Ted J.M. Sanders*, Vera Demberg, Jet Hoek,
Merel C.J. Scholman, Fatemeh Torabi Asr, Sandrine Zufferey
and Jacqueline Evers-Vermeul

Unifying dimensions in coherence relations: How various annotation frameworks are related

Abstract: In this paper, we show how three often used and seemingly different discourse annotation frameworks – Penn Discourse Treebank (PDTB), Rhetorical Structure Theory (RST), and Segmented Discourse Representation Theory – can be related by using a set of unifying dimensions. These dimensions are taken from the Cognitive approach to Coherence Relations and combined with more fine-grained additional features from the frameworks themselves to yield a posited set of dimensions that can successfully map three frameworks. The resulting interface will allow researchers to find identical or at least closely related relations within sets of annotated corpora, even if they are annotated within different frameworks. Furthermore, we tested our unified dimension (UniDim) approach by comparing PDTB and RST annotations of identical newspaper texts and converting their original end label annotations of relations into the accompanying values per dimension. Subsequently, rates of overlap in the attributed values per dimension were analyzed. Results indicate that the proposed dimensions indeed create an interface that makes existing annotation systems "talk to each other."

*Corresponding author: Ted J.M. Sanders, Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, Netherlands, E-mail: T.J.M.Sanders@uu.nl
Vera Demberg, Computer Science, Saarland Informatics Campus, Saarbrücken, Germany, E-mail: vera@coli.uni-saarland.de
Jet Hoek, Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, E-mail: j.hoek@uu.nl
Merel C.J. Scholman, Language Science & Technology, Saarland University, Saarbrücken, Germany, E-mail: m.c.j.scholman@coli.uni-saarland.de
Fatemeh Torabi Asr, Discourse Processing La, Department of Linguistics, Simon Fraser University, Burnaby, Canada, E-mail: fatorabi@indiana.edu
Sandrine Zufferey, Institut de Langue et de Littérature françaises, University of Bern, Bern, Switzerland, E-mail: sandrine.zufferey@rom.unibe.ch
Jacqueline Evers-Vermeul, Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, Netherlands, E-mail: j.evers@uu.nl

1 Introduction

In recent decades, linguistics has seen major developments in the area of corpus linguistics. Large corpora allow us to obtain qualitative and quantitative observations about language use. For a long time, corpus annotation was limited to annotation at the morphological, syntactic or semantic level, but over the last 15 years the annotation of corpora at the discourse level has been realized in large annotation efforts. Leading examples of discourse annotation frameworks include the Penn Discourse Treebank (Prasad et al. 2008), the Rhetorical Structure Theory Discourse Treebank (RST-DT; Carlson and Marcu 2001), and the Segmented Discourse Representation Theory (SDRT; Asher and Lascarides 2003).

This development enables us to take the study of coherence relations an important step forward. Corpora can now be searched for coherence relations, whether they remain implicit, or are linguistically marked by cue phrases or connectives. For instance, looking at all annotated occurrences of a connective like English *since* allows us to determine how often and under which circumstances *since* expresses a TEMPORAL relation, as in (1) or a CLAIM-ARGUMENT relation, as in (2). In addition, we can search corpora for cases in which alternative connectives are used to express the same relation, as in (3), which also expresses a claim-argument relation, or cases in which the same relation is conveyed implicitly as in (4). Analyzing such cases in a qualitative and quantitative way provides us with important insights into a connective's distribution over coherence relations.

(1)　　*Since* Cruijff[1] played on the team, they never lost a game.

(2)　　It was impossible they would lose the game, *since* Cruijff played on the team.

(3)　　It was impossible they would lose the game, *because* Cruijff played on the team.

(4)　　It was impossible they would lose the game. Cruijff played on the team.

From annotations in the Penn Discourse Treebank (PDTB Research Group 2008), we know that more than half of all CCR are not explicitly marked by a connective

---

1 Johan Cruijff (1947) was the best Dutch soccer player ever, and one of the best in the world; he passed away on 24 March 2016.

or cue phrase. This observation raises several important issues for discourse annotation. One question is whether some relation types are more often conveyed implicitly or by the use of alternative signals than others, and if so, what the causes of these differences are (Asr and Demberg 2012; Das and Taboada 2018; Hoek et al. 2017b; Taboada 2006). To address these issues, the development of extensive and comparable sets of annotated data with coherence relations across several languages and genres will represent a major step ahead.

In other words, the existence of discourse-annotated corpora is also crucial to the field of discourse studies and language use. Therefore, it would be worthwhile to make the various annotated corpora accessible to and comparable for all researchers in the field.[2] At present this is not yet possible. While there is a large consensus regarding the usefulness of discourse-annotated data, there are many alternative ways of annotating coherence relations, and discourse annotation schemes differ strongly in the type of coherence relations that are distinguished (Hovy and Maier 1995), varying from sets of approximately 20 relations (such as the original RST developed by Mann and Thompson 1988), others of only two relations (Grosz and Sidner 1986). The PDTB contains a three-tiered hierarchical classification of 43 sense tags (Prasad et al. 2008), and the annotation scheme used for the RST Treebank distinguishes 78 relations that can be partitioned into 16 classes (Carlson and Marcu 2001).

The annotation schemes do not only differ in granularity, but also in their choice of labels: different labels are used for the same conceptual relations, and the same labels are used for different relation sense definitions. This makes it extremely difficult to make comparisons across corpora that are annotated according to different frameworks.

1.1 Goal of the paper

This article proposes a way to "translate" annotation tags from one framework to the terminology of other frameworks, so that the different annotation systems can "talk to each other." Our concrete goal is to develop an interface that will allow researchers to find identical or at least closely related relations within a set of annotated corpora, even if these relations carry different names in the respective frameworks in which they were annotated. To make this goal even more concrete: imagine a discourse researcher who uses the PDTB framework, and who is interested in REASON relations in English. She might want to know how often these relations are made explicit with connectives like *because* or *since*, and how often

---

and under which circumstances they remain implicit. She knows the labels provided in the PDTB, but in order to benefit from other annotated corpora, she also needs to know what labels to search for in RST-DT or SDRT in order to retrieve similar relations. Our interface will allow her to do exactly that: start from the tag REASON, and find similar or closely related relations in other frameworks (for example, relations labeled as RESULT in SDRT for this particular case), so that her research corpus is larger. Being able to use several discourse-annotated corpora at the same time, instead of just one, multiplies the amount of available data and unlocks a whole new set of research possibilities for the whole community.

