherent. It can thus be conjectured that a way to analyze the coherence of a text is to permute its sentences and determine if its permutations improve upon the original coherence; if so, then it can be said that the original text was not very coherent to begin with. This approach assumes the two most common problems encountered thus far, coreference resolution and lexical cohesion, are incorporated into the model. Until that happens, one cannot test such a hypothesis quite yet.

Permutation 17 was rated the most coherent in NTSB-3452:

1. THE CEILING WAS OBSCURED DUE TO FOG.

2. NO EVIDENCE OF PRE-IMPACT MECHANICAL FAILURE OR MALFUNCTION WAS OBSERVED AT THE ACCIDENT SITE.

3. VISIBILITY IN THE VICINITY OF THE ACCIDENT SITE WAS REPORTED BY SEVERAL WITNESSES TO BE AS LOW AS 100 FEET AND AS HIGH AS 1/4 OF A MILE.

4. THE PILOT HAD ACCUMULATED A TOTAL OF 187 FLIGHT HOURS.

5. THE NON-INSTRUMENT RATED PRIVATE HELICOPTER PILOT ENCOUN-
   TERED FOG AND LOW CEILINGS WHILE EN ROUTE ON A 132 MILE VFR
   CROSS COUNTRY FLIGHT.

6. NO ANOMALIES WERE FOUND DURING THE INVESTIGATION THAT T COULD
   HAVE PREVENTED NORMAL FLIGHT OPERATION.

7. PHYSICAL EVIDENCE AND SIGNATURES AT THE ACCIDENT SITE INDI-
   CATE THAT THE HELICOPTER IMPACTED THE TREES AND TERRAIN IN
   A 45 DEGREE NOSE LOW ATTITUDE IN A SLIGHT RIGHT TURN.

8. WITNESSES NEAR THE ACCIDENT SITE REPORTED OBSERVING THE HE-
   LICOPTER IN A NOSE DOWN ATTITUDE AND HEARING THE SOUND OF
   IMPACT WITH TREES.

9. THE ACCIDENT OCCURRED AT 0807 LOCAL TIME AND SUNRISE OCCURRED
   AT 0619.

10. RESIDENTS IN THE AREA REPORTED THAT FOG HAS A TENDENCY TO
    FORM THAT TIME OF THE YEAR AROUND SUNRISE.

Permutation 17's entity grid is:

MALFUNCTION - X - - - - - - - -
WITNESSES - - X - - - - S - -
ATTITUDE - - - - - - X X - -
OPERATION - - - - - X - - - -
TERRAIN - - - - - - O - - -
TENDENCY - - - - - - - - - O
RESIDENTS - - - - - - - - - S
SITE - X X - - - X X - -
NON-INSTRUMENT - - - - S - - - - -
AREA - - - - - - - - - X
TIME - - - - - - - - X O
SOUND - - - - - - - O - -
NOSE - - - - - - X X - -

```
SUNRISE - - - - - - - - - S X
ANOMALIES - - - - - S - - - -
FAILURE - X - - - - - - - - -
VISIBILITY - - S - - - - - - -
MILE - - X - X - - - - -
VICINITY - - X - - - - - - -
INVESTIGATION - - - - - X - - - -
TOTAL - - - O - - - - - -
HELICOPTER - - - - S - S O - -
SIGNATURES - - - - - - S - - -
EVIDENCE - S - - - - S - - -
TREES - - - - - - O X - -
HOURS - - - X - - - - - -
ACCIDENT - X X - - - X X S -
COUNTRY - - - - O - - - - -
FOG - - - - O - - - - S
ROUTE - - - - X - - - - -
VFR - - - - S - - - - -
FEET - - X - - - - - - -
PILOT - - - S S - - - - -
FLIGHT - - - X O X - - - -
DEGREE - - - - - - X - - -
CEILING S - - - - - - - - -
IMPACT - - - - - - - X - -
CEILINGS - - - - O - - - - -
TURN - - - - - - X - - -
YEAR - - - - - - - - - X
```

Document NTSB-3402 only recognized half of the permutations as being less coherent than the original document. The original text is:

1. THE PILOT SAID HE DID NOT VISUALLY CHECK THE FUEL BUT STUCK HIS FINGER IN THE FUEL TANK AND FELT FUEL (IN HIS ACCIDENT REPORT HOWEVER HE SAID HE VISUALLY INSPECTED THE FUEL).

2. HE DEPARTED NORTH LAS VEGAS AIRPORT.

3. WHEN THE AIRPLANE WAS APPROXIMATELY 18 MILES WEST OF GRAND JUNCTION THE ENGINE LOST POWER FORCING THE PILOT TO LAND IN AN ONION FIELD.

4. THE AIRPLANE STRUCK A PILE OF ONIONS AND NOSED OVER.

5. THE PILOT SAID THERE WAS 62 GALLONS OF FUEL ON BOARD THE AIRPLANE.

6. ACCORDING TO THE AIRPLANE OWNER 'S MANUAL STANDARD TANKS HOLD 65 GALLONS (58 GALLONS USABLE).

7. THE PILOT SAID HE CRUISED AT 9,500 FEET MSL USED A POWER SETTING OF 18-19 INCHES OF MANIFOLD PRESSURE AND 2300 RPM AND LEANED THE MIXTURE.

8. ACCORDING TO THE CESSNA 207 CRUISE PERFORMANCE CHART FOR 10,000 FEET THE ENGINE SHOULD HAVE DEVELOPED BETWEEN 48 AND 51 PERCENT OF ITS RATED HORSEPOWER (300 HP) AND CONSUMED BETWEEN 10.3 AND 11.0 GALLONS OF FUEL PER HOUR.

9. THIS WOULD GIVE THE AIRPLANE A RANGE OF 695 MILES AND AN ENDURANCE OF BETWEEN 5.3 AND 5.6 HOURS.

10. THE PILOT 'S ACCIDENT REPORT INDICATES THE AIRPLANE HAD BEEN ALOFT APPROXIMATELY 2 HOURS 45 MINUTES.

11. AN FAA INSPECTOR WHO EXAMINED THE AIRPLANE AT THE ACCIDENT SITE SAID THE AIRPLANE HAD BEEN INVERTED PRIOR TO HIS ARRIVAL.

12. WHEN THE AIRPLANE WAS PLACED UPRIGHT HE FOUND NO EVIDENCE OF FUEL IN THE FUEL TANKS FUEL STAINS ON THE GROUND OR FUEL DYE STAINS ON THE WINGS.

13. THERE WAS NO ODOR OF FUEL IN THE AREA.

14. THE PILOT SAID THAT ALTHOUGH HE HAD FLOWN OTHER CESSNA SIN-
GLE ENGINE AIRPLANES THIS WAS HIS FIRST EXPERIENCE FLYING THE
CESSNA 207.

The original text's entity grid is:

GROUND - - - - - - - - - - - - X - -
JUNCTION - - X - - - - - - - - - - -
SETTING - - - - - - X - - - - - - -
CHART - - - - - - - X - - - - - -
GALLONS - - - - O O - O - - - - - -
FIELD - - X - - - - - - - - - - -
ENGINE - - S - - - - O - - - - - O
AIRPLANE - - S X O X - - O S S S - -
PRESSURE - - - - - - X - - - - - - -
SITE - - - - - - - - - - X - - -
AREA - - - - - - - - - - - - X -
MIXTURE - - - - - - O - - - - - - -
LAS - O - - - - - - - - - - - -
WINGS - - - - - - - - - - - X - -
VEGAS - O - - - - - - - - - - - -
PERCENT - - - - - - - O - - - - - -
MANUAL - - - - - X - - - - - - - -
BOARD - - - - X - - - - - - - - -
RPM - - - - - - X - - - - - - -
MILES - - X - - - - - X - - - - -
POWER - - O - - - X - - - - - - -
ENDURANCE - - - - - - - - X - - - - -
EVIDENCE - - - - - - - - - - - O - -
HE S S - - - - X - - - - S - S
CESSNA - - - - - - - X - - - - - O
HOUR - - - - - - - X - - - - - -
FINGER O - - - - - - - - - - - - -
AIRPORT - O - - - - - - - - - - - -
HOURS - - - - - - - - - X X - - - -

