A decompositional analysis of discourse relations

MSc Thesis *(Afstudeerscriptie)*

written by

Nikolas Nisidis
(born June 29th, 1986 in Athens, Greece)

under the supervision of **Dr Henk Zeevat**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
January 8th, 2016	Dr Floris Roelofsen
	Dr Maria Aloni
	Dr Raquel Fernández Rovira
	Dr Henk Zeevat

Institute for Logic, Language and Computation

Abstract

Current approaches to discourse coherence hypothesize a number of discourse relations which are used to link sentential units to one another. These relations, also called "coherence" or "rhetorical" relations, are defined either semantically in terms of their truth-conditional content or pragmatically in terms of speaker intention. Our enterprise consists in analyzing the constituting features of discourse relations and, more specifically, in investigating the inferences involved when a speaker or a hearer links two parts of a text. The framework we use to carry out our analysis is based on two main assumptions: agents reason in a Bayesian fashion and discourse is characterized by some sense of topicality. For that reason, we use a version of Update Semantics and combine it with Causal Probabilities along with Argumentation and Questions Under Discussion. Causal Probabilities allow us to model causal and identity inferences the interlocutors make as well as probabilistic inferences having to do with modalities, whereas Questions Under Discussion provide an elegant way to model topicality. We analyze the 32 discourse relations of Rhetorical Structure Theory due to the latter's longevity as a discourse structure theory and its application in a variety of linguistic fields. The main result of our analysis is that discourse relations can be reduced to specific inferences and there is no need to hypothesize any primitive relations constituting these relations. Given the nature of the inferences, it is also possible that a Bayesian interpreter be able to automatically assign a discourse relation just by answering certain questions pertinent to discourse structure.

Contents

Chapter 1

Introduction

In the first part of the chapter, I will present the nature of discourse relations and their relation to coherence, I will state the research question this thesis sets out to provide an answer to and, in the second part of the chapter, I will state the main research question of this thesis.

1.1 Coherence

Inquiries on discourse relations and, consequently, coherence have a rather traditional introduction. They begin with a question that was probably asked by the very first person who dealt with such a topic: *what makes a text a text*? That is, what is the element that differentiates a sequence of random sentences from another set of sentences that can be called *a text*? The answer is structure. But how can this structure be achieved? It can be achieved through discourse relations between sentences; these relations render the text coherent.

By reflecting a text's structure, discourse relations also provide additional information that stems from the inferences made. Consider example (1).

(1) John was late for the meeting. There was a traffic jam.

The combination of these two sentences seem to make sense. However, if one does not see that the two events described are causally linked, one ends up just adding the two events together. The inference that the two facts are causally related gives the text its intended meaning. In other words, the meaning of a set of discourse segments is greater than the meaning of the sum of its parts.

In order for a writer to convey her thoughts, she builds up a text by successively writing sentences. Her intention to be understood subconsciously leads her to link sentences to one another, otherwise she will either be misunderstood or not understood at all. As far as the reader is concerned,

sentences are perceived as succeeding one another in a smooth way. That smoothness comes from the reader drawing a relation that links them. As a matter of fact, even the first sentence of a text that has no precedent is linked in some way to the sentence(s) following it. In any case, should we randomly choose any sentence from a text, it must relate to another sentence; otherwise the text will not be coherent.

Coherence is, thus, a fundamental aspect of communication, for without it language becomes a chaos of sentences. Some researchers such as Hobbs (1976) and Kehler (2002) approach coherence from an anaphoric point of view: anaphoric elements can only be comprehended on the grounds that a text is coherent. For instance, the pronoun of the utterance

(2) John took a train from Paris to Istanbul. He has family there.

can only refer to *John* if one accepts that there is a connection between the two sentences. This need to resolve anaphora within coherence goes as far as trying to link sentences only on the terms of them being consecutive or close to each other as in (3).

(3) ? John took a train from Paris to Istanbul. He likes spinach.

Even though the two sentences initially seem totally unrelated to each other, one is very likely to try to find a relation between them; a fact which is surprising and shows how important coherence is in comprehending language as Hobbs (1979) first noticed. Both examples (2) and (3) are retrieved from the much discussed examples Hobbs (1979) used in his seminal paper.

Halliday and Hasan (1976) were the first to model *lexical coherence*, i.e. cohesion, on the basis of reference as well as other forms of underspecificity in discourse, such as substitution and ellipsis. They, thus, formed their conjunctive relations which they mainly tied to the overt marking of semantic relations.

However, Hobbs (1979) was the first to claim that coherence is a by-product of coreference, for the mechanisms behind anaphora resolution are mainly driven by world knowledge, semantics and inference. Later on, he presented his own taxonomy of discourse relations or coherence relations as he first called them: Occasion, Evaluation, Ground-Figure, Explanation and Expansion (Hobbs, 1990).

Following the hobbsian tradition, Kehler (2002) created a *NeoHumian categorization* of discourse relations divided in *Cause-Result*, *Contiguity* and *Resemblance* relations. His aim was to treat five diverse (as he called them) phenomena: ellipsis, gapping, extraction, anaphora and tense interpretation. Kehler et al. (2008) provide new experimental evidence supporting that coherence driven analysis. That seems to confirm the need of a holistic approach, for the six enterprises, i.e. these phenomena along

with coherence, cannot be separated from intention recognition; candidate solutions for everything need to be stochastically compared by Bayesian interpretation.

Sanders et al.'s (1992) approach was not only motivated by descriptive adequacy but also by psychological plausibility. They consider discourse relations as cognitive entities and they define four cognitive primitives to provide a taxonomy thereof: basic operation, source of coherence, polarity and order. Regarding each primitive, discourse relations are divided in causal and additive, semantic and pragmatic, and positive and negative respectively. The order refers to that of the event and cause for causal relations. All these will be extensively explained in Chapter 2.

To date, one of the most successful discourse relations taxonomies, should we take into account its longevity and its application to various disparate fields, is that of Mann and Thompson (1988): Rhetorical Structure Theory (henceforth RST). In its most recent account, Taboada and Mann (2006b) model coherence through a taxonomy of 30 relations, 6 more than the initial account.

Another approach to coherence with which this thesis shares similar aspirations is that of Segmented Discourse Representation Theory (SDRT) by Asher and Lascarides (2003). This theory provides a semantic-pragmatic framework to model the *Logics of Conversation*, i.e. how interlocutors reason when connecting parts of discourse. SDRT is based on Discourse Representation Theory (DRT) by Kamp and Reyle (1993) but also adds to that a default logic to allow non-monotonic reasoning. In the following chapter, SDRT will be presented in more detail along with the discourse theories mentioned before.

1.2 Research Question

The latest version of the RST website, where the relations are explicitly informally listed with explanations and examples, contains 32 relations forming the so called *Classical RST* (CRST). Other accounts not related to RST, lay out taxonomies containing up to 100 relations (Martin, 1992). Thus, the situation in this field could be characterized as rather unstable and undecided, even regarding fundamental enterprises such as defining the very basic concepts and tools of explicating coherence.

Given that *no single taxonomy seems suitable* (Mann and Thompson, 1988) as one can either divide a relation or merge two or more relations into one, a question that naturally occurs is what underlies discourse relations. That is, what are the basic mechanisms that writers and readers rely on in order to forge and perceive coherence respectively.

What kind of inferences do writers and readers draw? Sometimes, it is

very clear that these inferences reflect causal relations between facts as in (4).

(4) John fell. Bill pushed him.

However, that causality is not always immediately evident as it is in (2). The relation exemplified in (4) is usually tagged as *Result* or *Non-Volitional Cause*. Other relations such as *Result*, *Justification* or *Evidence*, are usually grouped under this category. However, in our exposition this will not be the case as we will see in Chapter 4.

Oftentimes, there might be an argumentative relation, namely when the writer is making an argument except for an assertion, such as in (5), which is usually tagged as *Contrast*, but in Chapter 4 it will be seen that it has more than one reading.

(5) John is tall but Bill is short.

The above converge to the need of a more general characterization of the inferences drawn by readers and writers concerning discourse structure. That characterization will serve not as another taxonomy of discourse relations but as a twofold tool: it will represent the dimensions governing the inferences made by a reader when comprehending discourse as well the inferences made by a writer when planning the discourse.

Our research question is the following: is it possible to see discourse relations as emergent on inferences that are made anyway by interpreters dealing with new information that arrives in through perception or in verbal communication? This would be so if all relations can be seen as bundles of possible outcomes of these inferential processes. That is precisely the treatment that Traum (1994) applied to speech acts.

While this may not be the right characterization of the speaker planning her next utterance, it may be that strategic moves for the speaker just exploit the inferences that a hearer would be making anyway.

The proper treatment of our research question requires the use of some methodological tools as they are summed up below and detailed in Chapter 3.

1.3 Methodology

The tools used to analyze discourse relations are motivated by two main assumptions: i) interlocutors reason in a Bayesian fashion when speaking or hearing; ii) there is a notion of topicality that characterizes discourse.

In Chapter 3, we follow the proposal of Zeevat (2014) and adopt the update semantics of Veltman (1996) whilst adding causal probabilities. These probabilities provide the mechanism to draw causal inferences, e.g. *A*

causes B, as well as identity inferences, e.g. A is B, that relate events and entities described in different sentences. The former inferences are mostly used when analyzing causal and anticausal relations such as *Volitional Cause* and *Condition*, whereas the latter ones are mostly used with identity relations such as *Restatement* and *Summary*.

The probabilistic enhancement also gives us the chance to formalize epistemic modalities such as *must* and *may*. These formalizations are proven to be quite useful when analyzing relations that explicitly refer to evidence that a speaker has supporting his beliefs as in the Evidence relation.

Another capacity of our probabilistic semantics is the integration of argumentation as introduced by Anscombre and Ducrot (1983) and further developed by Winterstein (2009). For instance, *Concession*, namely when a speaker compares arguments and prefers one, can easily be described by combining arguments with probabilities.

The other assumption crucial to our approach, i.e. that utterances have a topic and interlocutors tend to stick to a topic, leads us to incorporate a strictly defined notion of topichood in our framework. Roberts's (2012) questions under discussion, i.e. implicit or explicit questions that set a topic, are very useful in handling topics. For instance, questions under discussion are relevant when analyzing relations pertinent to content elaboration such as the Sequence (or Narration) relation.

These tools will be described in detail in Chapter 3 and they will be used to analyze the RST relations. We use the list of the RST relations, for it is the most extensive and most used list of discourse relations in various fields such as corpus linguistics and text linguistics.

1.4 Goal of the thesis

This thesis's goal is to analyze the discourse relations by decomposing them to basic inferences (also *dimensions*). It also aims to provide a novel analysis of discourse relations, for, as Zeevat (2015) observes, probabilistic reasoning in language has not been employed to a satisfying degree as yet.

The ultimate goal of the thesis is to contribute to the debate about the nature of discourse relations by exposing the nature of the inferences that those relations trigger. This will eventually provide a better insight and a deeper understanding of the way discourse structure works be it from a speaker or a hearer view.

Chapter 2

State of the Art

2.1 Naming discourse relations

Researchers in the field of discourse coherence have not yet reached a consensus with respect to terminology. Different approaches to discourse usually employ a different name for the relations. The most prevalent name is that of *coherence* relations. This name carries a preconception of what these relations stand for and what they aim at.

Discourse relations are also named rhetorical relations by Mann and Thompson (1988) and Asher and Lascarides (2003), referring not to rhetoric as such, but to an early taxonomy of different relations between ideas (as expressed in speech or text) by Aristotle in his *Metaphysics*: e.g. part to whole, same to opposite. Again, that naming of relations implies at least remotely that when one uses these relations, her rhetoric goal is to *show* something or make her speech or text more artful.

In this thesis, I use the term *discourse relations* which is mainly used in the field of Artificial Intelligence (see Webber et al., 2012) as a more general and neutral one, for it does not presuppose any other overt or hidden purpose that the relations might have. What primarily characterizes these relations is that they relate discourse parts.

2.2 What are discourse relations

Should one wish to answer this question, one must inquire how people approach these relations. An uncontroversial claim is that sentences which form what one would call a text are somehow related. But where does such a relation come from?

Most researchers either implicitly or explicitly acknowledge that relations are not the units of connecting discourse parts but they are formed either by synthesizing different primitives (Sanders et al., 1992) or by being

deduced by more general relation categories (Kehler, 2002). An alternative expression of the former argument is that relations' structural elements are to be found in different layers of knowledge organization (Hobbs, 1976). The reasoning of Asher and Lascarides (2003) is similar to Hobbs's (1976) but they differ in the way the former use constrained representations of defeasible knowledge.

Although Mann and Thompson (1988) and their Rhetorical Structure Theory offer a plethora of relations that seem unitary, the way these are formed is by combining four properties (Mann and Thompson call them *fields*) that can be viewed as the basic elements of each relation. These are the semantic and pragmatic constraints on the main clause (nucleus in their terminology), the constraints on the secondary clause (satellite), the constraints on the combination of these two, and the illocutionary effect, namely the writer's intention as far as the reader is concerned.

Hence, it is evident to the researchers that there are specific elements that underlie these relations. So, the question is transposed to identifying these elements.

2.3 Approaches to Discourse Relations

In this section, I present some of the discourse relations theories putting an emphasis on how each theorist perceives and construes the discourse relations. The emphasis will also be put on what will be relevant for the next chapter.