We see several advantages of such an interface. It will allow researchers in the field of discourse to answer research questions like the ones mentioned above, making use of all annotated corpora, from all frameworks. Furthermore, the mapping will be useful for researchers and engineers working on automatic coherence relation labeling. Many natural language processing tasks, such as information retrieval and question-answering systems (Bosma 2005; Verberne et al. 2007), text summarization systems (Louis et al. 2010; Marcu 2000; Mithun 2010), and machine translation systems (e.g., Koehn 2009; Meyer and Popescu-Belis 2012; Meyer et al. 2011) would improve from increased performance in automated coherence relation classification. Current state-of-the-art coherence relation classification systems (see Xue et al. 2015 for an overview) make use of human-annotated coherence relations in corpora for training, especially the large resources PDTB and RST-DT. The performance of these tools, and generalizability from one text type to another would likely improve if more training data could be used. The mapping proposed here would enable researchers to train their models on all of the annotated resources, and not just those corresponding to a specific framework. Finally, we believe that a mapping between frameworks might help us extend current theories of coherence relations, because it will improve our understanding of the features defining different types of coherence relations, and pinpoint the exact differences and similarities between existing frameworks.

1.2 How to achieve a mapping

What are possible ways in which we could achieve the goal of mapping relation labels from different frameworks onto one another? We can think of three alternative ways of achieving this goal, depicted in Figure 1, and will discuss their respective advantages and disadvantages below.

(1)　One-to-one
One way of achieving our goal would be to define one-to-one mappings for each combination of frameworks separately. Benamara and Taboada (2015) have set
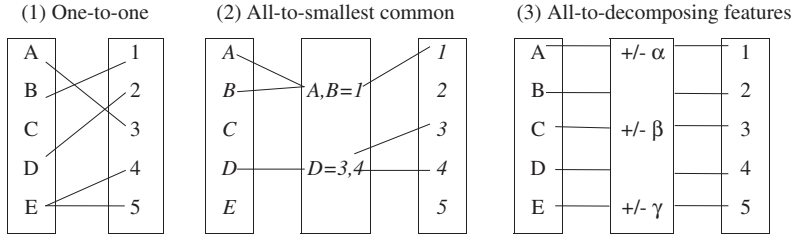
(1) One-to-one   (2) All-to-smallest common   (3) All-to-decomposing features

Figure 1: Possible mappings of relation labels from various annotation frameworks. (ABC = relation labels; 123 = relation labels; αβγ = decomposing features).

an inspirational example in this respect, by mapping RST and SDRT onto each other, and Chiarcos (2014) has made a similar effort by creating an ontological model of PDTB and RST labels. How would a one-to-one mapping work if we were to include more frameworks? If we had PDTB, RST, and SDRT, this would involve three mappings: PDTB to RST, RST to SDRT, and PDTB to SDRT. In order to map a fourth framework, we would need to create one-to-one mappings to the other three frameworks, and so forth for each new added framework. This solution would therefore be very time consuming.

An obvious idea might be to use the mappings transitively. For example, if we have a mapping from PDTB to RST, and one from RST to SDRT, we could automatically generate the mapping from PDTB to SDRT. While this might provide a good start, we would run into a problem: the mapping can only work out if RST can represent all the annotations in PDTB that are relevant for SDRT. If, on the other hand, there is an aspect of coherence relations that has been annotated in PDTB, but not in RST, then this information would get lost during the mapping to RST, and could not be recovered automatically for the mapping between PDTB and SDRT.

(2)   All-to-smallest-common

A second mapping option might be to come up with a scheme that comprises only those aspects that all the different frameworks have in common, and map all frameworks to this "shared" scheme. A problem with this approach is the difficulty of deciding what is the smallest shared among all. If we add all frameworks, would this boil down to Grosz and Sidner's (1986) set of just two relations? We would likely end up with just a very small, coarse set of coherence relations that all frameworks would map onto. For example, because SDRT makes fewer fine-grained distinctions than PDTB and RST, the "shared" scheme would cover fewer relation labels than PDTB and RST. If we were then interested in "translating" PDTB to an RST-format, we would end up with this smaller set

of labels, even though both PDTB and RST might make some of the same more fine-grained distinctions. For example, both RST and SDRT have labels for ELABORATION relations, but the PDTB does not. Such labels would therefore be lost by using this method. In sum, by mapping to a minimal shared scheme, we would lose information that has been very costly to obtain in hours of manual annotation labor.

(3)   All-to-a-set-of-decomposing-features

Finally, we could map everything to a scheme with features that can represent all the distinctions made by the various discourse annotation frameworks. This is the solution we pursue in this paper. To achieve this, we need a set of features or dimensions in terms of which all possible relations from all frameworks can be described. For instance, coherence relations marked by *although* have been described as negative causals, which indicates that they share the notion of causality with causal and conditional relations, and the notion of being "negative," "adversative," or "contrastive" relations with relations most typically marked by *but*.

    If we can uniquely describe all coherence relations in terms of their properties, this will allow us to map relation labels from different frameworks without any loss of information. All the information that has been annotated can be decomposed into a specific set of features. By using these features as an "interlingua," we can make translations between frameworks without having to do an all-to-all mapping.

    Here, the difficulty would be to express the common situation in which a framework does not make a distinction that is represented as a feature in the interlingua. For instance, consider an annotated resource with the label RESTATEMENT, such as RST, and a second one like PDTB, which distinguishes three subtypes of restatements: SPECIFICATION, EQUIVALENCE, AND GENERALIZATION. The mapping in terms of decomposing features should then also distinguish between these subtypes in order not to lose any information and be able to represent the annotations from both resources. However, how should the RST RESTATEMENTS be mapped onto PDTB's three subtypes? One solution that we put forward in this paper is to use different features for representing the positive, additive nature and specificity of the relation. This way, PDTB RESTATEMENTS are specified for each of these features, while RST relations are only specified as "positive additive," properties shared by all three relations, leaving the more fine-grained distinction on a "lower" level underspecified. This implies that in cases of underspecified features our interlingua will not generate a one-to-one mapping between RST labels and PDTB-labels. However, the "translation scheme"

does allow us to indicate which relation labels from RST are the most likely counterparts of the PDTB-labels, given their similar specification on other features. The interlingua also specifically indicates where the differences lie, which is something one-to-one mappings for different sets of frameworks would not do.

This approach requires a set of features that is extendable to include further subcategorizations. If we come across a language for which it becomes necessary to introduce a distinction that is not present in the existing annotation schemes, this can easily be accommodated in the mapping. For example, the adaptations from the PDTB that were made in the Prague Discourse Treebank (Poláková et al. 2013) could also be described in terms of the same dimensions. In other words, instead of assigning a very large set of different labels to coherence relations, we propose to describe coherence relations in terms of their common dimensions.