EXPERIENCE - - - - - - - - - - - - - - O

ACCIDENT X - - - - - - - - - S X - - -

FAA - - - - - - - - - - - S - - -

HP - - - - - - - X - - - - - -

NORTH - O - - - - - - - - - - - -

INSPECTOR - - - - - - - - - - - S - - -

ARRIVAL - - - - - - - - - - - X - - -

FEET - - - - - - X X - - - - - -

MINUTES - - - - - - - - - - X - - - -

AIRPLANES - - - - - - - - - - - - - O

PILOT S - S - S - S - - X - - - S

TANK X - - - - - - - - - - - - -

TANKS - - - - - S - - - - - X - -

FUEL O - - - X - - X - - - X X -

GRAND - - X - - - - - - - - - -

ONIONS - - - X - - - - - - - - -

DYE - - - - - - - - - - - X - -

INCHES - - - - - - X - - - - - - -

REPORT X - - - - - - - - S - - - -

PILE - - - O - - - - - - - - - -

STAINS - - - - - - - - - - - X - -

ODOR - - - - - - - - - - - - O -

HORSEPOWER - - - - - - - X - - - - - -

OWNER - - - - - X - - - - - - - -

CRUISE - - - - - - - X - - - - - -

PERFORMANCE - - - - - - - X - - - - - -

RANGE - - - - - - - - X - - - - -

Between sentences 1 and 2, even though 'pilot' is not resolved to 'he', there isn't as strong a penalty since the pronominal form exists in both sentences. But a penalty exists in the existence of an additional entity in sentence 1 that is not realized in sentence 2 (e.g., 'pilot' transitions from subject to unrealized). Between sentences 2 and 3, the non-resolved 'he' and 'pilot' do cause a maximum coherence gap. Also, not explicitly mentioned is that the pilot departed from the airport in an airplane, due to ellipsis and lack of recognizing a

lexical relationship. Between sentences 3 and 4, 'onion' was not recognized as an entity in sentence 3, whereas its plural was recognized in sentence 4, thus punishing the coherence transitions score for this seemingly very connected sentence pair. Between sentences 5 and 6, 'fuel' is not explicitly realized in sentence 6 (e.g., '65 gallons of fuel') but is realized in sentence 5, again punishing the transition score. Between sentences 6 and 7, there is a shift in focus from the airplanes fuel tank to details of the flight parameters. Sentence 8 elaborates on the technical specifications of the aircraft's engine and relates it to fuel usage. Hence, sentence 8 would serve the text better in terms of coherence by acting as a bridge between sentences 6 and 7. Sentence 9 should immediately follow sentence 8, but even better sentence 9 should be made subordinate to sentence 8, as in '... 11.0 gallons of fuel per hour, which would give the airplane a range of ...'. Sentence 10 returns to what the pilot reports, but there is a strong tie with sentence 9 in the entities 'airplane' and 'hours'. Sentence 11 produces another topic shift, the inspection of the accident site. Between sentences 11 and 12 there are lexical relationships that need to be captured with regards to the orientation of the airplane (e.g., 'airplane had been inverted' and 'airplane was placed upright'). Sentence 13 realizes 'fuel', which is also realized in sentence 12, but sentence 13 actually continues sentence 12 and should be appended to the end of its conjoined list of evidence i.e., '... he found no evidence of fuel in the fuel tanks, fuel stains on the ground, fuel dye stains on the wings, or odor of fuel in the area'. Sentence 14 is another topic shift that returns to the pilot's account, in this case he describes his aviation experience, which has no relationship whatsoever to the inspection of the accident site mentioned in the previous sentences. This sentence should ideally be placed as either the first or last sentence in the text to minimize the number of maximum coherence gaps. In this case, the sentence was placed last, but the sentence before has no entities in common which results in a maximum coherence gap. Instead, sentence 14 should be placed before sentence 1, which would result in a much more favorable transition since both sentences begin with 'The pilot said'. Hence, it is obvious this text is not maximally coherent in its sentence order, and a more coherent ordering is possible through a permutation. A sample is permutation 16 (shown above). Note that there is a major coherence problem with permutation 16, namely in the sentence 'This would give the airplane a range of 695 miles...' which *must* follow the sentence 'According to the CESSNA 207 cruise performance chart...'. This *result* relationship *cannot* be identified using entity coherence (or any type of cohesion) alone.

Permutation 16 was determined to be the most coherent text in NTSB-3402:

1. HE DEPARTED NORTH LAS VEGAS AIRPORT.

2. THE PILOT SAID HE CRUISED AT 9,500 FEET MSL USED A POWER SETTING OF 18-19 INCHES OF MANIFOLD PRESSURE AND 2300 RPM AND LEANED THE MIXTURE.

3. THE PILOT SAID THAT ALTHOUGH HE HAD FLOWN OTHER CESSNA SINGLE ENGINE AIRPLANES THIS WAS HIS FIRST EXPERIENCE FLYING THE CESSNA 207.

4. WHEN THE AIRPLANE WAS APPROXIMATELY 18 MILES WEST OF GRAND JUNCTION THE ENGINE LOST POWER FORCING THE PILOT TO LAND IN AN ONION FIELD.

5. THE AIRPLANE STRUCK A PILE OF ONIONS AND NOSED OVER.

6. THE PILOT SAID HE DID NOT VISUALLY CHECK THE FUEL BUT STUCK HIS FINGER IN THE FUEL TANK AND FELT FUEL (IN HIS ACCIDENT REPORT HOWEVER HE SAID HE VISUALLY INSPECTED THE FUEL).

7. WHEN THE AIRPLANE WAS PLACED UPRIGHT HE FOUND NO EVIDENCE OF FUEL IN THE FUEL TANKS FUEL STAINS ON THE GROUND OR FUEL DYE STAINS ON THE WINGS.

8. ACCORDING TO THE AIRPLANE OWNER 'S MANUAL STANDARD TANKS HOLD 65 GALLONS (58 GALLONS USABLE).

9. THE PILOT SAID THERE WAS 62 GALLONS OF FUEL ON BOARD THE AIRPLANE.

10. THE PILOT 'S ACCIDENT REPORT INDICATES THE AIRPLANE HAD BEEN ALOFT APPROXIMATELY 2 HOURS 45 MINUTES.

11. THIS WOULD GIVE THE AIRPLANE A RANGE OF 695 MILES AND AN ENDURANCE OF BETWEEN 5.3 AND 5.6 HOURS.

12. THERE WAS NO ODOR OF FUEL IN THE AREA.

13. AN FAA INSPECTOR WHO EXAMINED THE AIRPLANE AT THE ACCIDENT SITE SAID THE AIRPLANE HAD BEEN INVERTED PRIOR TO HIS ARRIVAL.

14. ACCORDING TO THE CESSNA 207 CRUISE PERFORMANCE CHART FOR 10,000 FEET THE ENGINE SHOULD HAVE DEVELOPED BETWEEN 48 AND 51 PERCENT OF ITS RATED HORSEPOWER (300 HP) AND CONSUMED BE-TWEEN 10.3 AND 11.0 GALLONS OF FUEL PER HOUR.