2.3.1 Hobbs's Coherence

Hobbs's approach to discourse coherence must be thought as part of his approach to discourse interpretation in general. According to Hobbs et al. (1993), to interpret a sentence one must perform the following actions:

1. Prove the sentence's logical form with any constraints imposed by predicates on their arguments

2. Merge redundancies where possible

3. Make assumptions where necessary

So, the basic idea is that the interpretation of a simple sentence can then be written in a logical form which sources data from a knowledge base. Discourse relations can be modeled in a similar fashion. Let's take a look at some relations described by Hobbs (1976). S_0 and S_1 stand for sentences of a text.

Background Infer from S_0 a description of a system of entities and relations, and infer from S_1 that some entity is placed or moves against that system as a background.

Explanation Infer that the state or event asserted by S_1 causes or could cause the state or event asserted by S_0.

Elaboration Infer the same proposition P from the assertions S_0 and S_1.

As shown above, Hobbs views relations as connecting the inferences made by the speaker and as such to be made by the hearer. These inferences stem from the assertions themselves and, afterwards, what is inferred is connected in the mind of the speaker. Discourse relations are not about connecting the sentences as such but connecting particular inferences. Namely, the sentences trigger certain inferences that the interlocutors make; the discourse relations regard the connection of these inferences.

Hobbs's framework requires a concrete theory of knowledge organization that allows the construing of the relations. Its basic pillars are the following:

1. Logical Notation or Knowledge Representation
   Any natural language has to be translated to some logical language in order to represent knowledge. Hobbs uses first-order predicate calculus.

2. Syntax and Semantic Translation
   This is the process that has natural language as its input and a logical language as its output.

3. Knowledge Encoding
   Knowledge needs to be encoded in order to form a knowledge base. That base is immense but this is not a hindrance in the analysis of Hobbs.

4. Deductive Mechanism
   This is abduction.

5. Discourse Operations or Specification of Possible Interpretations
   Identifying discourse problems and solving them in terms of what are the possible inferences.

6. Specification of the Best Interpretation
   Finding the best solution out of all the possible interpretations.

Hobbs seeks to identify the connection between the knowledge or belief system underlying a text and its interpretation. According to him,

the interpretation of a discourse is intertwined with coherence. Moreover, Hobbs deduces four classes of discourse relations based on the following assumptions:

- The speaker wishes to convey a message
- The message has a goal
- The speaker must connect what she says to what is already known to the hearer
- The speaker should alleviate the hearer's comprehension difficulties

These four classes of relations are:

- Relate events denoted by assertions, e.g. Occasion Relation (Causality).
- Relate what has already been said to the conversation's goal, e.g. Evaluation Relation.
- Relate a discourse part to the hearer's prior knowledge, e.g. Background Relation.
- *Expansion* relations easing the hearer's inference processes, e.g. Parallel Relation.

The assumptions used are rather uncontroversial, but the classes of relations seem virtually arbitrary, for they can be reduced to Hobbs's preconceptions about what relations stand for. One may come up with many other classes of relations based on these assumptions or reduce these four to three. The latter procedure was followed to some extent by Kehler (2002) as it can be seen in the following section.

2.3.2 Kehler's Coherence

Kehler (2002) tries to avoid or at least diminish the arbitrariness that some approaches to discourse relations seem to have by resorting to Hume (1748). Kehler claims there are three classes of relations under which all relations fall: Resemblance, Cause-Effect, and Contiguity. These three main relations are types of *connections among ideas* as Hume posited. The classes differ in two respects: a) the kind of arguments where the coherence constraints are applied; b) the type of inferences underlying this application. For instance, a resemblance relation such as *Contrast* is described as follows.

Contrast Infer $p(a_1, a_2, ...)$ from the assertion of S_1 and $\neg p(b_1, b_2, ...)$ from the assertion of S_2, in which for some property vector $\vec{q}$, $q_i(a_i)$ and $q_i(b_i)$ for all i., where a_i are entities and p_i, q_i are relations between entities corresponding to assertions.

The approach is a variant of that of Hobbs as Kehler admits. However, he uses Hume's insights on connections, thus concluding a better classification of the relations. Kehler's account differs from that of Hobbs's on the basis of relying on more tangible linguistic representation's aspects, such as that of the grammatical and thematic roles that are occupied by potential antecedents.

On the one hand, Kehler's endeavor seeks to provide a working classification of coherence relations but, on the other hand, he motivates the resulting relations by having them derived from more primitive notions. However, it is doubtful hat the notions are indeed as primitive as they should be. For instance, let us take into account how he describes the Resemblance relation (one of the three categories of relations).

Resemblance The hearer establishes such a relation by identifying a *common relation* p applying over a set of entities $a_1, ..., a_n$ from the first sentence and a set of entities $b_1, ..., b_n$ from the second sentence. Then, the hearer performs operations based on *categorization, comparison* and *generalization* on each parallel elements pair. The coherence, thus, is the outcome of a syntactic and semantic process.

The problem with the definition of *Resemblance* is what exactly constitutes a common relation and how common is that relation after all. This circularity is of crucial importance, because it is the heart of the problem: how to explicate the nature of discourse relations without resorting to minor, let us say, relations. Another significant matter rises regarding how these sophisticated operations are exactly carried out, namely no unambiguous way is proposed as to how one would compare two such entities. A rather covert answer by Kehler is that some inference processes relevant to resemblance relations are manifestations of a *more general cognitive ability to reason analogically.* Thus, the *Relations* stem from *common relations* and inferences that can be done thanks to our *cognitive ability.*

Therefore, although Kehler makes a step forward in further formalizing Hobbs' framework, he does not fully analyze the relations' components notwithstanding his claim about doing so. For, his analysis is not adequate to the extent of not being concretely founded and of not discovering the least minimal elements out of which relations, even the common ones in his terminology, can or could be deduced. At this point, it must be noted that Kehler et al. (2008) have brought to light new empirical evidence in support of such a coherence-driven (as they call it) analysis of pronoun interpretation, thus providing a type of foundation to Kehler's account and

consequently to Hobbs's.

Both Hobbs's and Kehler's approach deal with how coreference functions. Namely, the different types of inference processes used to establish different Relations affect the way in which pronouns are interpreted. Although the two approaches incorporate the intention of the speaker when conveying a message, their frameworks and their definitions of relations involve only entities and inferences between them. The following approach, i.e. Rhetorical Structure Theory, brings in the foreground the gricean intentions of the speaker (or writer), according to Taboada (2006).

2.3.3 Rhetorical Structure Theory

In RST of Mann and Thompson (1988), relations are defined based on four considerations:

- Constraints on the nucleus
- Constraints on the satellite
- Constraints on the combination of nucleus and satellite
- Effect (achieved on the text receiver)

As it can be deduced, when two or more sentences are related, the nucleus stands for the sentence the relation points at and it is indispensable; whereas the satellite plays a secondary role; i.e. even if omitted the passage is still understandable by the reader. To better comprehend how RST works, one should examine how a relation is defined, e.g. the relation named *Purpose* of RST which is classified as a causal relation by Mann and Thompson (1988).

Purpose

- Constraint on the nucleus: nucleus is an activity
- Constraint on the satellite: satellite is a situation that is unrealized
- Constraint on the combination of nucleus and satellite: nucleus is a solution to the problem presented in the satellite
- Effect (achieved on the text receiver): The reader recognizes that the activity in the nucleus is initiated in order to realize the satellite

On a first glance, the strict formalization of the previous frameworks is abandoned in favor of a more descriptive and all-embracing need of explicating the available linguistic data. Mann (2005) is more interested in

providing a complete description of relations capable of being used to describe coherence.

Furthermore, discourse relations are divided based on two groupings in RST. One of them regards whether there is one nucleus or many. That grouping is internal in the sense that it only necessitates out of the way the framework is formed and does not relate to anything else. The other is that of presentational and subject matter relations. The former's goal (e.g. *Background, Restatement*) is to increase some inclination in the hearer, such as to strengthen a positive regard for the nucleus. The latter's goal (e.g. *Cause, Result*) is to make the reader recognize the relation at issue.

RST achieves to transition the matter of discourse relations from simple inference processes to a matter of intentions of the speaker. In RST the corresponding inferences are to be made by the hearer recognizing the intention of the speaker. If one utters (6), one uses a subject matter relation.

(6) Bill fell. John pushed him.

What Mann and Thompson (1988) claim is that the intention of the speaker is to make the hearer recognize that intention. However, the problem is that, in this way, RST only conceals the real nature of discourse relations, for it does not explore what exactly the intention amounts to and what it means to recognize the intention. For instance, in the case of *Cause*, recognizing the intention of the speaker is equivalent to drawing the inference that the two events (i.e. entities) are causally related.

Let us put it otherwise. When I utter (6), my intention is to show that the two events are related, after having inferred that they are related. My goal, as far as the hearer is concerned, is not just to draw her attention to recognize my intention but to make her infer there is a causal link between the two sentences. The difference is subtle but significant, because one has to do with just acknowledging an intention and the other with drawing an inference. Consider a speaker uttering (7).

(7) John fell. The weather was nice.

The hearer, according to RST, would have to recognize my intention that the two events are related, but, in fact, she would try to draw an inference connecting the two events, which in that case would be quite difficult and demanding. Taboada (2006) clearly states that there is no one-to-one correspondence between relations and intentions. However, the way intentions are brought in their framework is rather confusing and misleading in explaining the inference processes the interlocutors engage in.

Despite its poor formalization, RST's approach to discourse relations is long-lived, for it provides a relatively clear way of tagging parts of a text with the relations and then making relations trees of the whole, which is very handy when it comes to analyzing corpora. The range of the relations

it offers renders it suitable for explaining coherence in a way that is also computationally implementable.

In Chapter 4, the relations of RST will be examined more closely, as they will be analyzed by the tools that I will provide in Chapter 3. At this point, let us present another theory of discourse relations that is relevant to our approach on the grounds of its explicit inquiring unitary elements that underlie discourse relations.

2.3.4 Sanders et al.'s Cognitive Primitives

This approach relates more to that of Hobbs's than RST but it is brought up here, because of its particular nature which constitutes in finding the underlying primitives of discourse relations. Sanders et al. (1992) begin from noting that two requirements are needed when considering a discourse structure theory: *descriptive adequacy* and *psychological plausibility*. A theory is descriptively adequate if it can describe the structure of all kinds of natural texts, whereas it is psychologically plausible if the discourse relations map to cognitive entities.

More particularly, Sanders et al. claim that a psychologically plausible discourse theory must at least generate *plausible hypotheses* on the role of discourse structure in constructing a cognitive representation. This need essentially sets the goal of their endeavor which is *not* to develop a complete discourse theory but mainly to identify the *primitives* underpinning discourse relations. These primitives are four.

- Basic operation
  Causal as implication or Additive as conjunction

- Source of coherence
  Semantic or Pragmatic

- Order of the segments
  Basic or Nonbasic

- Polarity
  Positive or Negative

For example, the relation which is causal, semantic, basic and positive is that of Cause-consequence, whereas the relation which is additive pragmatic and negative is that of Concession. Let us take a closer look to these cognitive primitives (as Sanders et al. call them).

Basic operation An additive operation exists as long as a conjunction relation $P\&Q$ is deductible between two discourse segments, i.e. if all that can be inferred is that the discourse parts are true for the speaker. A causal operation exists as long as an implication relation $P \rightarrow Q$ can be deduced between two segments of discourse, where P is the antecedent and Q is the consequent. In every case, P and Q are propositions expressed by assertions S_1 and S_2 respectively.

This definition, as Sanders et al. notice, fails to correctly predict what happens in (8).

(8) If Sweden is larger than Denmark, then Jurki is older than Lauri.

Although the antecedent is true, it does not relate to the consequent whatsoever. Hence, Sanders et al. are compelled to add that the basic operation is intended to be germane to relevant implication. A conditional "if P then Q" is considered true only if P is relevant to Q's conclusion. So, how is this relevance determined? This is a quite unexplained aspect of Sanders et al.'s framework. As a matter of fact, the existence of such a requirement indicates that the basic operation primitive can be analyzed further and be subject to a more elementary notion, that of relevance. That would doubtless mean that the basic operation ceases to be a primitive.

However, even if one acknowledges that the primitives must rely on some deeper concept of speech such as relevance, the question is how is this relevance measured and processed in such a framework. Sanders et al. provide little or no insight on how propositions which correspond to sentences can be deemed as relevant or not.

Source of coherence A relation is semantic should the discourse segments be related in terms of their propositional content, whereas it is pragmatic should the discourse segments relate due to the illocutionary meaning of one or both of them.

Again, the source of coherence is associated with a semantic and pragmatic dichotomy which is more or less vague, for there are no formal tools applied as to decide which relation of the two should be established. As in RST, this framework severely lacks a good formalization that would allow it to be approved or refuted. That happens because the goals of Sanders et al. are basically linguistic and having a strict semantics-pragmatics framework, that could, for example, be computationally implemented is not among their priorities.

The primitives of order and polarity are simple, although the former is a bit problematic. Basic is the order if S_1 expresses P in $P \rightarrow Q$ and S_2 expresses Q in the basic operation, whereas it is nonbasic if the reverse holds. This is rather counterintuitive, for, in speaking, an event is usually succeeded by its cause. For example, it is "more" or equally basic to say

John ate because he was hungry than say *John was hungry and he ate.* With regard to polarity, a relation is positive if S_1 and S_2 function in the basic operation as antecedent P and consequent Q respectively, whereas it is negative if $not - S_1$ or $not - S_2$ function in the basic operation.