In the remainder of this paper, we will explain in more detail how the proposed "coherence relation annotation interlingua" works. We will start from an existing proposal that already focuses on identifying dimensions that all relations share. We investigate to what extent this proposal actually succeeds in describing all relations present in leading annotation frameworks, that is, succeeds in decomposing all relations in such a way that these frameworks can be related to each other. We will see that we need additional features to end up with our ideal: an interlingua.

2 Candidate dimensions shared by all coherence relations

Our goal is to find decomposing features in terms of which we can describe all possible coherence relations, so that we can make various annotation systems "talk to each other." Finding such unifying dimensions shared by all possible relations has played a central role in earlier work on relation categorization, namely the Cognitive approach to Coherence Relations (CCR; Sanders et al, 1992, 1993; see also Sanders 1997; Sanders and Spooren 2009a, 2009b). Originally, CCR proposed four dimensions (originally called "cognitive primitives") to organize the set of relations: polarity (positive vs. negative), basic operation (additive vs. causal), source of coherence (objective vs. subjective) and order of the segments (basic vs. non-basic). CCR has recently been extended with a fifth dimension, namely temporal order (temporal vs. non-temporal) (Evers-Vermeul et al. 2017). All dimensions were proposed on the basis of a *relational criterion*. A property of a coherence relation satisfies this criterion if it concerns

the informational surplus that the relation adds to the discourse segments in isolation. Before we can answer the question of whether these dimensions can be used to fulfill the task we have set ourselves in this paper, we will briefly introduce them as well as some other central concepts in this section.

2.1 Definitions of central concepts

According to CCR, coherence relations connect at least two discourse segments, which are taken to be minimally clauses – the simplest way to operationalize the observation that relations connect "idea units" in discourse.[3] Following the surface order in which they appear, the first segment is called S_1, and the second segment S_2. It is assumed that S_1 and S_2 directly or indirectly express the underlying propositions P and Q that are conceptually related. The relation itself is defined by the way in which S_1 and S_2 map onto P and Q, as well as the way in which P and Q relate to each other. These propositions can be events, states, speech acts, claims, opinions and judgments – they can even be inferences derived on the basis of S_1/S_2. Unless otherwise indicated, the examples in this section are all English translations of fragments taken from the Dutch DiscAn-corpus (Sanders et al. 2012).

2.1.1 Coherence relations and their linguistic markers

A defining characteristic of coherence relations is that the interpretation of the related segments needs to provide more information than is provided by the sum of the segments taken in isolation. Coherence relations are conceptual and they can, but need not, be made explicit by linguistic markers. Protypical lexical markers are so-called connectives (*and, because, so, however, although*) and lexical cue phrases (*for that reason, as a result, on the other hand*). Although these coherence markers vary from connectives to cue phrases and signaling phrases, and thus vary with respect to the grammatical categories they belong to (e.g., coordinating and subordinating conjunctions, adverbs), they have a very similar function when it comes to establishing discourse coherence: they are regarded as "processing instructions" (Sanders and Spooren 2009b). Over the last three decades experimental evidence has shown how the presence of these markers facilitates discourse processing and often improves comprehension (see van Silfhout et al. 2015 for an overview).

---

3 There has been, and there still is, a lot of debate about the exact definition of discourse segments. On top of that, not all frameworks use the same definitions, which further complicates comparison. We refer to Hoek et al. (2017a) for further definition and discussion.

The meaning and use of connectives is an important issue for another reason: often, language users systematically prefer one lexical item (e.g., *because*) over another (even highly similar) one (e.g., *since*) to express a relation. Such choices provide a window on speakers' cognitive categorizations, and have been modeled in many corpus studies in various languages. To what extent does connective use show relation categorization? Many publications since Halliday and Hasan (1976) have dealt with the question whether connectives specialize in expressing certain types relations (see Section 2.2 on specializations of connectives in the domain of causal relations). The general picture emerging from these studies is that connectives do specialize, although their interrelations are more subtle than a simple one-to-one assignment from connectives to classes of coherence relations would suggest (Knott and Dale 1994; Knott and Sanders 1998). This specialization allows us to use connective substitution tests as indicators of the type of relation at hand.

As explained above, coherence relations are inferred; the amount of inference they require depends – among other things – on the question whether they occur with or without their typical linguistic markers. We know that relations occur without such information in the majority of cases. This is a central methodological issue for all work in discourse annotation and interpretation: How to decide which relation holds? In this paper, we use the methods generally used in the field: we apply paraphrase tests to determine the most plausible interpretation, and we make sure that at least two analysts agree on this (Spooren and Degand 2010). The annotation manuals available for the three frameworks use similar interpretation steps.

2.1.2 Polarity: *Positive* versus *negative* relations

The first dimension refers to the polarity of the relation; it distinguishes between positive and negative (or adversative) relations. A relation is positive if the relation holds between P and Q, expressed in S_1 and S_2. A positive relation is typically expressed by connectives such as and or *because*. For example, in (5), S_2 has a direct link to S_1: Q in S_2 is an expected consequence of P in S_1.

(5)   [The stocks can decrease tremendously in value]$_{S1}$ and [thereby result in a loss for the investor.]$_{S2}$

A relation is negative if a negated version of either P or Q, not-P or not-Q, functions in the relation. In terms of connectives, negative relations can be typically expressed by connectives such as *but* and *although*, as is illustrated in (6).

(6)   [The biofuel is more expensive to produce,]$_{S1}$ but [by reducing the excise-tax the government makes it possible to sell the fuel for the same price.]$_{S2}$

In (6), where P is in S_1, a logical positive S_2 would be that the biofuel costs more, as a consequence of the higher production costs. However, S_2 presents a denial of this expectation; the fuel is not sold at a higher price due to a reduced excise-tax. S_2 expresses not-Q, that is, the negation of the consequent of the relation. This results in a negative relation. Note that not-P or not-Q may, but need not involve explicit negation. In (6), for instance, S_2 expresses not-Q without using a negation marker (as would be the case in "*not* impossible"), but still reverses the polarity of the consequence.

Polarity distinguishes contrastive, adversative and concession relations from all other relations. Hence, in terms of connectives and cue phrases, polarity distinguishes negative relations signaled by, for example, *but* or *although* from positive relations expressed by conjunctions such as and or *because*.

2.1.3 Basic operation: *Causal* versus *additive* relations

In discourse, segments can be strongly or weakly connected. On the basis of this pre-theoretical intuition, CCR distinguishes between causal and additive relations. More specifically, this dimension concerns the operation that has to be carried out on the two discourse segments. A relation is causal if an implication relation ($P \rightarrow Q$) can be deduced between the two discourse segments, as in (7).

(7)   [The athletics union was forced to emigrate to Belgium,]$_{S1}$ because [there was no accommodation available in the Netherlands.]$_{S2}$

In (7), the consequence Q is presented in S_1, and the cause P in S_2: a lack of accommodation has led to the emigration of the athletics union.