Permutation 16's entity grid is as follows:

GROUND - - - - - - X - - - - - - -
SETTING - X - - - - - - - - - - - -
JUNCTION - - - X - - - - - - - - - -
CHART - - - - - - - - - - - - - X
GALLONS - - - - - - - O O - - - - O
FIELD - - - X - - - - - - - - - -
PRESSURE - X - - - - - - - - - - - -
ENGINE - - O S - - - - - - - - - O
AIRPLANE - - - S X - S X O S O - S -
SITE - - - - - - - - - - - - X -
AREA - - - - - - - - - - - X - -
MIXTURE - O - - - - - - - - - - -
LAS O - - - - - - - - - - - - -
WINGS - - - - - - X - - - - - - -
VEGAS O - - - - - - - - - - - - -
PERCENT - - - - - - - - - - - - - O
MANUAL - - - - - - - X - - - - - -
BOARD - - - - - - - - X - - - - -
RPM - X - - - - - - - - - - - -
MILES - - - X - - - - - - X - - -
POWER - X - O - - - - - - - - - -
ENDURANCE - - - - - - - - - - X - - -
EVIDENCE - - - - - - O - - - - - - -
HE S X S - - S S - - - - - - -
CESSNA - - O - - - - - - - - - X
EXPERIENCE - - O - - - - - - - - - - -
FINGER - - - - - O - - - - - - - -
AIRPORT O - - - - - - - - - - - - -

```
HOURS - - - - - - - - - - X X - - -
HOUR - - - - - - - - - - - - - X
ACCIDENT - - - - - X - - - S - - X -
FAA - - - - - - - - - - - - S -
INSPECTOR - - - - - - - - - - - - S -
NORTH O - - - - - - - - - - - -
ARRIVAL - - - - - - - - - - - - X -
HP - - - - - - - - - - - - - X
FEET - X - - - - - - - - - - - X
MINUTES - - - - - - - - - - X - - - -
AIRPLANES - - O - - - - - - - - - - -
PILOT - S S S - S - - S X - - - -
TANK - - - - - X - - - - - - - -
TANKS - - - - - - X S - - - - - -
FUEL - - - - - O X - X - - X - X
GRAND - - - X - - - - - - - - - -
ONIONS - - - - X - - - - - - - - -
DYE - - - - - - X - - - - - - -
INCHES - X - - - - - - - - - - - -
REPORT - - - - - X - - - S - - - -
PILE - - - - O - - - - - - - - -
STAINS - - - - - - X - - - - - - -
ODOR - - - - - - - - - - - O - -
HORSEPOWER - - - - - - - - - - - - - X
OWNER - - - - - - - X - - - - - -
CRUISE - - - - - - - - - - - - - X
PERFORMANCE - - - - - - - - - - - - - X
RANGE - - - - - - - - - - X - - -
```

Appendix C

Sample of Computer-Generated Summaries

For document set 31001, our system correctly ranked 4 of 10 summary comparisons correctly. The output of each system's summary is presented and discussed below. Our system ranked every summary in this document set as being very incoherent.

System 6

1. Torturers and bombers who carried out atrocities defending or fighting apartheid need counseling to ensure they do not repeat their crimes, The institution exploring apartheid's horrors will issue a report that finds the African National Congress shares blame for human rights violations as it struggled to overcome white rule.

2. One major flaw in the report is the absence of a section accusing F.W. de Klerk, the last apartheid president, of knowing about several bombings after the fact and keeping silent.

3. Already, three major political parties are calling for some kind of new amnesty, and the ANC isn't closing the door.

System 6 Coreference and Salience Output

1. NP#15/17 NP#21/16 atrocities/15 apartheid/14 need/13 counseling/12 crimes/11 institution/10 horrors/9 report/8 African/7 National/6 Congress/5 shares/4 rights/3 violations/2 rule/1

2. flaw/17 NP#12/16 absence/15 section/14 F.W./13 Klerk/12 apartheid/11 president/10 bombings/9 fact/8

3. parties/17 kind/16 amnesty/15 ANC/14 door/13

In System 6's summary, sentences 1 and 2 only share 'apartheid' in common. 'Report' and 'the report' were not resolved to the same entity. 'Bombings' and 'bombers' illustrates a need for lexical relatedness. Sentence 2 and 3 have no entities in common, with sentence 3 appearing to be a non-sequitur with reference to an 'amnesty' that wasn't mentioned previously.

System 16

1. Charles Villa-Vicencio, research head of the Truth and Reconciliation Commission, said ignoring the psychological needs of those who carried out human rights abuses could alienate a potentially dangerous sector of society.

2. Torturers and bombers who carried out atrocities defending or fighting apartheid need counseling to ensure they do not repeat their crimes.

3. The institution exploring apartheid's horrors will issue a report that finds the African National Congress shares blame for human rights violations as it struggled to overcome white rule.

4. Facing a court challenge, the Truth and Reconciliation Commission said Wednesday that it would withhold, at least temporarily, the parts of its final report that implicate South Africa's last apartheid-era president, F.W. de Klerk, in illegal acts.

System 16 Coreference and Salience Output

1. Charles␣␣Villa-Vicencio/16 research/15 head/14 Truth/13 Reconciliation/12 Commission/11 needs/10 rights/9 abuses/8 sector/7 society/6

2. Torturers/16 NP#17/15 atrocities/14 apartheid/13 need/12 counseling/11 crimes/10

3. institution/16 NP#12/15 horrors/14 report/13 African/12 National/11 Congress/10 shares/9 rights/8 violations/7 rule/6

4. court/16 challenge/15 Truth/14 Reconciliation/13 Commission/12 Wednesday/11 NP#5/10 parts/9 report/8 South␣␣Africa/7 apartheid-era/6 president/5 F.W./4 de/3 Klerk/2 acts/1

In System 16's summary, sentences 1 and 2 have no entities in common, though there is a lexical relationship between 'psychological needs' and 'counselling' as well as 'crimes' and 'human rights abuses'. Sentences 2 and 3 also have no entities in common, though there is an inferential relatedness between 'atrocities' and 'apartheid's horrors' and a lexical relatedness between 'crimes' and 'human rights violations'. Between sentences 3 and 4, only 'report' is shared. 'Apartheid' and 'apartheid-era' were not resolved to each other. There is a lexical relationship between 'human rights violations', 'horrors' and 'illegal acts' that is not captured in the entity model, as well as the relationship between 'South Africa' and 'African National Congress'.

System 13

1. (10/26/1998) Commission officials could not immediately be reached for comment.

2. (10/26/1998) The 3,500-page report by the Truth and Reconciliation Commission culminates more than two years of nationwide hearings.

3. (10/31/1998) The 3,500-page report of the South African Truth and Reconciliation Commission, released on 10/29/1998, is the most comprehensive and unsparing examination of a nation's ugly past that any such commission has yet produced.

4. (10/31/1998) Drawing from the commission's own investigations and the testimony of hundreds of applicants for amnesty and 20,000 victims, the report is a detailed look at the crimes of the apartheid era, and blames successive white governments for the vast majority.

System 13 Coreference and Salience Output

1. Commission/14 officials/13 comment/12

2. NP#6/14 years/13 hearings/12

3. report/14 South/13 African/12 Truth/11 Reconciliation/10 Commission/9 examination/8 NP#13/7 past/6

4. NP#13/14 investigations/13 testimony/12 hundreds/11 applicants/10 amnesty/9 victims/8 NP#6/7 look/6 crimes/5 apartheid/4 era/3 governments/2 majority/1

In System 13's summary, sentences 1 and 2 have no shared entities since 'commission' and 'South African Truth and Reconciliation Commission' were not resolved to refer to each other. Between sentences 2 and 3 there were no shared entities, resulting in maximum incoherence. Failures in coreference involve 'report' or 'hearings' with 'examination' or 'report', and 'South African Truth and Reconciliation Commission' with 'Truth and Reconciliation Commission'. Between sentences 3 and 4, the only common entity is 'commission'. A need for lexical relatedness is shown between 'examination' and 'investigations'. ALso, 'report' was not properly resolved to refer to each other.

System 18

1. JOHANNESBURG, South Africa – The deal that South Africa's Truth and Reconciliation Commission offered was simple enough: Confess your crimes, apply for amnesty and you will go free.

2. He concedes that his is a lonely voice.

3. The disagreement stems from Thursday's release of the report by the Truth and Reconciliation Commission on horrors committed during apartheid.

4. The institution exploring apartheid's horrors will issue a report that finds the African National Congress shares blame for human rights violations as it struggled to overcome white rule.

System 18 Coreference and Salience Output

1. NP#16/11 NP#22/10 deal/9 Truth/8 Reconciliation/7 Commission/6 crimes/5 amnesty/4

2. NP#16/11 voice/10

3. disagreement/11 Thursday/10 release/9 report/8 Truth/7 Reconciliation/6 Commission/5 horrors/4 apartheid/3

4. institution/11 NP#2/10 horrors/9 report/8 African/7 National/6 Congress/5 shares/4 rights/3 violations/2 rule/1

In System 18's summary, sentences 1 and 2 have a single shared entity due to 'his' being misresolved to 'Johannesburg'. Sentence 2 causes strong incoherence due to its dangling pronouns 'he' and 'his'. Sentence 2 also causes a maximally incoherent transition between

it and sentence 3 for the same reasons. Between sentences 3 and 4, there a two common entities which causes a boost in the overall coherence score.