As shown above, Sanders et al.'s framework lacks two properties: a) truly analyzing discourse relations to primitives in the sense of having underlying structure that does not (need to) refer to other more general concepts such as relevance and b) having a well-formed formalization that could be computationally valuable. Our approach, which ultimately aims to identify the inferences underlying discourse relations, will try to find solutions to these two demands to the best possible extent.

2.3.5 Segmented Discourse Representation Theory

Segmented Discourse Representation Theory (henceforth SDRT), as it is presented by Asher and Lascarides (2003), uses Discourse Representation Theory, a semantic representation framework developed by Kamp and Reyle (1993), by enriching it with a default logic specially developed for it in order to infer rhetorical relations. SDRT's dynamic updating brings semantics and pragmatics closer, for the non-monotonic reasoning allows the information to be sourced from various domains such as domain knowledge, context and interlocutors' cognitive states.

What differentiates SDRT from Hobbs et al.'s (1993) account is that the former's nature is modular: the different sources contributing information are separated yet interacting. Each one has its own logic and all together are combined with a *glue logic* in which the discourse's logical form is built. Hobbs et al., on the contrary, allows any information from the knowledge domains to be always accessible.

Essentially, one of the main purposes of SDRT is to do away with interpretation as abduction as defined by Hobbs (1990) and its methodological shortcomings. However, SDRT apparently returns to the need of a big database of conceptual knowledge attained not less cryptically than the abduction weights.

SDRT follows Stalnaker (1978) in defining discourse content based on its context change potential and attempts to spot inferences related to content resulting from the requirement of discourse coherence. The discourse becomes completely coherent if all context information is rhetorically connected and anaphoric elements are resolved.

The explanatory mechanism created unifies insights from AI on commonsense reasoning and dynamic semantics in order to explicate a variety of semantic phenomena: presuppositions, resolution of pronouns, disambiguation of word senses and temporal structure. SDRT also covers inferences relating to content such as dialogue disputes and conversational

implicatures.

One of the main claims of SDRT is that discourse relations are essentially speech acts as defined by Searle (1969): they are a fundamental part of meaning and successful communication heavily relies on identifying the speech acts performed. For instance, explaining an action by an Explanation Relation or contrasting two events by a Contrast Relation are to be thought as what interlocutors do with sentences. Hence, coherence can be seen as the requirement that every utterance is connected to another via an illocutionary contribution each one makes.

Let us see how a relation such as *Explanation* is described in SDRT. *Explanation* is a veridical relation in SDRT. Practically, veridical relations are the ones where the semantic content's truth is questioned by neither the hearer nor the speaker. A non-veridical relation is, thus, *Correction*, which is mostly present in dialogue.

Each veridical relation conforms to the *Satisfaction Schema for Veridical Rhetorical Relations* which essentially transforms the context. Let w, w' be possible worlds, f, g variable assignment functions, π_1, π_2 labels for discourse representations structures, K_{π_1}, K_{π_2} segmented discourse structures of these labels, then the schema is the following: $(w, f)[\![R(\pi_1, \pi_2)]\!]_M(w', g)$ iff

$$(w, f)[\![K_{\pi_1} \wedge K_{\pi_2} \wedge \phi_{R(\pi_1,\pi_2)}]\!]_M(w', g)$$

where $\phi_{R(\pi_1,\pi_2)}$ are the special constraints germane to the specific discourse relation $R(\pi_1, \pi_2)$ that holds between two propositions such as the ones in (6). (w, f) is the input context and (w', g) is the output context.

(6) α: Bill fell.
β: John pushed him.

Here, the special constraint of the relation is a meaning postulate called the *Temporal Consequence of Explanation*:

$$\phi_{Explanation(\alpha,\beta)} \Rightarrow (\neg e_\alpha \prec e_\beta) \tag{2.1}$$

$$\phi_{Explanation(\alpha,\beta)} \Rightarrow (event(e_\beta) \Rightarrow e_\beta \prec e_\alpha) \tag{2.2}$$

where e_α is a gloss for the event in α and e_β corresponds to the event in β. These postulates go beyond the typical compositional semantics of the propositions which are presented in SDRT in a box-style language as in DRT.

Considering the above and especially the postulate regarding temporal sequence of explanation, it is evident that SDRT resorts to some kind of world knowledge from which certain relations are abduced. In Chapter 3, we will describe an abduction method relying on Bayesian Interpretation that is more explanatory than the above. Of course, that will be combined

as in SDRT with a semantic framework to allow for simple content to be described in terms of compositional semantics.

Chapter 3

Method of Analysis

In this chapter, I will present and motivate the use of certain methodological tools to analyze the discourse relations. Our decompositional analysis requires the tools that will lay out the fundamental dimensions of discourse relations. Let us consider a simple example from a narrative.

(9) John got out of his car and entered his house.

This sentence describes a series of actions performed by John. As a matter of fact, it is very hard to conceive a situation where there would be no context at all for such a sentence. It could well be the case as in (9) that one starts narrating a story. Thus, the reader expects to read about events. So, we already have some kind of topicality. Information is added with respect to some topic. The first conjunct elaborates on a topic and the second conjunct elaborates further on the first conjunct.

The next observation is that we almost unconsciously draw a link between the two conjuncts of the sentence. It is not a link based only on general similarities, e.g. they both share *John* as the agent, but a link between the events that these sentences describe. Namely, we causally relate the action of *John getting out of his car* to the action of *John entering his house*. It is presumably so that if John hadn't got out of his car, he wouldn't have entered his house either.

Considering the above, a combination of three theories is needed. First, a dynamic semantics framework is needed so as to define the language, the information states and the updating of states. Second, an information representation is needed that would incorporate the notion of topicality. Third, on top of the above, an inferential mechanism is needed in order to draw causal links between events that the sentences refer to.

Unfortunately, this is still not enough. Consider example (10). We are quite certain that the second sentence has the same topic with the first one. Both refer to Bill's actions.

(10) Bill cut down a lemon tree. He destroyed a valuable fruit.

In this example, the kind of the aforementioned topicality is not enough to let us connect the event(s) described and, as it will be observed later, is rather inadequate. On the other hand, we cannot actually claim that the second event is caused by the first as it can be easily seen that both sentences refer to the *same* event. Thus, we need another inference: the identity inference. That is, our system must be economical in introducing new entities and new events in general. This will also be explained in what follows.

Along with that inference, we lastly need an inclusion inference. Namely, it must be possible that we infer that a sentence describes parts of an event or an entity preceding it. There must be a similar economy on hypothesizing that what is succeeded refers to something that is a part of what is preceded.

The last two inferences as described seem to overlap with the concept of a topic but we will show that the latter concept is still useful as an important dimension in discourse relations, for it can still give us a macroscopic picture of what the writer or the speaker writes or talks about.

3.1 Update Semantics with Causal Probabilities

Let us start with our basic update semantics. We define an update system based on the update semantics of Veltman (1996) with an additional dimension, i.e. causal probabilities, as Zeevat (2014) and Zeevat (personal communication, 2015) sketches it. As it will be explicated later, these probabilities are applied with respect to causes as far as learning is concerned. The update system combines a classical update system for first-order logic with a set of probability distributions over the worlds of the information state:

Definition 1. $\langle \mathcal{L}, \Sigma, P_\sigma, [] \rangle$
The language $\mathcal{L}$ corresponds to that of first-order logic with the operators $\wedge, \vee, \neg, \rightarrow,), ($. Additionally, there is a binary predicate: cause(a, b) where a, b are events or states. Σ contains information states σ which are subsets of a set of possible worlds W, 0 is the empty set $\emptyset$ and 1 is the maximal state W. $[]$ is an update function and we denote $\sigma[\phi]$ as the state σ updated by ϕ. P_σ is a set of probability distributions over worlds in σ.

Updates with factual information A eliminate the worlds in which the fact does not hold. Additionally, the distributions conditionalise on A and learn from A as an observation relevant for the causal order and get restricted to the non-eliminated worlds. Information states in this way

support factual information and partial order over the formulas that are not supported, defined by $\varphi < \psi$ iff for all $p \in P_\sigma$, $p(\varphi) < p(\psi)$.

This preference order can be exploited by a Natural Language interpretation system for making forced choices such as the following:

- Is e a new event or is it identical with or part of a given e'?
- Is e caused by or contingent on a given e'?

The preferences can solve the conflict between too much and too little identity and connection. If it works, the result is that we get extra inferred updates such as $e = e'$, $e \subset e'$, $cause(e, e')$, $contingent(e, e')$.

The preference relation can also be used for dealing with argumentation. That is, A is an argument for B in σ iff for all $p \in P_\sigma$, $p(B|A) > p(B)$.

In our system, support is defined as follows.

Definition 2 (Support)**.** *Let σ be an information state, w a world such as $w \in \sigma$.*

- $\sigma \models p$ *iff* $\forall w \in \sigma : p \in w$
- $\sigma \models \bot$ *iff* $s = \emptyset$
- $\sigma \models \varphi \wedge \psi$ *iff* $\sigma \models \varphi$ *and* $\sigma \models \psi$
- $\sigma \models \varphi \vee \psi$ *iff* $\sigma \models \varphi$ *or* $\sigma \models \psi$
- $\sigma \models \varphi \rightarrow \psi$ *iff* $\forall \tau \subseteq \sigma$*: if* $\tau \models \phi$ *then* $\tau \models \psi$
- $\sigma \models \varphi < \psi$ *iff for all* $p \in P_\sigma$, $p(\varphi) < p(\psi)$

Before we proceed in defining the *cause* predicate in our exposition, we may intuitively list the following types of inference, if A is given and B is new:

- B is caused by A
- B is partially caused by A
- B is part of A
- B is A

All of these inferences share some characteristics. They all add information instead of leaving the matter unsettled. Following Lewis (1973) in defining causation, a typical causal relation also carries a counterfactual: namely, if the antecedent were not true, then the consequent would not be the case. A causal relation between A, B implies that the events have to share spatiotemporal properties.

Consider example (9): it must be the case that the two events involving John must have taken place not only in the order of the first one and then the second one but their spatiotemporal distance must have also been considerably small contrary to the case where he would have to park his car far away from his house and, afterwards, walk 1 mile to get there.

Similarly, if an event or an entity is said to be the other's part as in the third type of inference, there must be some restrictions concerning space and time. For instance, if I describe my city, I may start with some general traits of it and then proceed to some details: I could say that its roads are big and then describe a specific road. If I have said the former, then what follows has to relate to what preceded. Not only in terms of the topic but also in terms of the similarities between what is succeeded to what is preceded.

(11) a. My city has big roads. Syngrou is one of its biggest ones having 6 lanes altogether.

b. My city has big roads. Its little streets are very nice.

Example (11a) seems to make more sense than (11b). Although (11b) refers to the same topic, it is evident that the restrictions of similarity between the part and the whole are not there. Thus, we cannot claim that the second sentence refers to the first. However, they do share the same topic, i.e. describing the city. Hence, (11b) cannot be rejected as nonsensical.

As a matter of fact, later, we will observe that examples (11a) and (11b) are characterized by different discourse relations holding between the two clauses. The first one could be conceived as that of *Elaboration*, whereas the second one could be thought as that of *List*; as in listing the city's different traits.

Another feature of all inferences must be that they win from their denial and their competitors regarding probability. So, it must hold that

$$cause(A, B) > \{C \wedge cause(C, B)\} \text{ for all } C \neq A \tag{3.1}$$

$$cause(A, B) > \neg cause(A, B) \tag{3.2}$$

Assuming that B has a cause

$$cause(A, B) > \bigvee_{C \neq A} (C \wedge cause(C, B)) \tag{3.3}$$

There are two cases

- C is given with $cause(C, B)$ being acceptably probable. It is unlikely that $cause(A, B)$ is more probable.

- no C is given with $cause(C, B)$ being acceptably high, then C itself -if it is not predicted- is unlikely so that it is likely that $cause(A, B) \gg (C \wedge cause(C, B))$ which makes it likely that it holds that

$$\{cause(A, B)\} > \{\exists C(C \neq A \wedge cause(C, B))\} \tag{3.4}$$

- **Exception**
  - A has a weak causal power, i.e. A causes B but only rarely
  - There are very many potential causes

Identity As was described before, the identity inference must have the following characteristics. The relation between X, Y denoted with $X > Y$ is used here as X *is more probable than* Y.

$$A = B > A \neq B \tag{3.5}$$

Find so many correspondences between A and B that it be improbable that A is another entity from B.

$$(A \subset B) > (A \not\subset B) \tag{3.6}$$

A fits to what we know of B so that it is improbable that A is not a part of B.

may & must Epistemic modalities can be smoothly integrated in this framework as Zeevat (2013) shows. More specifically, *may* can be formalized as $\phi > \bot$, whereas *must* can be formalized as $\phi > \neg\phi$. Namely, we have that

$$\text{may}(\phi) : \phi > \bot \Leftrightarrow P(\phi) > 0 \tag{3.7}$$

$$\text{must}(\phi) : \phi > \neg\phi \Leftrightarrow P(\phi) > P(\neg\phi) \tag{3.8}$$

3.1.1 Defining cause

Let us examine now, how our binary predicate is defined in order to have a stricter framework that we will use to explicate discourse relations. Firstly, we have to accept two assumptions upon which the predicate $cause(a, b)$ is defined. It cannot be the case that an event has no cause. Although not necessarily unique, everything has a cause.