The category of causal relations also comprises conditional relations. They share the implicational nature of the relation, but at the same time they differ in the status of the *antecedent* (P) in the implication ($P \rightarrow Q$) relation. In conditionals, P is not yet realized; in causals it is. (8) is an example of a conditional relation, where P is in S_1 and Q in S_2.

(8)   If [you take one more step,]$_{S1}$ [I will arrest you.]$_{S2}$

In (8), the speaker confronts the listener with a condition. If the listener takes one more step, there will be a consequence: he will be arrested.

A relation is additive if the segments cannot be connected by an implication operation, but rather in a logical conjunction (P & Q), as in (9).

(9)   [The quality of this fuel with bio component is completely similar to Shell's regular Euro 95]$_{S1}$ and [the price at the pump is the same as well.]$_{S2}$

The relation in (9) consists of a conjunction of two segments that both describe a fact about fuel with a bio component. Temporal relations such as SEQUENCE (which can typically be expressed as: S_1 *And then* S_2) and TEMPORAL OVERLAP (with a prototypical connective expressed as: S_1 *Meanwhile* S_2) are additive relations too, but the events expressed in the segments are ordered in time (see Section 2.1.5 on order).

In sum, basic operation distinguishes between causals and conditionals on the one hand, and additive and temporal relations on the other hand. In terms of prototypical lexical markers: *because, so* and *if*, versus *and, also*, and *then*.

Similarly, basic operation distinguishes between causal and additive relations within the category of negative relations; causal relations are typically DENIALS OF EXPECTATIONS, as illustrated in (10).

(10)   Although [they were officially assured the police would not be involved in the census] [many people are afraid of reprisals ...]

Here, the causal operation lies between the fact that people were assured the police would not be involved (P in S_1) and the fact that people are still afraid of reprisals, which is not-Q (S_2) because the implication would be: *police not involved* (P) → *no need to be afraid of reprisals* (Q). It is this expected implication that is denied, and that is why *although* is used. A connective like *although* can often be used as the negative counterpart of *because*.

A case like (11) shows a negative additive relation: passengers say X (P), and Easyjet says not-X (Q). This is a classical OPPOSITION relation (Lakoff 1971; Spooren 1989).

(11)   [Passengers reported in the British tabloid Daily Express that the pilot tried to get the landing gear in position with strange nose dives] but [this is denied by Easyjet].

Apart from *but*, which is the most general adversative marker in English (Knott and Dale 1994; Knott and Sanders 1998), *however* is often used for such cases.

To avoid misunderstandings: in the above, the notions of implication and conjunction are used to define relations, and not connectives. Connectives are

mentioned as prototypical markers of some categories; they do not define relation categories. Well-known examples showing we have to be careful with assuming a direct relationship between connective and relations include conditional relations expressed with *and*: "Take one more step and I will arrest you" (compare the conditional in (8)).

2.1.4 Source of coherence: *Objective* versus *subjective* relations

The third dimension is source of coherence, which distinguishes between objective and subjective relations.[4] A relation is objective if the discourse segments are connected at the level of their propositional content. In other words, both segments describe situations in the real world, as in (12). The speaker merely reports these facts, and is not actively involved in the construction of the relation.

(12) [A cloud moved before the sun,]$_{S1}$ so [suddenly the room turned darker.]$_{S2}$

Relations are subjective if speakers or authors are actively engaged in the construction of these relations, either because they are reasoning, or because they perform a speech act in one or both segments. Subjective relations, such as (13), usually express the speaker's opinion, argument, claim or conclusion.

(13) [Drugs destroy people's lives,]$_{S1}$ so [drugs have to be battled judicially.]$_{S2}$

In (13), the statement in S_1 is not the cause for S_2, but an argument that is given to support the claim in S_2.

Hence, source of coherence distinguishes between relations describing real world-situations such as temporal sequence (SUCCESSION) and cause-consequence (RESULT) or consequence-cause (REASON), from typical argumentative relations such as claim-argument (EVIDENCE, JUSTIFICATION) and CONCLUSION.

In various languages, the connective lexicons make a systematic distinction that can be described in terms of source of coherence. This has been investigated in more or less detail for causal connectives: Dutch *doordat* and *omdat* versus *want*, French *parce que* versus *car* and *puisque* and German *weil* versus *denn*

---

4 Originally, CCR labeled these categories as semantic versus pragmatic relations. Other terms have been used to refer to similar categories: external versus internal (Halliday and Hasan 1976; Martin 1992), propositional vs. illoucutionary (Sanders and Spooren 1999), subject matter versus presentational (Mann and Thompson 1988) or content versus epistemic and speech act use (Sweetser 1990), see Sanders and Spooren (2009a: 208).

(see Section 2.2). However, English *because* is well-known for being able to express all types of positive causal relations (Ford 1993; Knott and Dale 1994; Knott and Sanders 1998; Sweetser 1990). Still, English *since* seems to have a preference for epistemic/subjective relations, whereas *for that reason* and *as a result* seem to be cue phrases for typical objective causal relations (Andersson 2016).

2.1.5 Implication order: *Basic* versus *non-basic* relations

The final CCR dimension concerns what was originally called the order of the segments, we will however refer to this dimension as the implication order, as it only applies to causal relations. The implication order dimension is not applicable to additive relations, as they are symmetrical by definition. Given two discourse segments in a causal relation, they can be connected in a basic or a non-basic order. This implication order is defined in terms of the mapping of P and Q onto the segments. A relation with a basic order has an antecedent P as S_1, followed by a consequent Q in S_2, as in (14). The antecedent is the cause or the argument, and the consequent is the consequence or the claim. In a relation with a non-basic order, such as (15), the consequent precedes the antecedent.

(14)　Sometimes children tease me. But [I don't reply,]$_{S1}$ that's why [they don't do it anymore.]$_{S2}$

(15)　[Universities supposedly cancel subscriptions to scientific journals more often]$_{S1}$ because [there is more information available through the internet.]$_{S2}$

Thus, implication order distinguishes cause-consequence relations (S_1 *As a result/So* S_2) from consequence-cause relations (S_1 *because* S_2) and claim-argument (S_1 *since* S_2) from argument-claim or conclusion (S_1 *therefore* S_2) relations.

2.1.6 Temporality: *Temporal* versus *non-temporal* relations

Temporality defines whether two segments are ordered in time. When S_1 and S_2 display a temporal sequence (e.g., SEQUENCE or PRECEDENCE), or temporal overlap, the relation is temporally ordered. If temporal order is not relevant to the relation at hand, as is the case in for example purely additive relations, we talk about a non-temporal relation. Within the category of temporal relations, a further distinction, referred to as temporal order, is made between chronological (with event 1 in S_1 and event 2 in S_2), anti-chronological (with event 2 in S_1 and

event 1 in S_2) and synchronous relations (where events 1 and 2 described in S_1 and S_2 happen simultaneously).