System 26

1. Following are excerpts from the final report issued by South Africa's Truth and Reconciliation Commission on Thursday: Madikizela-Mandela herself was responsible for committing such gross violations of human rights. ..

2. In choosing a truth commission rather than a Nuremberg-style trial with which to close out its brutal past, South Africa hoped for an idealistic reconciliation.

3. In its report, the commission called the apartheid state the "primary perpetrator" of gross human rights violations and held such high ranking former officials as former President P.W. Botha, his former army chief, Georg Meiring, and the former intelligence chief, Niel Barnard, responsible.

System 26 Coreference and Salience Output

1. excerpts/14 report/13 South___Africa/12 Truth/11 Reconciliation/10 Commission/9 Thursday/8 Madikizela-Mandela/7 violations/6 rights/5

2. truth/14 commission/13 trial/12 NP#10/11 past/10 NP#8/9 reconciliation/8

3. NP#10/14 report/13 NP#9/12 NP#8/11 perpetrator/10 rights/9 violations/8 NP#4/7 NP#5/6 Georg/5 Meiring/4 intelligence/3 chief/2 Niel___Barnard/1

In System 26's summary, sentences 1 and 2 share the entities 'commission' and 'truth'. The instances of 'South Africa' were not resolved to each other. The lexical relationship between 'gross violations of human rights' and 'brutal past' was not captured. Between sentences 2 and 3, the only entity in common was 'its'. The instances of 'commission' failed to be resolved to each other. The relationship between 'South Africa' and 'apartheid state' was not captured.

For document set 31033, our system correctly ranked 4 of 10 summary comparisons correctly. The output of each system's summary is presented and discussed below. System 6 was determined to be very incoherent by human evaluators, whereas our system ranked it as the most coherent summary. Similarly, humans ranked System 16 as the most coherent, whereas our system considered it to be very incoherent.

System 6

1. In one 1995 e-mail message, Gates wrote other Microsoft executives: 'I think there is a very powerful deal we can make with Netscape.

2. If Netscape rejected Microsoft's deal, Barksdale testified, the Microsoft team told Netscape that it could expect the technical information to be delayed for three months.

3. After declining Microsoft's offer, Netscape did not obtain the necessary information until three months later, which Barksdale says was "precisely what Microsoft had threatened at the June 21 meeting".

4. Attempting to refute a central allegation in the government's antitrust case, Microsoft Corp. offering to stay out of its way and suggesting that Microsoft invest in Netscape.

System 6 Coreference and Salience Output

1. NP#19/10 Gates/9 Microsoft/8 executives/7 deal/6 netscape/5

2. NP#3/10 NP#2/9 deal/8 NP#10/7 Microsoft/6 team/5 NP#12/4 months/3

3. NP#2/10 offer/9 NP#3/8 NP#12/7 months/6 NP#10/5 June/4 meeting/3

4. allegation/10 government/9 case/8 Microsoft/7 Corp./6 offering/5 NP#3/4 way/3 NP#2/2 netscape/1

In System 6's summary, between sentences 1 and 2 'Netscape' was not properly resolved, nor was the first instance of 'Microsoft' in the second sentence resolved to 'Microsoft' in the first sentence. Between sentences 3 and 4, 'Microsoft' and 'Netscape' were failed to be resolved to each other, resulting in a maximally incoherent transition. After inspecting the text, we are not sure why this summary was ranked so poorly for its coherence by human evaluators (compare to the summary produced by System 16, which humans ranked to be most coherent). This summary appears to be one of the more coherent texts produced. There were quite a few strong entity transitions as determined by our entity model.

System 16

1. After months of noisy prelude, the antitrust trial against the Microsoft Corp. opened in federal court Monday morning with a pointed personal attack on Bill Gates, the company's chairman.

2. Microsoft, the world's leading maker of computer software, said demand for the Windows 98 operating system exceeded its projections, particularly in Japan and Europe.

3. A lawyer for Microsoft Corp. on Thursday portrayed the company's competitor in the Internet software business, Netscape Communications Corp., as a willing and eager participant in deal-making that culminated in a June 1995 meeting at which the government and Netscape now say that Microsoft illegally offered to divide the market.

System 16 Coreference and Salience Output

1. months/18 prelude/17 trial/16 NP#11/15 court/14 Monday/13 morning/12 attack/11 NP#23/10

2. NP#2/18 NP#18/17 projections/16 Japan/15 Europe/14

3. lawyer/18 NP#11/17 Thursday/16 competitor/15 Internet/14 software/13 business/12 Netscape/11 Communications/10 Corp./9 participant/8 deal-making/7 June/6 meeting/5 government/4 netscape/3 NP#2/2 market/1

In System 16's summary, sentences 1 and 2 shared no entities due to 'Microsoft' not being properly resolved. Also missing was the relationship between 'Windows 98', 'Microsoft', and 'Bill Gates'. Sentences 2 and 3 only had the mention of 'Microsoft' in common, although it was the final instance in the second sentence, and not the first, that was resolved properly, resulting in a poor sentence transition.

System 13

1. (10/18/1998) SAN FRANCISCO – In a recent move that indicated he has no intention of backing away from his company's aggressive business tactics, Microsoft Corp.'s chairman, Bill Gates, has proposed acquiring the software business behind 3Com Corp.'s popular Palm Pilot hand-held computer, according to people who have been briefed on the discussions.

2. (10/18/1998) The antitrust suit, initiated last October, originally focused on Microsoft's linking of its Internet Explorer World Wide Web browser with the industry-dominant Windows 95 operating system.

3. (10/19/1998) Microsoft replies that the prosecutors are misinterpreting a routine business meeting.

4. (10/20/1998) Demonstrating yet again the power of its franchise, the Microsoft Corp.

System 13 Coreference and Salience Output

1. NP#22/17 move/16 intention/15 company/14 business/13 tactics/12 Microsoft_Corp./11 chairman/10 Bill__Gates/9 software/8 3Com/7 Corp./6 Palm/5 Pilot/4 computer/3 people/2 discussions/1

2. suit/17 October/16 NP#2/15 Internet/14 Explorer/13 World/12 Wide/11 Web/10 browser/9 Windows/8 operating/7 system/6

3. NP#2/17 prosecutors/16 business/15 meeting/14

4. power/17 NP#2/16 franchise/15 Microsoft/14 Corp./13

In System 13's summary, sentences 1 and 2 have no shared entities due to the failure of resolving the instances of 'Microsoft' and 'Microsoft Corp'. Also not captured is the inferential relationship between 'Windows 95' and 'software'. Between sentences 2 and 3, the relationship between 'suit' and 'prosecutors' is not captured. Between sentences 3 and 4, there was again a failure to resolve 'Microsoft' with 'Microsoft Corp', resulting in an extra entity in the last sentence.

System 18

1. WASHINGTON _ The New Economy is on trial here along with Microsoft Corp. Sure, the company is the one in the dock.

2. Every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations, is hereby declared to be illegal.

3. SAN FRANCISCO _ In a recent move that indicated he has no intention of backing away from his company's aggressive business tactics, Microsoft Corp'.s chairman, Bill Gates, has proposed acquiring the software business behind 3Com Corp'.s popular Palm Pilot hand-held computer, according to people who have been briefed on the discussions.

System 18 Coreference and Salience Output

1. WASHINGTON/17 NP#7/16 trial/15 Microsoft/14 Corp./13 Sure/12

2. contract/17 combination/16 form/15 trust/14 conspiracy/13 restraint/12 trade/11 commerce/10 states/9 nations/8

3. SAN/17 FRANCISCO/16 _/15 move/14 intention/13 NP#7/12 business/11 tactics/10 NP#6/9 software/8 3Com/7 Corp'/6 Palm/5 Pilot/4 computer/3 people/2 discussions/1

In System 18's summary, sentences 1 and 2 have no entities in common, though there is a relationship between 'trial' and 'contract'. Between sentences 2 and 3, there again are no shared entities, though there is a lexical relationship between 'business' and 'commerce'.