The second proposition we must assume is that subjects develop intuitive theories or *schemata* with causal regularities. For instance, we know that if we drop a vase on the floor, it will break. However, these causal regularities have exceptions in the sense that the floor might be too soft or the vase might be too strong to break by such an impact.

The updates of causal inferences occur in two situations: if A is the case then

- all probabilities $p(X)$ should change to $p(X|A)$
  Eliminate the distribution of p if $p(A) < 1$ in σ, i.e.

$$p(A) = \sum\nolimits_{w \in \sigma \models A} P(w) \tag{3.9}$$

- A can be an observation subject to change.

Defining causal power The causal power of A, B in world w is denoted as $cp(A, B, w)$. Let θ be a ground unifier. The steps for the update are the following.

1. Look at θ and ϕ, ψ such that $\theta(\phi, \psi) = (A, B)$
   where

$$X = \{g \mid w \models \phi(g)\} \tag{3.10}$$
$$Y = \{g \mid w \models cause(\phi, \psi)[g]\} \tag{3.11}$$

2. Choose a generalization for which $\frac{|Y|}{|X|}$ is maximal and where the numbers are high enough (e.g. more than 10 for X).

3. Set $cp(A, B, w) = \frac{|Y|}{|X|}$

4. Otherwise, undefined.

Now, the inference of an observation O on a distribution p can be defined as

$$p(w|O) = \frac{p(O|w)p(w)}{p(O)} \tag{3.12}$$

Given that O is observed, we have $p(O) = 1$. Additionally, $p(w)$ is also given by the common ground. $p(O|w) = cp(X, O, w)$ where X is the case of O in w.

Update Process $\langle W, P \rangle$ where W is a finite set of finite worlds and P a set of probability distributions over W. $P_{\sigma[\varphi]}$ is defined in three steps.

$$P^0{}_{\sigma[\varphi]} \tag{3.13}$$
$$P^1{}_{\sigma[\varphi]} \tag{3.14}$$
$$P_{\sigma[\varphi]} \tag{3.15}$$

${P^0}_{\sigma[\varphi]}(w) = P_\sigma(w|A)$ as defined. In this step, the influence of A on the strength of causal connections is evaluated, using causal power.

$$ {P^1}_{\sigma[\varphi]} = \{p \restriction_{\sigma[\varphi]} : p \in {P^0}_{\sigma[\varphi]}\} \tag{3.16} $$

$p \restriction$ is the restriction of p to a subset of its domain, i.e. the distributions are no longer defined on the domain σ, but on the subset $\sigma[\varphi]$.

$P_{\sigma[\varphi]} = \{fp : p \in {P^1}_{\sigma[\varphi]}\}$ where $fp(w) = \frac{p(w)}{\Sigma_{w \in \sigma[\varphi]} p(w)}$. In this step, the distributions are normalized back to 1.

3.1.2 Arguments

Bayesian reasoning is not only useful for causal and identity inferences but also for the linguistic argumentation as it was introduced by Anscombre and Ducrot (1983) and extended by Winterstein ((2009), (2012), (2015)). More specifically, Winterstein (2015) shows how arguments can be modeled on probabilistic grounds.

Should we look into what is the consensus of the different accounts of argumentation regarding its properties, we may grasp the importance of that dimension in language. Argumentation has to do with persuasion: the aim of the speaker is to convince the hearer about a case. To achieve that, the speaker creates a syllogism: a conclusion is reached by starting from a set of premises and applying an inferential process. The conclusion is usually labeled as a *goal* in Anscombre and Ducrot's (1983) terminology. Lastly, an argument is defeasible, i.e. it can be refuted by another more convincing argument.

The argumentation dimension is also important because semantics alone cannot account for different uses of a single discourse marker. For instance, there are at least three different uses of *but*: contrastive (12), concessive (13) and argumentative (14).

(12) Jim plays the piano, but Mary plays the guitar.

(13) John is a lawyer but he is honest.

(14) The ring is nice but expensive.

The contrastive case might be easy to label it this way, as when asking *What do your kids play?* Thus, there is no need to assume an additional context. In example (13) though, it is safe to assume that the proposition *lawyers are normally not honest* lies somewhere in the context and the speaker proceeds to a denial of expectation with the second conjunct.

That use of *but* cannot still be generalized, should we examine example (14). The context is highly unlikely, although not altogether impossible, to contain the proposition *nice rings are normally not expensive*. Thus, the speaker tries to make an argument in favor of not buying the ring

by acknowledging one counterargument, comparing them and putting the winning argument in the place of the second conjunct.

A stochastic approximation of the argumentative use of *but* can be thought as follows (Merin, 1999): if a and b are the opposing arguments and c is the argumentative goal, then the presence of a makes c more probable, i.e. $cause(a, c)$ and the presence of b makes $\neg c$ more probable, i.e. $cause(a, \neg c)$. Adding to that the comparison of the two arguments, that can be modeled in detail as it will be shown in the next chapter, along with other uses of *but*.

Concerning the examples above, it must also be noted that example (12) can also accept an argumentative reading similar to that of example (14): if there is a context in which playing the guitar is better than playing the piano; e.g. for when members of a metal band are searching for another musician to join them, then the speaker probably makes an argument in favor of Mary who plays the guitar.

3.2 Questions Under Discussion

What our framework still misses is a notion of topic. It is not pleonastic to assume within our framework as we have described it until now that discourse normally evolves around the same topic and agents tend to keep on discussing the same events and objects. This idea has been expressed early on by von Stutterheim and Klein (1989) who introduced the concept of the *quaestio*: an implicit question that the utterance has to answer. That question might also be explicit or inferred.

Consider example (12): a question that is more or less salient is *Who plays what?* That question, as Büring (2003) has showed with the use of d(iscourse)-trees, can be divided in sub-questions: e.g. *who plays the piano?* and *who plays the guitar*?

Following the stalnakerian tradition, interlocutors engage in conversation to find out how the world is. Thus, it can be said that there always is a *vague* general question or better a conversational goal to be achieved. That big goal is divided in smaller ones, namely more specific questions that are to be answered by the interlocutors, because not all information is relevant at a given moment. The question that is accepted by both parts, i.e. the speaker and the hearer, becomes the discussion's topic, i.e. the immediate question under discussion.

Roberts (2012) developed a pragmatic semantic framework to handle the concept of the question under discussion in a concise way. If there is a conversational goal, then there are strategies to answer questions. Thus, the interlocutors make moves in the discourse; such moves are the assertions and the questions thereof.

The moves are totally ordered with a precedence relation and each move updates the common ground of the interlocutors, so no information that comes up in the discourse is lost or forgotten. In that framework, the question under discussion stack is a function from the moves to ordered subsets of accepted questions.

Definition 3 (Roberts 2012)**.** *Information Structure $\mathcal{D}$ is a tuple*

$$\mathcal{D} = \langle M, Q, A, <, Acc, CG, QUD \rangle \tag{3.17}$$

Where

- *M is the set of moves in the discourse (assertions and questions)*
- *Q is the set of questions, i.e. set of propositions*
- *A is the set of assertions assertions, i.e. set of possible worlds*
- *$<$ is a precedence relation*
- *Acc is the set of accepted moves in M*
- *CG is a function from M to sets of propositions, yielding for each $m \in M$ the common ground of $\mathcal{D}$ just prior to the utterance of m A general requirement for CG is that CG is always updated incrementally.*
- *QUD: the questions-under-discussion stack is a function from M to ordered subsets of $Q \cap Acc$ such that for all $m \in M$*
  1. *For all $q \in Q \cap Acc$, $q \in QUD(m)$ iff*
     (a) *$q < m$*
     (b) *$CG(m)$ fails to entail an answer to q and q has not been determined to be practically unanswerable.*
     (c) *$QUD(m)$ is (totally) ordered by $<$.*
     (d) *for all $q, q' \in QUD(m)$ if $q < q'$ then the complete answer to q' contextually entails a partial answer to q.*

From the definition, we have that questions are entailed, exactly in the sense of discourse trees as we already mentioned. The entailment of questions is actually defined à la Groenendijk and Stokhof (1984): question q_1 entails question q_2 iff answering q_1 yields a complete answer to q_2. Additionally, a question can be contextually entailed, i.e. if the union of the common ground and an answer to a previous question entails it.

As Roberts (2012) explicitly mentions, that pragmatic framework *is intended to be coordinated with a dynamic compositional semantics.* Our

compositional semantics carries a probabilistic aspect: there is a probability distribution over possible worlds in Σ. So, the common ground CG may also contain causal regularities and schemata as the ones described above. For instance, the CG might contain two propositions as "John hit Bill" (proposition a) and "Bill fell" (proposition b) and a causal relation between the events that these propositions describe (A and B respectively): $cause(A, B)$.

The probabilistic approach can well be extended to cover questions too. Roberts, wanting to address the Gricean (1970) Maxim of Relation, characterizes Relevance as follows: a move m is *Relevant* to the QUD iff m is a partial answer to the question or it is part of a strategy to answer the question. Instead of partial and complete answers we can think of possible answers, i.e. a probability distribution over all answers.

However, this is out of the scope of the present endeavor, as we want to keep our analysis as simple as possible. So, here, probabilities will not interact with answers of the various questions under discussion.

Chapter 4

Analysis of discourse relations

In this chapter, I will go through the RST relations as they are defined by Mann (2005), labeled as the *Classical Rhetorical Structure Theory* (all relations are listed in Table 4.1). These relations are the start pointing for the formal reference which is used by researchers working on RST by expanding it as Taboada and Mann (2006b) do or by applying it to various corpora. Taboada and Mann (2006a) have compiled an extensive overview of the applications of RST in which the relations below are used.

The various terms related to RST that will be mentioned in this chapter, have already been explained in Section (2.3.3) of Chapter 2, so the reader can look that part whenever that is necessary.

Table 4.1: Discourse (Rhetorical in RST terminology) relations in RST

Presentational Relations
Antithesis, Background, Concession, Enablement, Evidence, Justify, Motivation, Preparation, Restatement, Summary
Subject Matter Relations
Circumstance, Condition, Elaboration, Evaluation, Interpretation, Means, Non-volitional Cause, Non-volitional Result, Otherwise, Purpose, Solutionhood, Unconditional, Unless, Volitional Cause, Volitional Result
Multinuclear Relations
Conjunction, Contrast, Disjunction, Joint, List, Multinuclear Restatement, Sequence

Our analysis does not follow the grouping above: RST relations are grouped according to how the various events and entities described in the connected sentences are related. It must be noted that Mann and Thompson (1988) present a table of the *organization of the relations definitions*, where relations are grouped on the basis of sharing some characteristics; e.g. *Volitional* and *Non-Volitional Cause* and *Result* respectively along with *Purpose* comprise the *Relations of Cause*. However, we have not

taken into account their categories in analyzing the relations, for we want to be as neutral as possible in our analysis.

1. Causal Relations
   Events or entities described in utterances are arguments of the *cause* relation. For instance, if A is an event described by one utterance and B an event described by another one, then it holds that $cause(A, B)$. That entails that A and B are the case and in addition that B would not have occurred, if A had not.

2. Anticausal Relations
   Events or entities described in utterances are arguments of the $\neg cause$ relation or the *cause* predicate involves an event and a negation of another event, e.g. $cause(A, \neg B)$.

3. Identity Relations
   Events or entities are linked by identity inferences.

4. Question Under Discussion Relations
   The question under discussion and its maintenance is the most crucial aspect of the relation holding among the sentential units.

5. Other Relations
   Relations that do not fall under any of the categories above.

6. Too Vague Relations
   Relations that cannot be analyzed because of their problematic definition in RST.

The terms of *nucleus* and *satellite*, will be used only because RST relations are defined in this manner. In general, we do not adopt the RST terminology, for our goal is to critique the relations and find out the inferences that trigger them. RST definitions also refer to readers and writers but, again, in most cases the the terms *speaker* and *hearer* can be used in their place respectively.

Most relations are exemplified using the example sentences that Mann and Thompson (1988) and Mann (2005) employ. Although some example sentences are not clear, they were all retrieved from those sources, for these are the relations' exemplars according to the official version of RST.

The relations will be analyzed using the tools outlined in Chapter 3: update semantics with causal probabilities and questions under discussion will be our equipment in investigating the relations.

4.1 Causal Relations

The causal relations of RST as we have grouped them are the following: *Condition, Enablement, Means, Purpose, Non-volitional Cause, Volitional Cause, Non-Volitional Result, Volitional Result, Evaluation* and *Interpretation*. As it will be seen later, *Sequence*, analyzed in section 4.3.7, can also be classified as a causal relation.

4.1.1 Condition

In the Condition relation, the satellite describes a hypothetical scenario whose realization decides the realization of the nucleus. The writer's intention is to make the reader recognize the dependence between the nucleus and the satellite. (15) illustrates the relation.

(15) a. Employees are urged to complete new beneficiary designation forms for retirement or life insurance benefits

b. whenever there is a change in marital or family status.

(Mann and Thompson, 1988)

The satellite (15b) is a causal precondition of the nucleus (15a) and it is the case that neither of them have happened, for the situation that they refer must be unrealized. If X is the satellite, then (4.1) should be the case.

$$\sigma[X] \models cause(X, useful(F)) \tag{4.1}$$

That is, a change in the marital status of an employee makes the filling of a relevant form useful. This can be stated in a more general way, namely $cause(X, A)$, where X is the condition that helps in bringing A about.