As Evers-Vermeul et al. (2017: 13) illustrate using example (16), implication order may but need not coincide with temporal order. For example, the implication order in (15) is non-basic and the temporal order is anti-chronological, so the two orders are in agreement. By contrast, the implication order in (16) is basic, because the antecedent P is expressed in S_1 and the consequent Q in S_2, but the temporal order is anti-chronological, since getting off the couch in S_2 would have to occur before getting rich in S_1.

(16)   If [you want to become rich someday]$_{S1}$ [you should probably get off the couch.]$_{S2}$

2.2 Empirical evidence for CCR-dimensions

CCR was inspired, among others, by Hobbs (1979), Mann and Thompson (1986), and Meyer (1985), and is closely related to Kehler's (2002) take on coherence relations. Over the last decennia, evidence was found in favor of the CCR dimensions, both in psycholinguistic and corpus-based studies.

The order in which children acquire connectives and relations confirms the cognitive relevance of dimensions such as polarity and basic operation – children produce positive connectives before negative ones and additives before causals (Bloom et al. 1980; Evers-Vermeul and Sanders 2009) – as well as source of coherence: children master objective relations before subjective ones (Evers-Vermeul and Sanders 2011; Zufferey 2010). In studies on discourse processing and representation, the difference between positive and negative relations is a classical finding, going back to Clark and Clark (1977). The special status of causal relations found by many researchers (Graesser et al. 1994; Noordman and Vonk 1998; Sanders and Noordman 2000; Sanders and Spooren 2009a; Singer et al. 1992; van den Broek 1990) corroborates the relevance of basic operation, and recent eye-tracking studies have shown that subjective relations are harder to process than objective ones, in English (Traxler et al. 1997), Dutch (Canestrelli et al. 2013) and French (Zufferey et al. 2018). Evidence for the primitive implication order comes from processing experiments (Noordman and de Blijzer 2000), in which it is shown for objective or content-relations that a basic cause-consequence order is easier to process than a non-basic consequence-cause order. Furthermore it has been shown that chronologically ordered sentences are processed faster and remembered better than sentences in which the order of mention does not correspond to the order in which the events take place in the real world (Baker 1978; Clark and Clark 1968; Townsend 1983).

Results of corpus-based studies suggest that the lexicons of connectives in various languages are organized according to similar dimensions. Even though some languages have been studied more intensively than others, recent research (see Degand and Pander Maat 2003; Stukker and Sanders 2012; for an overview on causal connectives) suggests that the relevance of the CCR-dimensions not only holds for closely related languages like Dutch, English, French and German, but also for typologically less-related languages like Mandarin (Li et al. 2013; Li et al. 2016). This research program is far from finished, however, and is currently being tested on more languages. In addition, the CCR-dimensions have been used to annotate corpora of coherence relations (Sanders and Scholman 2015; Sanders et al. 2012), children's language use (Evers-Vermeul 2005; van Veen 2011), parallel corpora (Hoek et al. 2017b), and to organize the complex task of relation annotation in a step-wise approach (Scholman et al. 2016).

2.3 Using CCR-dimensions to decompose coherence relations

In order to achieve a mapping between different frameworks for discourse annotation, we will decompose their relation labels in terms of the five CCR-dimensions. This in turn will allow us to find out whether and how relation labels from different frameworks can be mapped onto each other. In this section, we illustrate this mapping attempt on examples taken from RST-, SDTR-, and PDTB-corpora.

First, decomposing labels from other frameworks into CCR-dimensions allows us to check whether identical labels in different systems should actually be treated as identical. If this is the case, these labels can be expected to have the same values along the CCR-dimensions. Take, for example, the RST label REASON for cases like (17), and the PDTB label REASON for instances such as (18).

(17)   Earlier this year, DPC Acquisition made a $15-a-share offer for Dataproducts, [which the Dataproducts board said it rejected] [because the $283.7 million offer was not fully financed.] (wsj_0661)

(18)   [Use of dispersants was approved] [when a test on the third day showed some positive results,] officials said. (1347)

Indeed, both types of relations can be described as positive, causal, objective, in an anti-chronological, non-basic order. This similarity between relations holds, even though they are marked by different connectives. The use of the temporal connective *when* in (18) is a typical case of underspecification (Spooren 1997):

causal relations can be marked by temporal connectives (and in some cases also by very general connectives like *and*), but still the causal interpretation prevails because the segments describes facts or events that are causally related.

Second, decomposing labels into CCR dimensions can also be applied to labels that appear to cover similar relations but do not have the exact same label across annotation systems. Example (19) expresses a CONTRAST relation according to RST, whereas the relation expressed in PDTB-fragment (20) is one of OPPOSITION.

(19)　[But from very early on, Tiger's workers unionized,] [while Federal's never have.] (wsj_1394)

(20)　[Most bond prices fell on concerns about this week's new supply and disappointment that stock prices didn't stage a sharp decline.] [Junk bond prices moved higher, however.] (wsj_1464)

These different relation labels can be considered similar, as both would be described in identical CCR-terms: negative, additive, objective, synchronous ("implication order" does not apply).

Third, the decomposition into CCR dimensions can be applied to labels that at first glance appear to cover different relations. These may turn out to be different indeed, or to actually represent highly similar relations. We will illustrate the latter situation with instances of implicit relations. Fragment (21) is a case of JUSTIFICATION according to PDTB, whereas (22) expresses EVIDENCE in RST. The PDTB-analysis notes an implicit *because* between the segments.

(21)　[Mrs Yeargin is lying.] [They found students in an advanced class a year earlier who said she gave them similar help.] (wsj_0044)

(22)　[That system has worked.] [The standard of living has increased steadily over the past 40 years ...] (wsj_1120)

In both cases an author's claim is expressed, which is supported by an argument. In CCR-terms, these cases would be analyzed as positive, causal, subjective, synchronous, non-basic. This would place the different labels JUSTIFICATION (PDTB) and EVIDENCE (RST) in the same category, thus uncovering the underlying similarity between them that is not obvious from the end labels alone.

The fourth and final type of illustration concerns cases in which it is unclear how relation labels match, because one framework seems to make less specific distinctions than another. Decomposing relations in terms of

dimensions can then help determining how seemingly different relations are related. For example, in the SDRT-framework, a case like (23) is labeled with a generic CONTRAST tag, because there is only one type of contrastive relationship.