System 26

1. The government says Netscape rejected Microsoft's collusion pact, while Microsoft replies that the government is misrepresenting what was a routine business meeting.

2. According to the government, Microsoft offered to make an investment in Netscape and give Netscape's software developers crucial technical information about the Windows operating system if Netscape would agree not to make a browser for Windows 95 operating system, which Microsoft released two months later.

3. Microsoft will offer its own opening remarks on Tuesday.

4. In court Thursday, Microsoft presented memos and e-mail messages suggesting that Netscape had eagerly sought the June meeting.

System 26 Coreference and Salience Output

1. NP#17/12 NP#2/11 NP#6/10 collusion/9 pact/8 business/7 meeting/6

2. NP#17/12 NP#6/11 investment/10 NP#2/9 software/8 developers/7 information/6 Windows/5 operating/4 system/3 browser/2 months/1

3. NP#6/12 remarks/11 Tuesday/10

4. court/12 Thursday/11 NP#6/10 memos/9 e-mail/8 messages/7 NP#2/6 June/5 meeting/4

In System 26's summary, this had strong entity coherence throughout. Humans ranked this text with an average coherence score of 4.08 out of 7, which appears very low after closer inspection.

For document 30040, System 6 received a low score in our evaluation even though it was ranked very highly by humans. The output from System 6 is:

System 6

1. Taking a major step toward statehood, the Palestinians on Tuesday inaugurated Gaza International Airport, their first gateway to the world, with cheers, tears and an outpouring of patriotism.

2. The first Palestinian commercial flight landed at Amman's Marka Airport on Saturday, inaugurating an air route between Jordan and the autonomous Gaza Strip.

3. Established in 1996, the tiny Palestinian airline was an outgrowth of the accords with Israel on Palestinian self-rule.

4. Israel has threatened to close down the Palestinian-run Gaza airport over a security violation, an Israeli official said Tuesday, a move that could further undermine the already fragile peace process.

System 6 Coreference and Salience Output

1. step/12 statehood/11 palestinians/10 NP#3/9 Gaza___International___Airport/8 NP#27/7 gateway/6 world/5 cheers/4 tears/3 outpouring/2 patriotism/1

2. flight/12 Amman/11 Marka___Airport/10 Saturday/9 air/8 route/7 Jordan/6 Gaza/5 Strip/4

3. airline/12 outgrowth/11 accords/10 NP#7/9 self-rule/8

4. NP#7/12 Palestinian-run/11 Gaza/10 airport/9 security/8 violation/7 official/6 NP#3/5 move/4 peace/3 process/2

In System 6's summary, there are no common entities between sentences 1 and 2 even though "airports" and "flights" are explicitly mentioned. Also, "Palestinian" was not identified as an entity in the second sentence even though "Palestinians" was identified in sentence 1. Between sentences 2 and 3 there again are no entities in common. Between 3 and 4, the

only entity in common is "Israel". This document illustrates the need for lexical cohesion to capture the relationships between "airport", "flight" and "airline".

For document set 31050, our system correctly ranked 6 of 10 summary comparisons correctly. The output of each system's summary is presented and discussed below. This document set did not receive high coherence scores from human evaluators.

System 6

1. One, Qin Yongmin, 45, of the central city of Wuhan, and the other, Xu Wenli, 55, of Beijing, are being held on suspicion of subversive activities.

2. Qin and Xu became known as democracy advocates in the late 1970s.

3. Foreign Ministry spokesman Zhu Bangzao refused to specify what laws were broken or how Xu Wenli and Qin Yongmin endangered the state.

4. The two more prominent dissidents, Xu and Qin, are likely to face a much longer haul since both have been charged with "criminal acts".

5. Xu's and Qin's trouble almost certainly stems from their efforts to gain recognition for the China Democratic Party,

System 6 Coreference and Salience Output

1. NP#13/8 city/7 Wuhan/6 NP#14/5 suspicion/4 activities/3

2. NP#4/8 NP#5/7 democracy/6 advocates/5 1970s/4

3. Foreign/8 Ministry/7 spokesman/6 Zhu␣␣Bangzao/5 laws/4 NP#14/3 NP#13/2 state/1

4. dissidents/8 NP#5/7 NP#4/6 haul/5 acts/4

5. NP#5/8 NP#4/7 trouble/6 efforts/5 recognition/4 China/3 Democratic/2 Party/1

In System 6's summary, there are no shared entities between sentences 1 and 2. The persons "Qin" and "Xu" were not resolved to "Qin Yongmin" and "Xu Wenli" (same problem occurs between sentences 2 and 3). Sentences 2 and 3 also do not have any entities in common (as well as sentences 3 and 4), again resulting in a maximum incoherence transition. Sentences 4 and 5 have two highly salient entities in common, resulting a comparatively stronger coherence link.

System 16

1. With attorneys locked up, harassed or plain scared, two prominent dissidents will defend themselves against charges of subversion Thursday in China's highest-profile dissident trials in two years.

2. Releasing Liu Nianchun appeared to be an attempt by the government to blunt international criticism over Monday's upcoming trial of prominent dissident Xu Wenli.

3. Xu is the third leading member of a would-be opposition political party put on trial for subversion in a three-week crackdown that has seen 30 dissidents arrested or interrogated.

4. The dissident, Yao Zhenxian is a leader of the China Democracy Party, which was formed in June during President Clinton's visit to China.

System 16 Coreference and Salience Output

1. attorneys/11 NP#25/10 charges/9 subversion/8 Thursday/7 NP#1/6 trials/5 years/4

2. Releasing/11 Liu/10 Nianchun/9 attempt/8 government/7 criticism/6 Monday/5 trial/4 NP#14/3

3. NP#14/11 member/10 opposition/9 party/8 trial/7 subversion/6 crackdown/5 dissidents/4

4. Yao/11 Zhenxian/10 leader/9 China/8 Democracy/7 Party/6 June/5 President/4 Clinton/3 visit/2 NP#1/1

In System 16's summary, there are no entities shared between sentences 1 and 2. The coreference system did not resolve "two prominent dissidents" to "dissident". There is only one entity in common between sentences 2 and 3, "Xu", but this is a fairly incoherent transition with the entity changing in salience from least salient to most salient. Between sentences 3 and 4 there are no shared entities in common. "Dissident" in sentence 4 was not identified as an entity since the part-of-speech tagger assigned it the tag of *adjective*. The general overall coherence of the text is questionable since the text begins by mentioning "two prominent dissidents", yet continues to mention three dissidents.

System 13

1. (12/07/1998) The charges that Wang faces, in a trial that will almost certainly be secret and quick, could bring a sentence of five years or more.

2. (12/09/1998) NEW YORK : A Chinese dissident fleeing a new round of arrests of democracy activists in Shanghai arrived here on 12/09/1998 and announced that he and other opponents of the Chinese government plan a demonstration on 12/09/1998 at the United Nations to protest the crackdown.

3. (12/09/1998) Last week the Chinese government arrested 10 members and sympathizers of the China Democracy Party, one of whom, Wang Youcai, is to go on trial Dec. 17.

System 13 Coreference and Salience Output

1. charges/15 Wang/14 trial/13 sentence/12 years/11

2. NEW___YORK/15 _/14 NP#18/13 round/12 arrests/11 democracy/10 activists/9 Shanghai/8 opponents/7 government/6 plan/5 demonstration/4 United/3 nations/2 crackdown/1

3. week/15 government/14 members/13 sympathizers/12 China/11 Democracy/10 Party/9 Wang___Youcai/8 trial/7 Dec./6

In System 13's summary, sentences 1 and 2 have no entities in common. Between sentences 2 and 3, the only shared entity is "government". It can be seen that the coherence of the text can be improved if sentences 1 and 2 had their order reversed, as well as having "Wang" resolved to "Wang Youcai".