4.1.2 Enablement

When using the Enablement relation, the writer's intention is to increase the reader's potential ability to carry out the yet unrealized action in the nucleus by comprehending the satellite. Example (16) illustrates such a case. An interesting fact about this relation is that it is never marked (Taboada and Mann, 2006b).

(16) a. Training on jobs: A series of informative, inexpensive pamphlets and books on worker's health discusses such as filing a compensation claim, ionizing radiation, asbestos, and several occupational diseases. [...]

b. For a catalog and order form write WIOES, 2520 Milvia St., Berkeley, CA 95704. (Mann and Thompson, 1988)

Here, the reader is presented with an interesting action and she is then told how to perform it. Let M be the motive of the reader to bring about A, i.e. what is described in the nucleus (16a). If X is the action in the satellite (16b), then the presence of X will help to bring A about. This can be captured by the formula (4.2).

$$\sigma \models cause(X \wedge M, A) > cause(\neg X \wedge M, A) \tag{4.2}$$

4.1.3 Means

The Means relation is used whenever the writer's intention is to make the reader acknowledge that a method or an instrument renders the realization of an event more likely. Consider example (17).

(17) a. ... the visual system resolves confusion
b. by applying some tricks that reflect a built-in knowledge of properties of the physical world. (Mann, 2005)

The nucleus (17a) is an activity and the satellite (17b) is the means to achieve it. This relation is a causal one and is very close to the Condition relation; the difference between them is that the former one refers to a means that is needed to generate an outcome, whereas the latter one refers to a situation necessary for achieving the same effect.

Considering the above, the Means relation may be formalized as the Condition relation. If X, i.e. the satellite, is given and qualifies as a means of achieving A, i.e. the nucleus, then it is quite certain that there is a causal link between them. Consider example (18).

(18) Tom became a millionaire by winning the lottery.

Winning the lottery is a means to become a millionaire, thus a causal link between becoming a millionaire and winning the lottery is established.

$$cause(X, A) \tag{4.3}$$

Here, it must be noted that the means is not necessarily unique in causing A, as there may be other means that help in bringing about A.

Considering the analysis of the relation, *Means* can be merged with *Condition*, as there is a causal link between the events described by the clauses involved.

4.1.4 Purpose

The Purpose relation is used when the writer presents an activity (in the nucleus) that realizes a situation yet unrealized (in the satellite). The writer's intention is that the reader recognize such a dependence, e.g. (19).

(19) a. To see which Syncom diskette will replace the ones you're using now,

b. send for our free "Flexi-Finder" selection guide and the name of the supplier nearest you. (Mann and Thompson, 1988)

The Purpose relation is actually an inverted Means relation, as an activity in the nucleus (19b) is the means to achieve the outcome in the satellite (19a). However, it is a bit more complicated as one must account for the connection between the desire of the interlocutor and the purpose: the desire of the purpose causes the desire of the action and the action causes the purpose.

$$\sigma \models cause(desire(P), desire(A)) \wedge cause(A, P) \tag{4.4}$$

4.1.5 (Non-)volitional Cause/Result

The relations *Non-volitional Cause, Non-volitional Result, Volitional Cause* and *Volitional Result*, seem to overlap in the RST definitions. Let us discuss their difference.

In the non-volitional Cause relation, the nucleus is the non-volitional action caused by the satellite and the focus of the writer, whereas in the non-volitional Result relation the satellite is the non-volitional action caused by the nucleus. So, this distinction is a matter of focus of the writer, i.e. the focus is the nucleus and the satellite can be omitted in each case.

Should we fix the parameter of volition, things are the same as before, but, according to RST, the satellite in the Volitional Cause relation *could have caused* the agent in the nucleus to perform the volitional action. Similarly, the nucleus in the volitional result *could have caused* the satellite.

Although the Cause relation is actually the inverse of the Result one, RST supports that this is not exactly the case. In both the Cause relations, the writer's intention is that the reader recognize the satellite as a cause of the nucleus, whereas, in both the Result relations, the writer's intention is that the reader recognize that the nucleus *could be a cause* of the satellite.

If we carefully examine (20), i.e. how *Non-volitional Result* is exemplified in RST, we can conclude that it describes a typical causal relation. If E_1 and E_2 are the events described by the nucleus (20) and the satellite (20b) respectively, then given E_1, we get that $cause(E_1, E_2)$.

(20) a. The blast, the worst industrial accident in Mexico's history, destroyed the plant and most of the surrounding suburbs.

b. Several thousand people were injured, and about 300 are still in hospital. (Mann and Thompson, 1988)

Considering the RST examples, what really differentiates the Non-volitional Result from the Non-volitional Cause relation is that in the latter

there is always an agent as in (21)'s satellite (21b), whereas that is not necessary in the former case.

(21) a. [...] we've been able to mine our own iron ore, coal, manganese, dolomite, all the materials we need to make our own steel. And because we can mine more than we need,

b. we've had plenty of manganese and iron ore for export.

(Mann and Thompson, 1988)

Probably because of the salience of the agent in the Non-volitional Cause relation, the definition explicitly requires that the satellite should present a situation that, "by means other than motivating a volitional action", causes the situation in the nucleus (Mann and Thompson, 1988). So, in the case of the Cause Relation we can talk about events causing actions, for events are good enough to describe the causes.

Although the difference between *Volitional Result* and *Volitional Cause* is stated in the definitions of RST, the example used to illustrate both of them is the same, that is example (22). Therefore, it is safe to consider them as the same relation, with a different focus each time.

(22) a. Writing has almost become impossible

b. so he had the typewriter serviced and I may learn to type decently after all these years.

(Mann and Thompson, 1988)

A better illustration of the volitional cause is (23). This can be opposed to the non-volitional cause relation, exemplified in (24).

(23) John pushed Bill. He was angry.

(24) John fell. Bill pushed him.

Above, it is easy to spot the volition of the action or to verify its absence. However, sometimes there is an ambiguity concerning the volition of an action as in example (25).

(25) John hit Bill because Bill annoyed him.

The ambiguity of the situation above is the following: either Bill annoys John on purpose, thus, making John want to hit him or John is negatively predisposed towards Bill and it is John's volition to hit Bill. The latter can be modeled as $cause(V, A)$, whereas the former can be presented as

$$\exists V\{cause(E, V) \wedge cause(V, A)\} \tag{4.5}$$

where V is the volition of the agent of the action A, in our example John, and E is the event that caused that volition, in our example Bill annoying John.

4.1.6 Evaluation

The evaluation relation is used whenever the writer has a positive regard for something and wants the reader to recognize that regard of hers. Consider example (26).

(26) a. Features like our uniquely sealed jacket and protective hub ring make our discs last longer. And a soft inner liner cleans the ultra-smooth disc surface while in use.

b. It all adds up to better performance and reliability.

(Mann and Thompson, 1988)

This is a causal relation between the nucleus (26b) and the impression it makes on the writer, expressed in the satellite (26a). This can be modeled as

$$cause(E, like(W, E, d)) \tag{4.6}$$

which means that E (the nucleus) causes the writer to like E to the degree d.

4.1.7 Interpretation

Although this RST relation is originally defined in an obscure manner, should one see the example used, it is evident that the Interpretation relation is a more general Evaluation relation and, as a consequence, a causal relation. Consider example (27).

(27) a. Steep declines in capital spending commitments and building permits, along with a drop in the money stock pushed the leading composite index down for the fifth time in the past 11 months [...]

b. Such a decline is highly unusual at this stage in an expansion.

(Mann and Thompson, 1988)

The writer does not necessarily like what has preceded as she does in the Evaluation relation, but interprets it by relating it to other data available. That connection might have positive or negative connotations. In (28a) the writer might be comparing cars and means of transport used in the past, whereas in (28b) she might be comparing cars to airplanes.

(28) a. An average car can easily reach the speed of 100 kilometers per hour. With such a speed, long distances are covered in a long time.

b. An average car can easily reach the speed of 100 kilometers per hour. With such a speed, long distances are covered in a short time.

In example (28a), the writer has a negative stance towards the car's speed, for it takes a long time to cover long distances. On the other hand, in example (28b), the writer connects the first sentence with something positive and the framework of the interpretation changes. Let X be an entity described in the first sentence which is common in examples (28a) and (28b), i.e. "the speed of 100 klm/hour". If P is a predicate involving X and d is a degree of that predicate, then we get (4.7).

$$cause(E, P(X, d)) \tag{4.7}$$

In (28a), $P(X, d)$ stands for "the speed of 100klm/hour" (X), "are covered" (P) in "long time" (d). Analogously, for (28b), d is "short time".

As commented in the RST definitions, the Evaluation relation reflects the writer's positive regard for something, whereas the Interpretation relation relates the nucleus (27b) via the satellite (27b) to any other frame of reference, i.e. to any other view.

4.2 Anticausal Relations

The RST relations that we classified as anticausal are the following ones: *Concession, Unconditional, Otherwise* and *Unless*.

4.2.1 Concession

As defined in RST, in *Concession* the writer favors the nucleus while acknowledging the truth of the satellite which is opposed to the nucleus. Thus, the writer concedes to something via the satellite part, and shows the incompatibility between the satellite and the nucleus in order to increase the reader's positive regard for the latter.

(29) Tempting as it may be, we shouldn't embrace every popular issue that comes along. (Mann, 2005)

In (29), the writer concedes that embracing every popular issue that comes along is tempting but then presents her view that that should be avoided. In RST, *Concession* and *Antithesis* are very much related but differ in the level of incompatibility between the nucleus and the satellite. In *Antithesis*, the incompatibility is evident, whereas in *Concession* the incompatibility is *potential or apparent*. Introducing the Contrast relation in the RST framework complicates even more the situation as it will be seen later, as the lines defining when an annotator should use one of the three relations get really blurred.

These three relations are usually grouped as the contrastive relations of RST and are treated altogether. However, as we will show there are major differences between the concession and the contrast relation.

König (2006) distinguishes three kinds of concessive relations: *standard*, *rhetorical* and *rectifying* concession. The first one is exemplified in (30), where there is an apparent incompatibility between the two clauses.

(30) Although [it is raining]$_p$, [Fred is going out for a walk]$_q$.
(König, 2006)

The rhetorical concession is more argumentative, where the speaker compares two arguments over deciding a matter, e.g. going to a restaurant as in (31).

(31) Although the food is nice, it is expensive.

The rectifying concession involves a cancellation of an inference that might be made in the first clause as in (32).

(32) John is intelligent, but sometimes he acts in a silly way.

König's distinctions will be our guide to our investigation of *Concession* in what follows.

Standard Concession

This type of concession is described as being dual to causality but it has been shown by Iten (1998) that this is not true. Here, we will provide another view on why Iten is right. In example (30), the speaker accepts that raining is normally a cause for not going out for walk but in this particular case the rain is not a sufficient reason for not going out for a walk. This reasoning is formalized as follows:

$$(\text{although } p, q) \equiv (\neg cause(p, \neg q))$$

However, this does not convey information about the utterance, for we already have q and we can easily get (4.8).

$$q \models \neg cause(p, \neg q). \tag{4.8}$$

What really happens here is that one finds a generalization by making a connection between the current instance of p and q, where $cause(p, \neg q)$ and the general causal knowledge of the speaker as it is shown in (4.9).

$$\forall t, x\{time(t) \wedge person(x) \wedge rain(t) \rightarrow cause(rain(t), \neg goout(x, t))\} \tag{4.9}$$

So, the generalization is that whenever it rains, one is more likely not to go out.

This causality relationship can be seen as presupposed or belonging to the common ground or being a belief of the speaker that might be

accommodated by the hearer in case it is not shared. This may be better illustrated in example (33): the hearer might be unaware that the speaker lives in a country of the southern hemisphere, where it normally snows in June; hence, the hearer should accommodate that.

(33) a. Although it is June, it's not snowing.
b. It is June. Still, it's not snowing.
c. Sure it is June, but it's not snowing.

Other concessive constructions are the ones with adverbs such as *still, sure, well* which are most likely to be met within a dialogue. Additionally, the connectives *but, however* and *nevertheless* may well be used for the same type of concession.

(34) a. It is June but it's not snowing.
b. It is June. However, it's not snowing.
c. It is June. Nevertheless, it's not snowing.

Rhetorical Concession

The second type of concession is the rhetorical one, where there is no causal connection between the two clauses as in example (35). This type corresponds to that of Anscombre and Ducrot (1989). The main reason that this is regarded as concession is that the two clauses contrast to one another regarding the QUD.

For instance, there is a vacancy at a French basketball team where one needs to be tall and know French. Although example (35) might seem totally irrational in isolation, in this context, it makes perfect sense. The speaker acknowledges that *being tall* is an argument in favor of hiring Tom, whereas *not speaking French* is an argument against hiring him and she asserts that the latter argument is stronger than the former.

(35) Although [Tom is tall]$_p$, [he doesn't speak French]$_q$.

What is crucial here and differentiates argumentation from standard concession is that, in the latter, the QUD might be introduced by the sentence itself, whereas in the case of argumentation, the QUD must already be in the context. This happens because, although the two clauses are contrasted, there is no an immediate connection between them and the hearer cannot make something out of the utterance without some sort of a predefined QUD.