(23)   [While European allies have closed ranks behind the United States in the latest showdown with Iraqi President Saddam Hussein,]$_{S1}$ [there are lingering apprehensions among NATO governments about the Clinton administration's insistence on recognizing the spread of nuclear, biological and chemical weapons as the alliance's most urgent priority.]$_{S2}$ (Reese et al. 2007: 18)

This example illustrates the case in which an expectation is not fulfilled. The paraphrase of the causal basic operation is roughly equivalent to: "European allies agree with the USA (P), so (Q) they will also be enthusiastic about the USA's insistence on recognizing the spread of weapons as an urgent priority." Under CCR, this example would be analyzed as a negative causal relation: Q in the basic operation is denied because they are *not* enthusiastic about the idea.

In a PDTB CONTRA-EXPECTATION relation such as (24), we also see that an expectation on the basis of S_1 does not work out in S_2. This is often signaled by *although*, but in this case the more general contrastive marker *but* is used.

(24)   [The Texas oilman has acquired a 26.2% stake valued at more than $1.2 billion in an automotive-lighting company, Koito Manufacturing Co.] [But he has failed to gain any influence at the company.] (wsj_0082)

Again, using paraphrases to analyze this relation in CCR-dimensions, this type of DENIAL OF EXPECTATION would be characterized as negative causal. Hence, both (23) and (24) would be labeled as: negative, causal, synchronous, non-basic. For these specific examples, this implies that PDTB relations labeled as CONTRA-EXPECTATION can be mapped onto SDRT's label CONTRAST. However, because SDRT's CONTRAST also covers other types of negative relations (e.g., negative additive ones), no one-to-one mapping between CONTRAST and CONTRA-EXPECTATION is possible.

In Section 3, we will systematically investigate how far the CCR dimensions bring us when we are looking for an interlingua between existing annotation systems. In Section 4, we will investigate which additional criteria we need to consider.

3 "Translation" of RST, PDTB, and SDRT labels into CCR dimensions

In this section, we apply the CCR dimensions to relation labels from three leading discourse annotation frameworks: RST, PDTB, and SDRT. Specifically, we have looked at the relation labels mentioned in the following annotation manuals: Carlson and Marcu (2001) for RST, PDTB Research Group (2008) for PDTB, and Reese et al. (2007) for SDRT, in order to test our "translation effort" on data from available resources.[5] An additional benefit of using these manuals is that the accompanying corpora are all in English, which facilitates a comparison between the application of relation labels. Finally, there are many corpora, both in English and in other languages, with relation inventories based on or similar to these three manuals, to which many of the observations in this paper are relevant as well, for instance the RST Basque Treebank (RST; Iruskieta et al. 2013), the Prague Discourse Treebank (PDT; Poláková et al. 2013), and ANNODIS in French (SDRT; Afantenos et al. 2012).

For each CCR dimension, we addressed three questions:

(i) Do the three frameworks make this distinction? If so, how?

(ii) If not, is it possible to map the dimension on their classification?

(iii) Where do the problems come from when the dimension is not easily translatable between frameworks?

The translation of the relation labels of the three frameworks into dimensions is made on the basis of the definitions of the relation labels and accompanying examples provided by the manuals. Four authors of this paper discussed the CCR classification of each relation label until consensus was reached. We adhered to the labels and information provided in the manuals, even if in some cases we disagreed with the categorization of certain examples. A complete overview of the classification of SDRT, PDTB, and RST relation labels in terms of dimensions is provided in Appendix A, B, and C, respectively.

3.1 Polarity

As we have explained in Section 2.1.1, polarity pertains to whether the relation features P and Q, or a negative counterpart of either P (not-P) or Q (not-Q). In

---

5 Unless stated otherwise, we resort to the 2.0 version of the PDTB. Although the 3.0 version (Prasad et al. in preparation; Webber et al. 2016) has not been officially released, we will extend our discussion to that version when crucial differences can be observed.

general, RST, PDTB, and SDRT all distinguish relations with a positive value for polarity from relations with a negative value for polarity. As such, determining polarity is fairly straightforward for most of these frameworks' relation labels. For instance, the SDRT category CONTRAST involves negative relations, just like PDTB categories such as OPPOSITION and EXPECTATION, and RST categories such as CONTRAST and CONCESSION.

CCR takes the stance that all positive relations have a negative counterpart, but the polarity dimension does not seem to have been implemented to the same extent in the other frameworks. RST, PDTB, and SDRT all include many more positive relations than negative relations, with SDRT at the extreme with only two negative versus ten positive relations. For SDRT's CONTRAST, a negative value for polarity is even the only defining relation characteristic, while no relation can be found for which a positive value for polarity is cited as the main relation characteristic. A negative value for polarity thus appears to be a very prominent relation attribute, and negative relations are often found to be underspecified for other dimensions (see Sections 3.2–3.4).

There are a few relation labels that seem to include both positive and negative relations. One example is RST's BACKGROUND, in which one segment "establishes the context or grounds with respect to which [the other segment] is to be interpreted" (Carlson and Marcu 2001: 47). This definition seems to straightforwardly capture positive (additive) relations. In (25), one of the examples of BACKGROUND provided by Carlson and Marcu (2001: 47), this is indeed the case. However, another example that Carlson and Marcu include for BACKGROUND is (26).

(25)   [The Voting Rights Act of 1965 was enacted to keep the promise of the Fifteenth Amendment and enable Southern blacks to go to the polls, unhindered by literacy tests and other exclusionary devices.] [Twenty-five years later, the Voting Rights Act has been transformed by the courts and the Justice Department into a program of racial gerrymandering designed to increase the number of blacks and other minorities – Hispanics, Asians and native Americans – holding elective office.] (wsj_1137)

(26)   [Banco Exterior was created in 1929 to provide subsidized credits for Spanish exports.] [The market for export financing was liberalized in the mid-1980s, however, forcing the bank to face competition.] (wsj_0616)

In (26), the relation between the two segments is signaled by *however*, a prototypical marker of negative relations. In CCR, the relation in (26) would indeed be

annotated as a negative relation. It appears, therefore, that Carlson and Marcu allow for some form of contrast between a segment and its "background." Since RST's CIRCUMSTANCE is only distinguished from BACKGROUND in the "strength" of the relation between the two segments, and the perceived "co-temporality" of the two segments (see Carlson and Marcu 2001: 47–48), we conclude that CIRCUMSTANCE will most likely also include negative relations.