System 18

1. China released a respected, but ailing labor rights campaigner from a prison work camp Sunday and immediately sent him into exile in the United States.

2. BEIJING : Protesting the lack of a defense lawyer, the father of a prominent dissident is to seek a delay in his son's subversion trial, scheduled to start on Thursday in the central city of Wuhan.

3. The defendant is Qin Yongmin, 45, a democracy advocate who has spent 10 years in prison and labor camps and has recently promoted an alternative, non-Communist political party.

4. China's government said Thursday that two prominent dissidents arrested this week are suspected of endangering national security : the clearest sign yet Chinese leaders plan to quash a would-be opposition party.

System 18 Coreference and Salience Output

1. NP#8/11 labor/10 rights/9 campaigner/8 prison/7 work/6 camp/5 Sunday/4 exile/3 United__States/2

2. beijing/11 _/10 lack/9 NP#20/8 delay/7 son/6 subversion/5 trial/4 NP#7/3 city/2 Wuhan/1

3. defendant/11 Qin/10 Yongmin/9 democracy/8 advocate/7 years/6 prison/5 labor/4 camps/3 party/2

4. NP#8/11 government/10 NP#7/9 dissidents/8 week/7 security/6 clearest/5 sign/4 leaders/3 opposition/2 party/1

In System 18's summary, sentences 1 and 2 have no entities in common. Between sentences 2 and 3, there are no common entities again. Between sentences 3 and 4, the only entity in common is "party", but it is the least salient entity in both sentences. This text exhibits a need for incorporating lexical cohesion, since entities such as "China", "China's government", "democracy", "Beijing", and "Wuhan" show a connectedness, as well as the entities "security", "prison", "lawyer", "trial", "subversion", "labor camps", and "dissident".

System 26

1. Qin, 44, and Wang, 31, are accused of inciting the subversion of state power, apparently for helping to organize the China Democracy Party.

2. Wang argued that his activities were legal under China's constitution and under an international covenant on political rights that China has signed, Ms. Hu said.

3. As the founder of the suppressed new political party, the China Democracy Party, he publicly announced its charter in June during President Clinton's visit to China.

4. He said he had no information from Chinese authorities about Qin or Wang.

5. Three other Democratic Party organizers were also detained on Monday, although two of them were released early Wednesday.

System 26 Coreference and Salience Output

1. NP#6/9 NP#10/8 subversion/7 state/6 power/5 NP#11/4

2. NP#10/9 activities/8 NP#11/7 constitution/6 covenant/5 rights/4 Ms./3 Hu/2

3. founder/9 party/8 NP#11/7 NP#10/6 charter/5 June/4 President/3 Clinton/2 visit/1

4. NP#10/9 information/8 authorities/7 NP#6/6 Wang/5

5. NP#2/9 Monday/8 Wednesday/7

In System 26's summary, sentences 1 and 2 have two entities in common, but one of the entity pairs were wrongly resolved to each other ("China" and "China Democracy Party"). Between sentences 2 and 3, two entities are common to both sentences. Between sentences 3 and 4, only one entity is common to both sentences, even though the coreference resolves erred with resolving "he" to "Wang" in sentence 4. Between sentences 4 and 5, there are no entities shared between the two.

Bibliography

Achananuparp, Palakorn, Xiaohua Hu and Xiajiong Shen. 2008. The Evaluation of Sentence Similarity Measures. In *Proceedings of the 10th International Conference on Data Warehousing and Knowledge Discovery*, pages 305-316.

Althaus, Ernst, Nikiforos Karamanis, and Alexander Koller. 2004. Computing Locally Coherent Discourses. In *Proceedings of ACL*, pages 399-406.

Angheluta, Roxana, Marie-Francine Moens and Rik De Busser. 2003. K.U. Leuven Summarization System - DUC 2003. *Workshop on Text Summarization, Document Understanding Conference 2003.*

Ariel, Mira. 1988. Referring And Accessibility. *Journal of Linguistics*, 24(1):65-87.

Arnold, Jennifer. 1998. *Reference Form And Discourse Patterns.* Ph.D. Dissertation, Stanford University.

Asher, Nicholas and Alex Lascarides. 2003. *Logics Of Conversation.* Cambridge University Press, Cambridge.

Barzilay, Regina and Mirella Lapata. 2008. Modeling Local Coherence: An Entity-Based Approach. *Computational Linguistics*, 34(1):1-34.

Barzilay, Regina and Lillian Lee. 2004. Catching the Drift: Probabilistic Content Models, with Applications to Generation and Summarization. In *Proceedings of HLT-NAACL*, pages 113-120.

Berger, Adam. 1996. "A Brief Maxent Tutorial" http://www.cs.cmu.edu/afs/cs/user/aberger/www/html/tutorial/tutorial.html.

Carlson, Lynn and Daniel Marcu. 2001. *Discourse Tagging Reference Manual.* ISI, Technical Report ISI-TR-545.

Cherry, Colin. 2008. Cohesive Phrase-based Decoding for Statistical Machine Translation. In *Proceedings of ACL-HLT*, pages 72-80.

Choi, Freddy, Peter Wiemer-Hastings, and Johanna Moore. 2001. Latent Semantic Analysis for Text Segmentation. In *Proceedings of EMNLP*, pages 109-117.

Noam Chomsky. 1981. *Lectures On Government and Binding.* Foris, Dordrecht.

Clark, Herbert H. 1977. Bridging. In *Thinking: Readings in Cognitive Science*, P. N. Johnson-Laird and P. C. Wason, Ed. Cambridge University Press, London and New York. pages 411-420.

Clark, Herbert H. and C.J. Sengul. 1979. In Search of Referents for Nouns and Pronouns. *Memory and Cognition*, 7:35-41.

Clarke, James and Mirella Lapata. 2007. Modelling Compression with Discourse Constraints. In *Proceedings of EMNLP-CoNLL*, pages 1-11.

Collins, Michael. 1997. Three Generative, Lexicalised Models for Statistical Parsing. In *Proceedings of ACL*, pages 16-23.

Conroy, John, Judith Schlesinger, Dianne O'Leary and Jade Goldstein. 2006. Back To Basics: Classy 2006. In *Proceedings of Document Understanding Conference*, vol. 6.

de Beaugrande, Robert Alain de and Wolfgang Dressler. 1981. *Introduction To Text Linguistics.* Longman, London.

Deerwester, Scott, Susan Dumais, George Furnas, Thomas Landauer, Richard Harshman. 1990. Indexing By Latent Semantic Analysis. *Journal of the American Society for Information Science*, 41:391-407.

Di Eugenio, Barabara. 1998. Centering in Italian. In *Centering Theory in Discourse*, Walker, Marilyn, Aravind Joshi, and Ellen Prince, Ed. pages 115-137. Oxford University Press, Oxford.

Dunlavy, Daniel, Sarah Goodman, Mary Ellen Okurowski, Hans van Halteren, John Conroy, Judith Schlesinger, Dianne O'Leary. 2003. Performance Of A Three-Stage System For Multi-document Summarization. *Workshop on Text Summarization, Document Understanding Conference 2003.*

Elsner, Micha and Eugene Charniak. 2007. A Generative Discourse-New Model for Text Coherence. Brown University, Technical Report CS-07-04.

Elsner, Micha, Joseph Austerweil and Eugene Charniak. 2007. A Unified Local and Global Model for Discourse Coherence. In *Proceedings of NAACL-HLT*, pages 436-443.

Fais, Laurel. 2004. Inferable Centers, Centering Transitions, and the Notion of Coherence. *Computational Linguistics*, 30(2):119-150.

Filipova, Katja and Michael Strube. 2004. Extending The Entity-grid Coherence Model To Semantically Related Entities. In *Proceedings of the Eleventh European Workshop on Natural Language Generation.*

Foltz, Peter. 1996. Comprehension, Coherence, And Strategies In Hypertext And Linear Text. In *Hypertext and Cognition*, Dillon Levonen and RJ Spiro Ed. Lawrence Erlbaum Associates, Hillsdale, New Jersey, pages 100-136.

Foltz, Peter, Walter Kintsch, and Thomas Landauer. 1998. Textual Coherence Using Latent Semantic Analysis. *Discourse Processes*, 25(2&3):285-307.