In spite of the fact that there is not an immediate connection between the two clauses as in *Standard Concession*, both clauses are causally connected to the QUD. In the example above, *being tall* can be seen as a cause for hiring someone on the French basketball team, whereas *not speaking*

French can again be seen as a cause for not hiring him. If *hiring Tom* corresponds to r, then (4.10) holds.

$$(\text{although } p, q) \equiv (cause(p, r) < cause(q, \neg r)) \tag{4.10}$$

We follow Winterstein (2015) by using a stochastic approximation to causality: a causes b, i.e. $cause(a, b)$, if and only if the probability of b raises if a is given, i.e. $P(b|a) > P(a)$. The *argumentative goal* (H in wintersteinian notation) in our example is r. Hence, example (35) can be modeled as follows:

$$\underbrace{P(r|p) > P(r)}_{cause(p,r)} \wedge \underbrace{P(\neg r|q) > P(\neg r)}_{cause(q,\neg r)} \wedge \underbrace{P(\neg r|q) > P(r|p)}_{\text{comparison}} \tag{4.11}$$

In this case, the speaker clearly makes an argument, apart from conceding to the truth of p, whereas in standard concession there is no arguing involved as q immediately decides the QUD. Utterance (35) can be imagined as the elliptic form of (36):

(36) Although [Tom is tall]$_p$, [he doesn't speak French,]$_{q \wedge r}$ [so we shouldn't hire him.]$_{ellipse=r}$

This example illustrates that the speaker has clearly taken sides with respect to what is the strongest argument. That inclination is captured when the ellipsis is undone, i.e. p argues against $q \wedge r$ and q argues for r:

$$cause(p, \neg(q \wedge r)) \wedge cause(q, r) \tag{4.12}$$

A more general example could be given if the QUD involves the question *who should be hired for the basketball team*? In both of the following examples, the second argument wins over the first one.

(37) a. Tom is tall, but Bill speaks French.
b. Although Tom is tall, Bill speaks French.

Their difference is that if (37a)'s clauses are inverted, it will be *Tom* that *should be hired*, whereas, if (37b)'s clauses are inverted, it will still be Bill that *should be hired* as it can be seen in the following.

(38) a. Tom is tall, but Bill speaks French. [So, we should hire Bill]
b. Bill speaks French, but Tom is tall. [So, we should hire Tom]
c. Although Tom is tall, Bill speaks French. [So, we should hire Bill]
d. Bill speaks French, although Tom is tall. [So, we should hire Bill]

The examples above illustrate that the order of clauses when using *but* is crucial in deciding the question, whereas when using *although* order is not important.

Rectifying concession

In this type of concession, the first clause is used to prevent not a causal but a conceptual inference made by the speaker on the basis of the second clause.

(32) John is intelligent, but sometimes he acts in a silly way.

In example (32), the hearer can well infer that *given John is intelligent, he does not act in a silly way* on the basis of the first clause. The second clause introduces an exception to the relevant conversational implicature.

Hence, this type of concession can be merged with *Standard Concession* and, analogously, (4.13) is the case.

$$\forall x\{(person(x)) \rightarrow cause(intelligent(x), \neg act\_silly(x))\} \tag{4.13}$$

The speaker apart from conveying the literal meaning of the two clauses shows that there is an exception to the relevant default reasoning, similarly to *Standard Concession*.

4.2.2 Unconditional

The Unconditional relation is related to the Concession relation, because the writer uses it when she acknowledges the existence of a causal relationship between two events or situations but, at the same time, makes room for an exception, which contradicts that causal regularity.

The Unconditional relation is grammaticized with the use of *even if* to connect two clauses. In RST, it is defined as the main clause, i.e. the nucleus, not being dependent on the complement clause, i.e. the satellite. However, this is not actually the case as we will see below. Consider example (39), where it might be the case that there is a conversation about Fred wanting to go out for a walk in unfavorable conditions such as raining.

(39) Even if [it snowed],$_{p_{max}}$ Fred would go out for a walk.

The function of *even if* can be mainly analyzed as the combination of *even* and *if*. That is, *even* creates a range of possibilities for what succeeds it: here, the cases under which Fred would go out for a walk. At the same time, it determines a maximally unlikely case which still causes Fred to go out for a walk.

Hence, the Unconditional relation can be modeled as $cause(p_{max}, \neg q)$ and $P(q|p_{max}) \ll P(q|p)$ for all p, where p are situations in which one would not do outdoor activities such as going out for a walk.

Via this construction the speaker also shows her surprise concerning two facts: first, that people normally do not go out for a walk when it snows,

whereas Fred does, and, second, that this is the maximally unlikely case among other ones, e.g. raining.

The peculiarity of the maximally unlikely case can be better outlined in example (40), where the speaker excludes an enormous set of cases but on the other hand leaves a margin of *doing such a thing* if *more than 1 million euros* is given to him.

(40) I wouldn't do such a thing, even if you gave me 1 million euros!

4.2.3 Otherwise

The Otherwise relation is used whenever the discourse marker *otherwise* is used. Both the nucleus and the satellite are unrealized and there is a dependence between them; namely the realization of the satellite is prevented by that of the nucleus. In this case, the writer wants the reader to recognize that dependency, as in example (41).

(41) a. Project leaders should submit their entries for the revised brochure immediately.
b. Otherwise the existing entry will be used. (Mann, 2005)

As it is commented in the definitions of RST the Otherwise relation may also describe cases of the following type: if A then B, otherwise (i.e. if not A) C. A formalization of that is (4.14).

$$cause(A, B) \wedge cause(\neg A, C) \tag{4.14}$$

So, A helps in bringing B about, whereas, $\neg A$ helps in bringing C about. A and B are events referred to in the nucleus and C is an event referred to in the satellite.

4.2.4 Unless

The Unless relation maps to the use of the discourse marker *unless*, which is usually explained in English as *if not*. In RST, it is explained as the satellite not being realized is a condition for the nucleus' realization.

(42) a. The following terms apply to all files associated with the software
b. unless explicitly disclaimed in individual files. (Mann, 2005)

Let A be the nucleus. It is generally the case that there is C which causes the nucleus (42a) to happen. At the same time, if P is the event described in (42b), then together with C cannot qualify as a cause of A.

$$(\exists C : cause(C, A)) \wedge (\neg cause((C \wedge P), A)) \tag{4.15}$$

Hence, the negation of the preventor is not a cause for the nucleus, as it is usually assumed, for the negation of preventors cannot be considered as causes. In other words, it certainly does not hold that $cause(\neg P, A)$.

4.3 QUD Relations

The RST relations relevant to a question under discussion are the following: *Antithesis, Contrast, Elaboration, Background, Circumstance, Disjunction, List, Sequence* and *Solutionhood*.

4.3.1 Antithesis

In this relation, the writer has a positive regard for the nucleus and wishes to increase the reader's positive regard for it. This is achieved by introducing a contrastive satellite which is incompatible to the nucleus and to which the reader disagrees.

Antithesis is one of the most weakly defined relations in RST as it resembles both a contrast and a concession relation. This weakness can be witnessed by taking into account the fact that *Antithesis* is rarely used to label passages in other researches where RST relations are used; e.g. according to Taboada (2006), the Antithesis Relation's frequency is 0.17% among 23 relations. In the RST definitions page, it is commented that the writer prefers one idea (nucleus) and not the other (satellite). The example given is (43) where the first sentence is the satellite and the second one is the nucleus.

(43) a. But I don't think [endorsing a specific nuclear freeze proposal is appropriate for CCC.]$_a$

b. We should [limit our involvement in defense and weaponry to matters of process, such as exposing the weapons industry's influence on the political process.]$_b$ (Mann, 2005)

So, in this relation, the question under discussion is raised and the writer initially presents the idea to which she is opposed. After she has given what she deems as the wrong answer to the question, she proceeds in giving the right answer to it.

Let q be a question and a and b two of its elements along with other ones $c, d, ...$, i.e. $q = \{a, b, c, ...\}$. The speaker makes that question the question under discussion, i.e. $QUD(q)$, and provides the following answer:

$$ans(QUD(q)) = \{\neg a, b\} \tag{4.16}$$

The incompatibility of the two propositions in q may be expressed as $(a \vee b) \wedge \neg(a \wedge b)$.

4.3.2 Contrast

Contrast bears some similarity to *Concession* in terms of the markers used: for instance, they both share the marker *but*. One of the most common type of examples (cf. Jasinskaja and Zeevat (2008), Winterstein (2012), Jasinskaja (2012)) is the utterance (12) repeated below. Example (12′) though, represents the rhetorical concession case.

(12) Jim plays the piano, but Mary plays the guitar.

(12′) Although Jim plays the piano, Mary plays the guitar.

It is interesting to note that (12′) has only one reading: the concession one. This may be due to the *although* marker being a subordinating one, which creates an *asymmetry* between the two clauses. On the other hand, *but* being a coordination marker can be used to convey contrast too.

So, although example (12) may well fall under the rhetorical concession case in the same manner (12′) does, it can have another reading as explicated in Jasinskaja and Zeevat (2008): it may respond to the QUD *who plays what*. Hence, it may be replaced by *and* or the order of the clauses may be reversed. This is called the *formal contrast* approach, where the contents of the conjuncts decide the interpretation of *but*.

Following Roberts (2012) we reason as follows: if there are only Jim and Mary and only two instruments, then the model of example (12) is:

1. Who plays what?
   (a) What does Jim play?
      i. Jim plays the piano.
      ii. Jim plays the guitar.
   (b) What does Mary play?
      i. Mary plays the piano.
      ii. Mary plays the guitar.

So, the question under discussion entails two questions:

$$\models q_1 = \{a_i, a_{ii}\} \tag{4.17}$$

$$\models q_2 = \{b_i, b_{ii}\} \tag{4.18}$$

The speaker answers accordingly that ans(QUD(q))=$\{a_i, b_{ii}\}$. Another typical example of contrast is a slightly altered example of (12), example (44).

(44) Jim plays the piano but Mary doesn't.

According to Jasinskaja and Zeevat (2008), this case is easy to model if we consider that the question under discussion is in the form *who "whether" plays the piano* and can be modeled in a similar way to the previous one:

1. Who "whether" plays the piano?

   (a) Does Jim play piano?

      i. Jim plays the piano.

      ii. Jim doesn't play the piano.

   (b) Does Mary play the piano?

      i. Mary plays the piano.

      ii. Mary doesn't play the piano.

Then, the same relations as before hold.

Analogously to the fact that (12′) does not convey contrast in the same way (12) does, (45) cannot be a rectifying concession in the same way (32) is. That's why *although* cannot be used instead of *but* here.

(45) Jack is not going to Paris, but to Berlin.

To sum up, contrast may be seen as a discourse relation that bears more similarity to typical conjunction with argumentative features. When these features become strong enough, then one can talk about *Concession* being brought forward.

4.3.3 Elaboration

The writer's intention, when using the Elaboration relation as defined in RST, is that the reader recognize that additional detail is provided for a topic. The nucleus identifies the subject matter, whereas the satellite carries the further details. There can be a series of pairs in which the nucleus presents the first part and the satellite the second: set and member, object and attribute, whole and part, generalization and specific are some of those pairs. Consider example (46).

(46) a. City, in Sweden, will be the site of the 1969 International Conference on Computational Linguistics, September 1-4.

b. It is expected that some 250 linguists will attend from Asia, West Europe, East Europe including Russia, and the United States.

c. The conference will be concerned with the application of mathematical and computer techniques to the study of natural languages, the development of computer programs as tools for linguistic research, and the application of linguistics to the development of man-machine communication systems. (Mann, 2005)

The nucleus (46a) presents the object, i.e. the linguistics conference, and the satellites (46b) and (46c) present its attributes. This can be generally represented as

$$(A \subset B) > (A \not\subset B) \tag{4.19}$$

where A is what is described in the nucleus and B what is described in the satellite(s). That is, it is more probable that the attributes listed are these of the conference than they are not.

In a way, the nucleus creates questions which are subsequently answered. Thus, it can be argued that the writer initially updates the question under discussion by adding questions and then provides the respective answers.

As Taboada (2006) notes, *Elaboration* is frequently used by journalists when writing an article: the article as a whole can be seen as its headline's elaboration. That implies that the Elaboration relation can also be modeled as an inverted Summary relation by using the identity inference. However, this time A, i.e. the nucleus, is strictly smaller in length than B, i.e. the satellite.

4.3.4 Background

The Background relation is used whenever the nucleus cannot be sufficiently understood without the satellite and the writer wants to increase the ability of the reader to understand it. The satellite, thus, makes the reader more able to comprehend an issue in the nucleus.

This relation is about clarifications that the writer gives to the reader by a preceding content elaboration rather than following one, as in example (47), where (47b) is the nucleus and (47a) is the satellite.

(47) a. ZPG's 1985 Urban Stress Test, created after months of persistent and exhaustive research, is the nation's first survey of how population-linked pressures affect U.S. cities. It ranks 184 urban areas on 11 different criteria ranging from crowding and birth rates to air quality and toxic wastes.

b. The Urban Stress Test translates complex, technical data into an easy-to-use action tool for concerned citizens, elected officials and opinion leaders. (Mann, 2005)

According to RST, the background relation pertains to general information that is conveyed to the reader. This means that it is not vital to the understanding of the sentence and is surely not necessary to it, but it nonetheless contributes to it.

In our analysis, the Background relation is an inverted elaboration relation which may be modeled on the basis of updates to the initial information state. These updates are relevant to a question under discussion and provide further details on the answer being given to it.

4.3.5 Circumstance

This relation is described as setting the framework for interpreting the nucleus. Its most frequent signals clarify what is meant by the word framework in RST: *when, as, after* and *following* show that most of the uses are to determine the time an event took place. Furthermore, this relation is also used to determine *where* an event happened. Time and place play the role of the circumstances or the framework in the explanation of RST. Consider example (48).