Other examples of RST relations that are underspecified for polarity are CONDITION and CONTINGENCY. CONDITION includes both positive conditional relations, prototypically signaled by *if*, and negative conditional relations, prototypically signaled by *unless*. In CONTINGENCY relations, one of the segments "suggests an abstract notion of recurrence or habituality" of the other segment (Carlson and Marcu 2001: 52). Similar to CONDITION, CONTINGENCY includes examples of both positive conditional relations and negative conditional relations. Both relation labels are therefore underspecified for polarity. However, using the connectives, if present, the majority of the relations with RST's CONDITION and CONTINGENCY labels can probably be easily specified for polarity.

RST's DISJUNCTION and SDRT's ALTERNATION appear to include both positive, or "inclusive," disjunctions, in which "P or Q" actually stands for "P and/or Q," and negative, or "exclusive," disjunctions, in which "P or Q" stands for "either P or Q, but not both P and Q." Both labels have therefore been indicated to be underspecified for polarity. PDTB, on the other hand, does have designated labels for inclusive and exclusive disjunctions: CONJUNCTIVE and DISJUNCTIVE, respectively.

On the basis of the relation definitions and the examples provided in the manual, all PDTB relations can be classified as either positive or negative. However, looking at the connective and cue phrase inventory in its appendices (PDTB Research Group 2008: 71 and 87), it appears that a small portion of CONJUNCTION relations, which we initially classified as relations with positive polarity, are signaled by connectives or cue phrases that typically mark negative relations.[6] Based on this observation, we decided to leave the polarity feature of PDTB's CONJUNCTION underspecified.

Although there are a few problematic categories, most relation labels in RST, PDTB, and SDRT can be classified as being either positive or negative

---

6 For explicit relations, at least 2.0% of all CONJUNCTION relations are negative, if we only include connectives predominantly signaling negative relations such as *but* and *nonetheless*. If we also include all relations signaled by *while*, a connective that can be used to signal both negative relations and positive (temporal) relations, the proportion reaches 3.3%. For implicit CONJUNCTION relations, 1.7–11% have a negative value for polarity.

relations. For specific relations from categories that are underspecified for polarity, it will often be possible to determine whether they have a positive or a negative polarity value on the basis of the connective used to signal the relation.

3.2 Basic operation

Basic operation as defined in Sanders et al. (1992) and Section 2.1.3 captures the basic intuition that discourse segments are either strongly connected (causality) or weakly connected (addition). In Section 3.2.1 we will discuss whether the different frameworks make this or a similar distinction in their inventory of relation labels. Within the class of causal relations, conditional relations can be distinguished as a subtype, which will be discussed in more detail in Section 3.2.2.

3.2.1 Causal versus additive relations

Other frameworks do not use the term "basic operation" explicitly, but they do generally present clusters of relations along this dimension or subclasses thereof. For example, all relations in the PDTB classes EXPANSION are additive – apart from DISJUNCTIVE, which can also be causal-conditional – whereas relations in the class CONTINGENCY all involve an implication relation, and hence can be classified as causal. Similarly, the RST classes BACKGROUND, ELABORATION, JOINT, SUMMARY, and TEMPORAL contain only additive relations, whereas the classes CAUSE, CONDITION, ENABLEMENT, and EXPLANATION are all causal in nature. SDRT does not present such clusters of relations, but the majority of the SDRT end labels can easily be assigned either an additive or a causal value for basic operation. For example, CONTINUATION and NARRATION are additive relations, whereas EXPLANATION, RESULT, and CONSEQUENCE are causal ones. For all three frameworks, however, there are a few relations or classes of relations that cannot be mapped as easily.

Two relations in the RST class EVALUATION, namely EVALUATION and INTERPRETATION, are underspecified for their basic operation. For example, the INTERPRETATION relations in (27) and (28), taken from Carlson and Marcu (2001: 60–61), both involve "the personal opinion of the writer or of a third party" (Carlson and Marcu 2001: 60). However, the example in (27) presents a simple list of such opinions, whereas in (28) the personal opinion is the claim in an argument-claim relation. In other words, (27) is additive and (28) is causal.

Similarly, an EVALUATION relation can contain a list of two or more evaluations or an evaluative conclusion for which an argument is presented.

(27) But John LaWare, a Fed governor, told the sub-committee [the evidence is mixed] [and that the Fed's believes (sic) the vast majority of banks aren't discriminating.] (wsj_1189)

(28) [Far from promoting a commonality of interests among black, white, Hispanic and other minority voters, drawing the district lines according to race suggests that race is the voter's and the candidate's most important trait.] [Such a policy implies that only a black politician can speak for a black person, and that only a white politician can govern on behalf of a white one.] (wsj_1137)

PDTB and SDRT would distinguish between these two INTERPRETATION relations based on their basic operation: in PDTB, these relations would be analyzed as CONJUNCTION and RESULT, respectively; and in SDRT, example (27) would be analyzed as a PARALLEL relation, and (28) as a RESULT relation.

A similar type of underspecification can be found in SDRT. CONTRAST in SDRT holds when two segments have "contrasting themes, i.e., sentence topics, or when one constituent negates a default consequence of the other" (Reese et al. 2007: 17). This definition and the accompanying examples on page 18 in the manual make it clear that this end label contains both additive and causal relations. While SDRT's CONTRAST is underspecified, both PDTB and RST distinguish these subtypes of negative relations along the dimension of basic operation: CONTRAST and CONCESSION, respectively.

Another instance of underspecification in terms of the basic operation dimension is found in both PDTB and SDRT. PDTB's category of DISJUNCTIVE as well as SDRT's ALTERNATION can be either negative additive, as in (29), or negative causal(-conditional), as in (30) (both examples taken from Reese et al. 2007: 19).

(29) either [by TWA's acquisition of USAir,] or [USAir's acquisition of TWA.]

(30) ["It looks like we're going to be there on the street,] unless [there is a miracle,"]

RST would analyze (29) as a DISJUNCTION, belonging to the class JOINT, and (30) as a CONDITION. Hence, RST categorizes these relations according to their basic operation, whereas SDRT and PDTB do not.

3.2.2 Conditional relations

CCR considers conditional relations to be a subtype of causal relations. PDTB makes a similar distinction by placing causals and conditionals in the same class of CONTINGENCY relations, but distinguishing between the classes of CAUSE and PRAGMATIC CAUSE on the one hand, and CONDITION and PRAGMATIC CONDITION on the other. In the annotation schemes of RST and SDRT, conditional relations are also distinguished from causal relations, although they are not clearly classified as a subtype of causals in these schemes. For example, RST distinguishes the positive causal classes CAUSE, ENABLEMENT, and EXPLANATION from the positive conditional class CONDITIONAL, and SDRT distinguishes the positive causal labels RESULT and EXPLANATION from the positive conditional CONSEQUENCE.