Gernsbacher, Morton Ann. 1990. *Language Comprehension As Structure Building.* Erlbaum, Hillsdale, New Jersey.

Givon, Talmy. 1983. *Topic Continuity In Discourse.* John Benjamins, Amsterdam.

Gordon, Peter, Barbara Grosz, and Laura Gilliom. 1993. Pronouns, Names and the Centering of Attention in Discourse. *Cognitive Science*, 17(3):311-348.

Graesser, Arthur, Danielle McNamara, Max Louwerse, Zhiqiang Cai. 2004. Coh-Metrix: Analysis of Text on Cohesion and Language. *Behavior Research Methods Instruments And Computers*, 36:193-202.

Graff, David. 1995. North American News Text Corpus. Linguistic Data Consortium, Philadelphia

Grefenstette, Gregory. 1994. *Explorations in Automatic Thesaurus Discovery.* Kluwer Academic Publishers, Dordrecht.

Grosz, Barbara, Aravind K. Joshi, & Scott Weinstein. 1995. Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*, 21(2):203-225.

Grosz, Barbara and Candace Sidner. 1986. Attention, Intentions, And The Structure Of Discourse. *Computational Linguistics*, 12(3):175-204.

Gundel, Jeanette. 1998. Centering Theory and the Givenness Hierarchy: Towards a Synthesis. Marilyn Walker, Aravind Joshi, and Ellen Prince, Ed., *Centering Theory in Discourse.* Oxford University Press, Oxford, pages 183-198.

Gundel, Jeanette and Thorstein Fretheim. 2001. Topic and Focus. Laurence Horn and Gregory Ward Ed. *The Handbook of Pragmatic Theory.* Blackwell, Oxford.

Gundel, Jeanette, Nancy Hedberg and Ron Zacharski. 1993. Cognitive Status And The Form Of Referring Expressions In Discourse. *Language*, 69:274-307.

Guo and Stylios. 2003. A New Multi-document Summarization System. *Workshop on Text Summarization, Document Understanding Conference 2003.*

Halliday, M.A.K. and Ruqaiya Hasan. 1976. *Cohesion In English.* Longman Group Ltd, London.

Harris, Zellig. 1952. Discourse Analysis. *Language*, 28(1):1-30.

Hirst, Graeme. 2008. The Future Of Text-Meaning In Computational Linguistics. *Proceedings of the 11th International Conference on Text, Speech and Dialogue* pages 3-11.

Hitzeman, Janet and Massimo Poesio. 1998. Long Distance Pronominalisation And Global Focus. In *Proceedings of ACL*, 36:550-556

Hovy, Eduard. 1988. Planning Coherent Multisentential Text. In *Proceedings of ACL*, pages 163-169.

Hudson-D'Zmura, Susan. 1988. The Structure of Discourse and Anaphor Resolution: The Discourse Center and the Roles of Nouns and Pronouns. Ph.D. Dissertation, University of Rochester, Rochester, NY.

Hurewitz, Felicia. 1998. A Quantitative Look at Discourse Coherence. In *Centering Theory in Discourse*, Walker, Marilyn, Aravind Joshi and Ellen Prince Ed. Oxford University Press, Oxford, pages 273-291.

Joachims, Thorsten. 2002. Optimizing Search Engines Using Clickthrough Data. In *Proceedings of ACM Conference on Knowledge Discovery and Data Mining*, pages 133-142.

Jones, Michael and James Martin. 1997. Contextual Spelling Correction Using Latent Semantic Analysis. In *Proceedings of the 5th Conference on Applied Natural Language Processing*, pages 166-173.

Kameyama, Megumi. 1985. Zero Anaphora: The Case Of Japanese. Ph.D. Dissertation, Stanford University, Stanford, CA.

Kameyama, Megumi. 1986. A Property-sharing Constraint In Centering. In *Proceedings of ACL*, pages 200-206.

Kameyama, Megumi. 1988. Japanese zero pronominal binding, where syntax and discourse meet. In *Papers from the Second International Workshop on Japanese Syntax*, William Poser Ed.

Kameyama, Megumi. 1998. Intrasentential Centering: A Case Study. In *Centering Theory in Discourse*, Walker, Marilyn, Aravind Joshi and Ellen Prince Ed. Oxford University Press, Oxford, pages 89-112.

Kameyama, Megumi, Rebecca Passonneau and Massimo Poesio. 1993. Temporal Centering. In *Proceedings of ACL*, pages 70-77.

Kamp, Hans and Uwe Reyle. 1993. *From Discourse to Logic: Introduction to Model Theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory.* Dordrecht, The Netherlands.

Karamanis, Nikiforos. 2004. Entity Coherence for Descriptive Text Structuring. Ph.D. Dissertation, University of Edinburgh.

Karamanis, Nikiforos. 2007. Supplementing Entity Coherence with Local Rhetorical Relations for Information Ordering. *Journal of Logic, Language and Information*, 16(4):445-464.

Karamanis, Nikiforos and Hisar Manurung. 2002. Stochastic Text Structuring Using The Principle Of Continuity. In *Proceedings of INLG*, 2:81-88.

Karamanis, Nikiforos, Massimo Poesio, Chris Mellish, & Jon Oberlander. 2004. Evaluating Centering-based Metrics Of Coherence For Text Structuring Using A Reliably Annotated Corpus. *Proceedings of ACL*, pages 391-398.

Karamanis, Nikiforos, Massimo Poesio, Chris Mellish, & Jon Oberlander. 2009. Evaluating Centering For Information Ordering Using Corpora. *Computational Linguistics*, 35(1):29-46.

Kehler, Andrew. 2002. *Coherence, Reference, and the Theory Of Grammar.* CSLI Publications, Stanford, CA.

Kibble, Rodger and Richard Power. 2004. Optimizing Referential Coherence in Text Generation. *Computational Linguistics*, 30(4):401-416.

Klare, George. 1974. Assessing Readability. *Reading Research Quarterly*, 10:62-102.

Knott, Alistair, Jon Oberlander, Michael O'Donnell, and Chris Mellish. 2001. Beyond Elaboration: The Interaction of Relations and Focus in Coherent Text. In *Text Representation: Linguistic and Psycholinguistic Aspects*, T. Sanders, J. Schilperoord, and W. Spooren, Ed. John Benjamins, Amsterdam/Philadelphia, pages 181196.

Landauer, Thomas and Susan Dumais. 1997. A Solution to Plato's Problem: the Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 104(2):211240.

Lapata, Mirella. 2003. Probabilistic Text Structuring: Experiments with Sentence Ordering. In *Proceedings of ACL*, pages 545-552.

Lapata, Mirella. 2006. Automatic Evaluation of Information Ordering: Kendall's Tau. *Computational Linguistics*, 32(4):471-484.

Lappin, Shalom and Herbert Leass. 1994. An Algorithm for Pronominal Anaphora Resolution. *Computational Linguistics* 20(4):535-561.

Lascarides, Alex and Jon Oberlander. 1992. Temporal Coherence and Defeasible Knowledge. *Theoretical Linguistics* vol. 18.

Leffa, Vlaidimir. 1992. Clause Processing in Complex Sentences. In *Proceedings of LREC*, pages 937-943.

Lin, Dekang. 1998. Automatic Retrieval and Clustering of Similar Words. In *Proceedings of ACL*, pages 768-774.

Lin, Dekang & Cherry, Colin. 2003. Word Alignment with Cohesion Constraint. In *Proceedings of NAACL-HLT*, pages 49-51.

Madnani, Nitin, Rebecca Passonneau, Necip Fazil Ayan, John M. Conroy, Bonnie Dorr, Judith Klavans, Dianne O'Leary, and Judith Schlesinger. 2007. Measuring Variability in Sentence Ordering for News Summarization. In *Proceedings of the 11th European Workshop on Natural Language Generation.*

Mani, Inderjeet. 2001. *Automatic Summarization.* John Benjamins, Amsterdam/Philadephia.

Mani, Inderjeet, Eric Bloedorn and Barbara Gates. 1998. Using Cohesion and Coherence Models for Text Summarization. *Intelligent Text Summarization Symposium*, pages 69-76.