(48) a. Probably the most extreme case of Visitors Fever I have ever witnessed was a few summers ago

b. when I visited relatives in the Midwest.

(Mann and Thompson, 1988)

According to the RST's definitions, the difference between the Background and the Circumstance relation is that in Circumstance the nucleus and the satellite refer to a single situation. So, again, the information state is updated with new information but the question under discussion obliges the writer to provide specific information that add to the topic in the form of an elaboration on it.

Another major difference is that *Circumstance* usually hides a causal relation, for without the satellite, the nucleus cannot happen, as it is the case in example (48). However, this contiguity requirement is not always there and, thus, *Circumstance* might fall under *Background* or even *Elaboration*: in example (49), there is no causal relation between the satellite (49a) and the nucleus (49b).

(49) a. Insisting that they are protected by the Voting Rights Act,

b. a group of whites brought a federal suit in 1987 to demand that the city abandon at-large voting for the nine-member City Council and create nine electoral districts, including four safe white districts. (Taboada, 2006)

Circumstance is comparable to a locative clause in a sentence that sets time or place for the main clause. In that respect, it sets a new question: what happened there or then. Namely, if a is (49a) and b is (49b) then the updates to an information state σ will be as follows:

$$\sigma[a] \models QUD(m) \tag{4.20}$$

$$\sigma[QUD(m)] \models ans(QUD(m)) \tag{4.21}$$

4.3.6 Disjunction

In *Disjunction*, the nuclei present two alternatives which might not necessarily be exclusive. Thus, it is about providing non-exhaustive answers to a

question under discussion that has been presented earlier in the discourse. The relation is grammaticized in English with the word *or*. Considering it is a multinuclear relation, both clauses are of equal importance. In other words, they are equally acceptable answers to the question under discussion.

The relation does not really decide the matter as it gives two possibilities on a single question under discussion. The two sentences a, b can be seen as partial answers and mutually exclusive ones (4.22).

$$\models \{\{a = ans(QUD(m))\} \vee \{b = ans(QUD(m))\}\} \wedge \{\neg(a \wedge b)\} \quad (4.22)$$

4.3.7 List and Sequence

When using *List*, the writer presents a series of events that might share some properties with each other, but are not causally related to one another and their order in which they appear does not play a role. *List* does not entail any temporal or spatial relationship among the recounted events.

On the other hand, in *Sequence*, order is important; this relation in other theories is usually labeled as *Narration* (Lascarides and Asher, 1993) or *Occasion* (Kehler, 2002). *Sequence* can be thought as representing the narration of a story, where the events are described one after the other and it is not necessary that much information apart from the sole happening of the events is provided. Another defining characteristic noted by Hobbs (1976) is that each event is contingent on the one preceding it. Consider the following example (50).

(50) Luke opened the fridge and grabbed the bottle of milk.

Considering the above, *List* can be seen as a content elaboration relation on a given question under discussion. On the other hand, *Sequence* is more complex than *List*, for the clauses have causal links among them. Each event seems to be a precondition for the next one to happen. Concerning example (50), if Luke hadn't opened the fridge, he wouldn't have grabbed the bottle of milk. That would be the counterfactual, whereas the contingency relationship is that opening the fridge allows Luke to grab the bottle of milk.

A food recipe may also be thought as a sequence relation between events described in each clause. It is important to note, however, that it is not necessary that the syntactic order of the sentences reflects the natural order of the events that they describe.

(51) Wash the chicken well and put it in the oven for 1 hour.

(51′) Put the chicken in the oven, after you have washed it well.

So, *Sequence* requires that the events happen in different times and that each one is finished before the other takes place. *Sequence* can be modeled as providing partial answers to a question under discussion set beforehand. If $q = \{e_n | n \in \mathbb{N}\}$ is a question of the form *What happened?* which was set as the question under discussion, the Sequence relation provides an answer to it, so

$$\models ans(QUD(q)) \tag{4.23}$$

The partiality of the answer can be expressed as

$$ans(QUD(q)) = \{e_f, e_2, e_3, e_4, e_l\} \tag{4.24}$$

where e_f can be the first event and e_l the last one. The temporal difference of the events in *Sequence* can be expressed as $t(e_1) \neq t(e_2)$.

However, given the contingency requirement for the events described in *Sequence*, one cannot disregard the fact that the relation can be classified as a causal one too. Thus, *Sequence* certainly qualifies for both categories of relations, i.e. causal ones and QUD ones.

4.3.8 Solutionhood

Solutionhood is about presenting a problem and providing a solution to it. Consider example (52). The problem is outlined in the nucleus (52a) and the solution is suggested in the satellite (52b).

(52) a. One difficulty ... is with sleeping bags in which down and feather fillers are used as insulation. This insulation has a tendency to slip toward the bottom.

b. You can redistribute the filler. (Mann and Thompson, 1988)

Here, the nucleus updates the question under discussion with a new question and the satellite subsequently provides an answer to it. As Taboada (2006) notes, the nucleus can be a proper question but also an assertion.

If m is the nucleus, then the information state is updated with $QUD(m)$, as the writer makes that move the question under discussion and then the new information state is updated with $n = ans(QUD(m))$, where n is the satellite, in our example (52b).

4.4 Identity Relations

The identity relations are three as we have grouped them: *(Multinuclear) Restatement* and *Summary*.

4.4.1 (Multinuclear) Restatement

We consider *Restatement* and *Multinuclear Restatement* as the same relation, for there is no difference neither in the inferences involved nor in the way the information state is updated. This relation is to be taken literally as it is one of the few relations in RST that the intention of the writer is circularly described, i.e. the reader recognizes that the satellite is a restatement of the nucleus. The nucleus is to be considered more central to the writer's goals than the satellite is. Example (53) the relation.

(53) a. [Title:] A WELL GROOMED CAR REFLECTS ITS OWNER

b. The car you drive says a lot about you. (Mann, 2005)

What matters in every restatement is to avoid hypothesizing a new entity in the second sentence. In example (53), both sentences refer to the same car that characterizes its owner. *Restatement* is, thus, an identity relation. So, if A is (53a) and B is (53b), (4.25) holds.

$$(A = B) > (A \neq B) \tag{4.25}$$

That is, it is more probable that the B is a restatement of A than B not being a restatement of A.

Interestingly, the restatement relation may be marked with *so*, as in example (54). This marker can be thought as an indicator of an inference to be made. However, it is not necessary to be present, for instance in the previous example (53).

(54) [N] Okay, next week, again, Thursday, or maybe Friday, [S] so the tenth or the eleventh. (Taboada, 2006)

In the example above, we infer that the tenth and the eleventh date of a month mentioned in the satellite sentence (marked with [S]) refer precisely to the two days (Thursday and Friday) that introduced before.

The restatement relation is the exact same relation labeled as *Elaboration* by Kehler (2002), in which two assertions constrain the corresponding relations and entities to be the same. What changes is the perspective or the depth of the details given. In (55), the understanding of the passage is established as long as the reader infers that there is only one event narrated here.

(55) A young aspiring politician was arrested in Texas today. John Smith, 34, was nabbed in a Houston law firm while attempting to embezzle funds for his campaign. (Kehler, 2002)

4.4.2 Summary

Similarly to the Restatement relation, in *Summary*, the writer's intention is that the satellite be recognized as a shorter restatement of the nucleus. Consider example (56). Although this example is marked, the Summary relation, in general, is rarely marked according to Taboada and Mann (2006b).

(56) [N] Many agencies roll over their debt, paying off delinquent loans by issuing new loans, or converting defaulted loan guarantees into direct loans. [S] In any case, they avoid having to write off the loans. (Taboada, 2006)

The equation that holds between the two parts is the same as that of *Restatement* (4.25) but the difference is that the length of A is always bigger than the length of B.

4.5 Other Relations

The relations that do not fall under the categories presented earlier are the following three: *Evidence, Justify* and *Motivation.*

4.5.1 Evidence

The Evidence relation is about convincing the reader about a proposition, by providing information, so that her belief towards this proposition is strengthened. The reader might not yet believe the nucleus but by presenting the satellite which is either more credible or part of the reader's belief, she may be persuaded. Mann and Thompson (1988) provides an illustration of the relation with example (57).

(57) a. The program as published for calendar year 1980 really works.

b. In only a few minutes, I entered all the figures from my 1980 tax return and got a result which agreed with my hand calculations to the penny. (Mann and Thompson, 1988)

If X is the nucleus (57b), A (57a) is a successful argument. Hence, given an information state σ, after updating with A, we get (4.26).

$$\sigma[A] \models must(X) \Leftrightarrow \sigma[A] \models X > \neg X \tag{4.26}$$

So, in our framework, the Evidence relation is an argumentative one and not a causal relation as Sanders et al. (1992) want it to be.

Because of the fact that the Evidence relation and causal relations such as *(Non-)Volitional cause* share the same marker, i.e. *because*, the two relations are labeled as instances of the Explanation relation, e.g. (Hobbs, 1976). Modern Greek's discourse marker *epeidi* is a typological argument on why that is not a good practice. As Kitis (2006) has shown, *epeidi* is a "direct cause" marker, i.e. it is used only to show a causal relation between two events contrary to *giati* which is loosely used in the same way *because* is in English.

(58) John came back because he loved her.
   a. O Giannis epestrepse giati/epeidi tin agapuse.

(59) John loved her. Because he came back.
   a. O Giannis tin agapouse, giati epestrepse
   b. * O Giannis tin agapouse, epeidi epestrepse.

(Bardzokas, 2014)

Example (58) illustrates a volitional cause relation and the two Greek discourse markers can interchangeably be used. John's volition results from his love and this qualifies as a cause for his return. In example (59) though, it cannot be accepted that coming back qualify as a cause of John loving her. The speaker most probably just desires to substantiate her claim by presenting a fact that the hearer might ignore.

4.5.2 Justify

This relation together with the Evidence relation form a subgroup according to Mann and Thompson (1988), for they both refer to the reader's predisposition toward the nucleus. When using the Justify relation, the writer's intention is to increase the reader's readiness to accept the writer's right to present the nucleus. The reader comprehending the satellite is the means to do that. The definition of this relation is very problematic as well as its example below (60).

(60) a. The next music day is scheduled for July 21 (Saturday), noon-midnight.
   b. I'll post more details later, but this is a good time to reserve the place on your calendar. (Mann and Thompson, 1988)

Here, the satellite (60b) supposedly gives the right to the reader to present the nucleus (60a). If X is the nucleus and A the satellite which stands for an argument, we will get

$$\sigma[A] \models must(K_s(X)) \tag{4.27}$$

$K_s(X)$ can be thought in the simplest terms as "the speaker s knows that X". Two things must one bear in mind with respect to this formulation: I use knowledge as Veltman (1996) does, i.e. the speaker knowing something does not imply that it holds; it is often the case that what an agent regards as knowledge is false.

Second, $K_s(X)$ might seem abundant when the speaker has already uttered X, but this is not quite right. As van der Sandt (2010) notes, when one utters ϕ, one is expected to *act as if* it is true, but he neither explicitly nor implicitly states that one knows (or even believes) that ϕ.

Considering the above, $K_s(X)$ does not carry as much information (or implications) and just reflects what the agent *deems* as knowledge regardless of it being true or not. The Justify relation is thus weaker than the Evidence relation, where both the speaker and the hearer accept X as true.

4.5.3 Motivation

The Motivation relation is used when the writer intends to reinforce the reader's desire to perform an action. The satellite contributes to that, whereas the nucleus is an action to be fulfilled by the reader as it is exemplified in example (61).

(61) a. Ask for SYNCOM diskettes, with burnished Ectype coating and dust-absorbing jacket liners.

b. As your floppy drive writes or reads, a Syncom diskette is working four ways [...] (Mann, 2005)

In RST, the Motivation relation is always grouped with the Enablement relation, because they are used in similar contexts and their respective definitions have only one difference to one another: the effect of the Enablement relation is to increase the potential ability of the reader to perform the action in the nucleus, whereas the effect of the Motivation relation is to increase the desire of the reader to perform the action in the nucleus.

Hence, the Enablement relation is about providing such information that make the action in the nucleus more likely, whereas the Motivation relation is about providing more information considered appealing to the reader. As Skoufaki (2009) underlines, the Motivation relation is not listed in the coherence relations of the RST Annotation Tool and in her corpus study this relation is tagged as 'preference'. The relation can be modeled as follows:

$$\sigma[A] \models must(want_R(X)) \tag{4.28}$$

The predicate $want_R(X)$ means "the reader wants X" and can be thought as any other proposition that updates an information state, as

it does not have other implications on the actual world. That is, the fact that the reader wants X does not render X more probable. At best, the desire of something makes it more tangible but without any further action the desire remains a desire. For instance, if one *simply* wants to be an astronaut, the desire on its own does not increase any probability of becoming one.

4.6 Too vague Relations

There are three RST relations that fail to qualify as discourse relations: *Preparation*, *Joint* and *Conjunction*. Preparation makes the reader more interested or ready to read what follows. This relation is a later addition to the RST and covers parts of text that is impossible to be related otherwise, such as titles that are not summing up what follows but just prepare the reader about it. However, what it means to be prepared to read or hear something is questionable, and, as a consequence, the relation at hand is discarded as too vague.

Joint corresponds to the absence of any relation, according to RST, which means that it cannot be taken seriously as a proper discourse relation but rather as the discourse analyst's inability to make a decision on what relation to choose. Ideally, *Joint* can be pictured as a *garbage bin* wherein relations are thrown; then, thorough investigation may lead to new relations to be incorporated in RST.

Lastly, *Conjunction* is defined as conjoining items to form a unit in which each one plays a comparable role. The writer's intention is that the reader recognize the conjunction. This relation does not carry much information and likely falls under *List*. Hence, there is no need to assume an additional relation.

Chapter 5

Conclusion

In Chapter 4, we analyzed the discourse relations going through the list provided by Rhetorical Structure Theory (Mann, 2005). The analysis of the relations was carried out with the tools presented in Chapter 3: *Update Semantics with Causal Probabilities* along with *Argumentation* and *Questions Under Discussion*.

The analysis at hand can be seen as part of a larger project, that of Bayesian Natural Language Semantics and Pragmatics (BNLSP) as introduced by Zeevat (2015). That is, the hypothesis that cognition's nature is Bayesian is tested on human communication, for the later is crucially characterized by the underdetermination of meaning by form. The Gricean intention recognition and the Hobbsian interpretation as abduction are combined in Bayesian interpretation which is applied to actual corpora apart from its traditional use in Artificial Intelligence. Our stochastic model also incorporated causal probabilities to capture the particular inferences interlocutors made which cannot be captured by BNLSP alone.

By definition, RST relations refer to written text, i.e. monologue, and not to dialogue. Nevertheless, the relations analyzed here can be applied in spoken text as Taboada (2004) has showed. For instance, the discourse relation between a question and its answer can be thought as an Elaboration relation and be modeled with respect to QUD.

However, there are some relations that seem exclusively pertinent to dialogue, such as *Acknowledgment* which connects the utterance of a speaker and a positive responding by another speaker like *yeah* or *uh huh*. Yet even *Acknowledgment* can be adapted to monologue, as one can, for example, acknowledge suggestions which are attributed to others. The approach of Asher and Lascarides (2003) is wider with respect to explicating the role of discourse relations in dialogue proper and our enterprise could be thought as a starting point of a larger theory based on Bayesian Interpretation and competing with SDRT.

Although our approach can be applied to dialogue, it is not sufficient,

for we haven't accounted for prosody and its role in distinguishing new and given information, in setting the focus and the topic, and in conveying speech acts that a written text could not possible convey. For instance, pauses in speech can decide whether an utterance is subordinated or coordinated to another one (Tyler et al., 2011), so interlocutors may draw different inferences respectively.

As far as discourse markers are concerned, it was seen that discourse relations except anticausal ones are asyndetic. That fact makes room for an analysis not having as its starting point the discourse markers, for the later are very likely to be ambiguous; e.g. *but* is used both contrastively and concessively. Although discourse markers can be correlated with discourse relations, it is not wise to define discourse relations based on them. Even in anticausal relations, where a discourse marker is needed, there are important differences in its use; e.g. interlocutors' inferences differ regarding whether a concession is a standard one or a rhetorical one.

Considering that our enterprise was inspired by an idea of Zeevat (2011), i.e. to carry out a feature analysis of discourse relations, we will address some of the questions on dialogue act taxonomies posed by Traum (2000). As Traum explicitly mentions, the term *dialogue acts* is used as an umbrella term for the different names researchers have given to speech acts in the past, such as *communicative acts* and *conversation acts*.

Defining discourse relations In our approach, the speaker intention plays an important role in analyzing discourse relations insofar its recognition is incorporated in BNLSP. As Traum notes, that feature requires mind-reading as far as the hearer is concerned and that is why we tried to refrain from adding many propositions to the commitment slate of the speaker, following van der Sandt (2010).

The point of view in analyzing discourse relations is also an issue for Traum as long as we applied to them a logical semantics. Given that we focused on monologue, the hearer's view was the prominent one, for the addressee uptake was discussed as far as the text's interpretation is concerned.

Discourse relations components Traum's *Question 7*, i.e. how actions are used in a logic, was presented in Chapter 3, where we showed how events and actions can be causally linked by incorporating causal probabilities in update semantics and that was the tool to define the components of causal and anticausal relations.

Context was also defined with the help of Roberts (2012)'s Questions Under Discussion and that was used to model the respective QUD relations, where topicality is important in representing the effect of a discourse relation.

Relationships and complex discourse relations By using a topicality tool such as QUD, we also answered in Traum's question on how dialogue acts interact with dialogue structure. For instance, by analyzing *Solutionhood*, we showed that the relation is about setting a new question under discussion and providing an answer to it.

The most central question of Traum as far as our enterprise is concerned is whether dialogue acts can be "composed" of more primitive acts. We showed that the decomposition of discourse relations does not need to contain more primitive acts, in the sense, for example, Sanders et al. (1992) want them to. That decomposition may be realized in other terms and our framework was constructed with that idea in mind.

Bayesian Interpretation of Discourse Relations

Our analysis ultimately tackles the following question: is it possible for a Bayesian interpreter who comes to know a new element A to automatically assign a discourse relation to it? Such an interpreter will have to check the following possibilities:

1. Whether A is caused by or is contingent on a recent given event X
2. Whether A causes a recent given event or makes it causally possible
3. Whether A is arguing for or against a statement under discussion
4. Whether A is identical to a given event or entity
5. Whether A is part of a given event or entity
6. Whether A is relevant to a question under discussion

If 1 or 2 is the case, the relation will be one of causal or anticausal ones. If 3 is the case, the relation will be an anticausal one and then by checking which marker is used, one could decide which discourse relation holds.

If 4 is the case, then there is an identity inference involved, so one should calculate the length of each sentence involved to decide if it is *Summary* or *Restatement*. For case 5, that would be *Elaboration* as well as *Sequence*. If 6 is the case, then A must be relevant to a question under discussion as in *Background*.

If none of the above is the case, then there might be some modality involved, e.g. *may*, and, consequently, one would have to decide which one of the three relations (*Evidence, Justify* or *Motivation*) is the case. Therefore, it could be possible for an interpreter who reasons in a Bayesian manner to decide which discourse relation holds between a new element in the discourse and a given one.

Bibliography

Jean-Claude Anscombre and Oswald Ducrot. *L'argumentation dans la langue*, volume 2. Editions Mardaga, 1983. ISBN 287009177X.

Jean-Claude Anscombre and Oswald Ducrot. Argumentativity and Informativity. In Michel Meyer, editor, *From Metaphysics to Rhetoric*, pages 71–87. Springer Netherlands, Dordrecht, 1989. ISBN 978-94-010-7672-2.

Nicholas Asher and Alex Lascarides. *Logics of Conversation*. 2003. ISBN 0521650585.

Valandis Bardzokas. Linguistic constraints on causal content: The case of Modern Greek markers. *Journal of Pragmatics*, 60:160–174, 2014.

Daniel Büring. On D-Trees, Beans, And B-Accents. *Linguistics and Philosophy*, 26(5):511–545, 2003.

Paul Grice. *Logic and conversation*. 1970.

Jeroen Groenendijk and Martin Stokhof. *Studies on the semantics of questions and the pragmatics of answers*. PhD thesis, University of Amsterdam, 1984.

Michael Alexander Kirkwood Halliday and Ruqaiya Hasan. *Cohesion in English*, volume 7. 1976. ISBN 0582550416.

Jerry R. Hobbs. On the Coherence and Structure of Discourse. *Structure*, pages 1–36, 1976.

Jerry R. Hobbs. Coherence and coreference. *Cognitive Science*, 3(1):67–90, January 1979.

Jerry R Hobbs. *Literature and Cognition*, volume 18. 1990. ISBN 0937073539.

Jerry R. Hobbs, Mark E. Stickel, Douglas E. Appelt, and Paul Martin. Interpretation as abduction. *Artificial Intelligence*, 63(1-2):69–142, 1993.

David Hume. *An Inquiry Concerning Human Understanding*. The Liberal Arts Press, New York, 1955 edition, 1748.

Corinne Iten. Because and although : a case of duality? *Pragmatics and Beyond New Series*, 9:59–80, 1998.

Katia Jasinskaja and Henk Zeevat. Explaining additive, adversative and contrast marking in Russian and English. *Revue de Semantique et Pragmatique*, 24:65–91, 2008.

Katja Jasinskaja. Correction by adversative and additive markers. *Lingua*, 122(15):1899–1918, December 2012.

Hans Kamp and Uwe Reyle. From Discourse to Logic, 1993.

Andrew Kehler. Coherence, reference, and the theory of grammar. *CSLI Publishers. Kibble, R., and R. Power*, pages 187–192, 2002.

Andrew Kehler, Laura Kertz, Hannah Rohde, and Jeffrey L Elman. Coherence and Coreference Revisited. *Journal of semantics*, 25(1):1–44, January 2008.

Eliza Kitis. Causality and subjectivity: The causal connectives of Modern Greek. In H. Pishwa, editor, *Language and Memory: Aspects of Knowledge Representation*, chapter 8, pages 223–267. Mouton de Gruyter, Berlin/New York, 2006. ISBN 3110895080.

Ekkehard König. Concessive Clauses. In Keith Brown, editor, *Encyclopedia of Language & Linguistics*, pages 820–824. Elsevier, 2006. ISBN 9780080448541.

Alex Lascarides and Nicholas Asher. Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy*, 16: 437–493, 1993.

David Lewis. Causation. 1973.

William C. Mann. RST web site, 2005.

William C. Mann and Sandra Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281, 1988.

J. R. Martin. *English Text: System and Structure*. 1992. ISBN 9027220794.

Arthur Merin. Information, relevance and social decision-making. *Logic, Language and Computation*, volume 2:179–221, 1999.

Craige Roberts. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6):1–69, 2012.

Ted J. M. Sanders, Wilbert P. M. Spooren, and Leo G. M. Noordman. Toward a taxonomy of coherence relations. *Discourse Processes*, 15(1): 1–35, January 1992.

John Searle. *Speech acts: An essay in the philosophy of language*, volume 0. 1969. ISBN 052109626X.

Sophia Skoufaki. An Exploratory Application of Rhetorical Structure Theory to Detect Coherence Errors in L2 English Writing : Possible Implications for Automated Writing Evaluation Software. *Computational Linguistics and Chinese Language Processing*, 14(2):181–204, 2009.

Robert Stalnaker. Assertion. In Peter Cole, editor, *Syntax and Semantics*, volume 9. New York, academic p edition, 1978.

Maite Taboada. Rhetorical Relations in Dialogue. In Carol Lynn Moder and Aida Martinovic-Zic, editors, *Discourse Across Languages and Cultures*, pages 75–97. John Benjamins, Amsterdam/Philadelphia, 2004.

Maite Taboada. Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics*, 38(4):567–592, April 2006.

Maite Taboada and William C. Mann. Applications of Rhetorical Structure Theory, 2006a.

Maite Taboada and William C Mann. Rhetorical Structure Theory: Looking Back and Moving Ahead. *Discourse Studies*, 8:423–459, 2006b.

David R. Traum. *A Computational Theory of Grounding in Natural Language Covnersation*. PhD thesis, University of Rochester, Rochester, New York, 1994.

David R. Traum. 20 Questions on Dialogue Act Taxonomies. *Journal of Semantics*, (17):7–30, 2000.

Joseph Tyler, Jason Kahn, and Jennifer Arnold. Speakers use prosody to communicate discourse structure, and listeners use that prosody in comprehending discourse structure. In *Experimental and Theoretical Advances in Prosody (ETAP) 2*, January 2011.

Rob van der Sandt. Pragmatic strategies. In R. Bauerle, U. Reyle, and T. E. Zimmermann, editors, *Presuppositions and Discourse: Essays offered to Hans Kamp*, chapter 1, pages 1–25. Emerald, 2010.

Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3), June 1996.

Christiane von Stutterheim and Wolfgang Klein. Referential movement in descriptive and narrative discourse. pages 39–76, 1989.

Bonnie Webber, Markus Egg, and Valia Kordoni. Discourse structure and language technology. *Natural Language Engineering*, 18(04):437–490, December 2012.

Gregoire Winterstein. Argumentative properties of pragmatic inferences. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5447 LNAI: 161–176, 2009.

Grégoire Winterstein. What but-sentences argue for: An argumentative analysis of but. *Lingua*, 122(15):1864–1885, December 2012.

Grégoire Winterstein. Layered Meanings and Bayesian Argumentation: The Case of Exclusives. In Henk Zeevat and Hans-Christian Schmitz, editors, *Bayesian Natural Language Semantics and Pragmatics*, volume 2 of *Language, Cognition, and Mind*, pages 179–200. Springer International Publishing, Cham, 2015. ISBN 978-3-319-17063-3.

Henk Zeevat. Rhetorical relations. In C. Maienborn, K. Von Heusinger, and P. Portner, editors, *Semantics*, chapter 39, pages 946–970. de Gruyter, 2011. ISBN 978-3-11-022661-4.

Henk Zeevat. Implicit Probabilities in Update Semantics. *M. Aloni, M. Franke & F. Roelofsen (eds.), Festschrift for Jeroen Groenendijk, Martin Stokhof, and Frank Veltman, Amsterdam*, 2013.

Henk Zeevat. Bayesian Foundations for Presupposition, 2014.

Henk Zeevat. Perspectives on Bayesian Natural Language Semantics and Pragmatics. In Henk Zeevat and Hans-Christian Schmitz, editors, *Bayesian Natural Language Semantics and Pragmatics*, pages 1–24. Springer International Publishing, 2015.