Mann, William and Sandra Thompson. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text*, 8(3):243-281.

Marcu, Daniel. 2000. *The Theory and Practice of Discourse Parsing and Summarization.* MIT Press, Cambridge, MA.

McCarthy, Diana, Rob Koeling, Julie Weeds, and John Carroll. 2004. Finding Predominant Senses in Untagged Text. In *Proceedings of ACL*, pages 280-287.

McCarthy, Philip, Stephen Briner, Vasile Rus, Danielle McNamara. 2006. Textual Signatures: Identifying Text-Types Using Latent Semantic Analysis to Measure the

Cohesion of Text Structures. In *Natural Language Processing and Text Mining*. A. Kao and S. Poteet Ed. Springer-Verlag, UK.

McNamara, Danielle. 2001. Reading Both High and Low Coherence Texts: Effects of Text Sequence and Prior Knowledge. *Canadian Journal of Experimental Psychology*, 55:51-62.

Miltsakaki, Eleni. 1999. Locating Topics in Text Processing. In *Computational Linguistics in the Netherlands: Selected Papers from the Tenth CLIN Meeting*, pages 127-138.

Miltsakaki, Eleni. 2002. Toward an Aposynthesis of Topic Continuity and Intrasentential Anaphora. *Computational Linguistics* 28(3):319-355.

Miltsakaki, Eleni. 2003. The Syntax-Discourse Interface: Effects of the Main-Subordinate Distinction on Attention Structure. Ph.D. Dissertation, University of Pennsylvania.

Miltsakaki, Eleni & Karen Kukich. 2004. Evaluation of Text Coherence for Electronic Essay Scoring Systems. *Natural Language Engineering* 10(1):25-55.

Mitchell, Jeff and Mirella Lapata. 2008. Vector-based Models of Semantic Composition. In *Proceedings of ACL*, pages 236-244.

Morris, Jane and Graeme Hirst. 1991. Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text. *Computational Linguistics*, 17(1):21-48.

Mosegaard Hanse, Maj-Britt. 1998. *The Function of Discourse Particles: A Study with Special Reference to Spoken Standard French.* John Benjamins, Amsterdam/Philadelphia.

Ng, Vincent and Claire Cardie. 2002. Improving Machine Learning Approaches to Coreference Resolution. In *Proceedings of ACL*, pages 104-111.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing. 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of ACL*, pages 311-318.

Paivio, Allan, John Yuille, and Stephen Madigan. 1968. Concreteness, Imagery and Meaningfulness Values for 925 Words. *Journal of Experimental Psychology Monograph Supplement*, 76(3, Part 2).

Poesio, Massimo, Rosemary Stevenson, Hua Cheng, Renate Henschel, & Janet Hitzeman. 2000. Specifying the Parameters of Centering Theory: A Corpus-Based Evaluation Using Text from Application-Oriented Domains. In *Proceedings of ACL*, 38(1):400-407.

Poesio, Massimo, Rosemary Stevenson, Barbara Di Eugenio, & Janet Hitzeman. 2004. Centering: A Parametric Theory and its Instantiations. *Computational Linguistics*, 30(3):309-363.

Poesio, Massimo, Rosemary Stevenson, Barbara Di Eugenio, & Janet Hitzeman. 2004. Centering: A Parametric Theory and its Instantiations (technical report). University of Essex, NLE Technical Report TN-02-01.

Poesio, Massimo, Rahul Mehta, Axel Maroudas, & Janet Hitzeman. 2004. Learning to Resolve Bridging References. In *Proceedings of ACL*, pages 143-150.

Prince, Ellen. 1981. Toward A Taxonomy Of Given-New Information. In *Radical Pragmatics*. Academic Press, New York, pages 223-255.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language.* Longman Publishing, Oxford.

Renkema, Jan. 2000. Clause Order as a Discourse Marker. *The Berkeley Conference on Dutch Linguistics 1997: The Dutch Language at the Millennium*, Thomas Shannon and Johan Snapper Ed. University Press of America, Lanham, MD, pages 63-77.

Renkema, Jan. 2004. *Introduction To Discourse Studies* John Benjamins Publishing Company.

Salton, Gerard, A Wang, and C Yang. 1975. A Vector-Space Model for Information Retrieval. *Journal of the American Society for Information Science*, 18:613-620.

Schiffrin, Deborah. 1994. *Approaches To Discourse.* Blackwell, Oxford, England.

Schutze, Hinrich. 1998. Automatic Word Sense Discrimination. *Computational Linguistics*, 24(1):97-124.

Schwarm, Sarah and Mari Ostendorf. 2005. Reading Level Assessment Using Support Vector Machines and Statistical Language Models. In *Proceedings of ACL*, pages 523-530.

Siddharthan, Advaith. 2006. Syntactic Simplification and Text Cohesion. *Research on Language & Computation*, 4(1):77-109.

Smyth, Ron and Craig Chambers. 1996. Centering, Parallelism, and Pronoun Interpretation. Poster presented at the 9th CUNY Conference on Sentence Processing, NYC, NY.

Soricut, Radu and Daniel Marcu. 2003. Sentence Level Discourse Parsing Using Syntactic and Lexical Information. In *Proceedings of ACL-HLT*, pages 149-156.

Strube, Michael and Udo Hahn. 1999. Functional Centering: Grounding Referential Coherence in Information Structure. *Computational Linguistics*, 25(3):309-344.

Taboada, Maite and Loreley Hadic-Zabala. 2008. Deciding on Units of Analysis Within Centering Theory. *Corpus Linguistics and Linguistic Theory*, 4(1):63-108.

Teufel, Simone. 2005. Argumentative Zoning for Improved Citation Indexing. In *Computing Attitude and Affect in Text: Theory and Applications*, pages 159-170.

Thagard, Paul and Karsten Verbeurgt. 1998. Coherence as Constraint Satisfaction. *Cognitive Science*, 22:1-24.

Thompson, Sandra and Elizabeth Couper-Kuhlen. 2005. The Clause as a Locus of Grammar and Interaction. *Discourse Studies*, 7(4-5):481-505.

Tofiloski, Milan, Julian Brooke and Maite Taboada. 2009. A Syntactic and Lexical-Based Discourse Segmenter. In *Proceedings of the ACL-IJCNLP*, pages 77-80.

Toglia, Michael, and William Battig. 1978. *Handbook of Semantic Word Norms.* Erlbaum, New York.

Walker, Marilyn. 1998. Centering, Anaphora Resolution, and Discourse Structure. In *Centering Theory in Discourse*, Walker, Marilyn, Aravind Joshi, and Ellen Prince, Ed. pages 401-435. Oxford University Press, Oxford.

Walker, Marilyn, Masayo Iida and Sharon Cote. 1990. Centering in Japanese Discourse. In *Proceedings of COLING*, pages 1-6.

Walker, Marilyn, Masayo Iida and Sharon Cote. 1994. Japanese Discourse and the Process of Centering. *Computational Linguistics*, 20(2):193-233.

Walker, Marilyn and Ellen Prince. 1996. A Bilateral Approach to Givenness: A Hearer-Status Algorithm and a Centering Algorithm. In *Reference and Referent Accessibility*, Thorstein Fretheim and Janet Gundel Ed. John Benjamins, Philedelphia/Amsterdam, pages 291-306.

Wellner, Ben, James Pustejovsky, Anna Rumshisky, Catherine Havasi and Roser Sauri. 2006. Classification of Discourse Coherence Relations: An Exploratory Study Using Multiple Knowledge Sources In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, pages 117-125.

Wolf, Florian and Edward Gibson. 2005. Representing Discourse Coherence: A Corpus-Based Study. *Computational Linguistics*, 31(2):249-287.

Yang, Dongqiang and David M.W. Powers. 2006. Word Sense Disambiguation using Lexical Cohesion in the Context. In *Proceedings of ACL-COLING*, pages 929-936.

Zhou, Liang and Eduard Hovy. 2003. Headline Summarization at ISI. In *Workshop on Text Summarization, Document Understanding Conference 2003*.