There are a few end labels that comprise both conditional relations and non-conditional causal relations or even additive relations. Two of these were discussed in Section 3.2.1: PDTB's DISJUNCTIVE relation and SDRT's ALTERNATION relation. These relation labels contain both negative additive and negative conditional relations. Another relation label that is underspecified in terms of conditionality, involves RST's PROPORTION: it can be conditional causal, as in (31), or non-conditional causal, as in (32).

(31)   When [the little guy gets frightened,] [the big guys hurt badly.] (wsj_2386)

(32)   [The higher the risk,] [the higher the return.] (financial concept)

In RST, these relations are categorized in the same class because they express a proportionality or equivalence of tendency between two segments. In other frameworks, these relations would likely be classified differently, thereby disregarding the degree of proportionality. For example, in PDTB, example (31) would be classified as GENERAL CONDITION, and (32) as a RESULT relation. In SDRT, example (31) would most likely be classified as a CONSEQUENCE relation, and (32) as a RESULT relation. Since PDTB and SDRT do not have end labels similar to RST's PROPORTION, a mapping for PROPORTION to these frameworks would require additional manual annotation.

To conclude, an additional feature for CONDITIONALITY should facilitate the mapping of different frameworks onto each other, but more features might be necessary in order to be able to accommodate the more fine-grained distinctions that RST and PDTB 2.0 make (see Section 4.2.1).

3.2.3 Conclusion on basic operation

In sum, RST, SDRT, and PDTB all mostly contain relations or classes of relations with a specific basic operation. For some relations, the basic operation remains underspecified. However, in these cases connectives – explicitly mentioned or added by the annotator – can be used to determine the basic operation. For example, markers such as *unless* and *if* indicate that the relation is causal or causal-conditional rather than additive, and can hence be used to make a further distinction within SDRT's category of ALTERNATION relations or within PDTB's DISJUNCTIVE relations. Note that not all markers are equally suitable for this dimension. For example, markers such as *but* can be used to mark negative additive as well as negative causal relations, and *when* can be used to mark both conditional (i.e., causal) and temporal (i.e., additive) relations.

3.3 Source of coherence

The third dimension is source of coherence, which distinguishes between objective (semantic) and subjective (pragmatic) relations (see Section 2.1.4). PDTB appears to make a similar distinction within the classes of CONTINGENCY and COMPARISON: it discriminates CAUSE (REASON and RESULT), CONDITION (six types of conditional relations), CONTRAST (JUXTAPOSITION and OPPOSITION), and CONCESSION (EXPECTATION and CONTRA-EXPECTATION) on the one hand, from PRAGMATIC CAUSE (JUSTIFICATION), PRAGMATIC CONDITION (IMPLICIT ASSERTION and RELEVANCE), PRAGMATIC CONTRAST, and PRAGMATIC CONCESSION on the other.

On closer inspection, however, a simple one-to-one mapping between pragmatic PDTB-labels and the value "subjective" on the one hand, and between non-pragmatic PDTB-labels and the value "objective" on the other does not seem to do justice to the actual source of coherence that is covered by the PDTB-labels at hand. The mismatch in the mapping arises mainly from the fact that certain non-pragmatic PDTB-labels are actually subjective or underspecified for source of coherence instead of objective. For example, the labels FACTUAL PAST and FACTUAL PRESENT are presented as non-pragmatic and might therefore – at first glance – be considered objective conditional relations. However, following Sweetser's line of reasoning (Dancygier and Sweetser 2005; Sweetser 1990), we propose to treat conditional relations with given or true *if*-clauses as epistemic and therefore subjective relations. For example, in (33), the information in the *if*-clause is presented as a background for the subjective prediction or conclusion in S_2. On the basis of this line of

reasoning, all relations with the label FACTUAL PAST and FACTUAL PRESENT were given the value "subjective" on the dimension of source of coherence.

(33) If [they had this much trouble with Chicago & North Western,]$_{S1}$ [they are going to have an awful time with the rest.]$_{S2}$ (wsj_1464)

For other relation labels no such one-to-one mapping can be achieved. For instance, it is conceivable that UNREAL PAST and UNREAL PRESENT conditionals can contain subjective as well as objective relations, despite the fact that PDTB classifies both types of relations as non-pragmatic. PDTB lists (34) as an example of an UNREAL PRESENT relation; in terms of CCR, this relation can indeed be considered objective. The adapted version in (35) would receive the PDTB-label UNREAL PAST, but in terms of CCR is classified as a subjective relation, since S_2 presents an evaluation of the speaker given the condition in S_1. In our mapping overview in Appendix B, we therefore labeled these two types of conditionals as underspecified for source of coherence, despite the fact that PDTB classifies them as non-pragmatic.

(34) If [the film contained dialogue,]$_{S1}$ [Mr Lange's Artist would be called a homeless person.]$_{S2}$ (wsj_0039)

(35) If [the film had contained dialogue,]$_{S1}$ [that would have been great.]$_{S2}$

A similar mismatch in the mapping appears to occur in the area of causal relations. While an objective REASON relation such as (36) has a subjective counterpart in (PRAGMATIC CAUSE) JUSTIFICATION (compare (37)), no such counterpart can be found in PDTB's label inventory for RESULT. JUSTIFICATION only applies when the clause following the connective provides a justification for the claim presented in the other clause, as in (37). The PDTB Research Group (2008: 29) admits that "no instances have been found in the corpus of an explicit or implicit connective in which 'Pragmatic cause' holds in the opposite direction," that is, with the connective clause expressing the claim and the other clause its justification. This could imply two things: either relations such as (37)'s counterpart in (38) did not occur at all in the corpus, or they simply ended up in the category of RESULT relations, which would turn this label into one that is underspecified for source of coherence because it also includes objective relations such as (39).[7]

---

7 Evidence for the latter option comes from Appendix A in the PDTB Annotation Manual (PDTB Research Group 2008: 69), in which 262 out of 263 *so*-cases are classified as RESULT relations.

(36) [Use of dispersants was approved]$_{S1}$ when [a test on the third day showed some positive results]$_{S2}$ (wsj_1247)

(37) [Mrs Yeargin is lying]$_{S1}$ (because) [They found students in an advanced class a year earlier who said she gave them similar help]$_{S2}$ (wsj_0044)

(38) [They found students in an advanced class a year earlier who said Mrs Yeargin gave them similar help]$_{S1}$ so [she is lying]$_{S2}$

(39) In addition, [its machines are typically easier to operate,]$_{S1}$ so [customers require less assistance from software]$_{S2}$ (wsj_1887)

Although PDTB distinguishes pragmatic labels for causal and conditional relations, there are no pragmatic counterparts for additive relations, which PDTB clusters under the class of EXPANSION relations, nor for the class of TEMPORAL relations. All additive relations therefore remain underspecified for source of coherence, while all relations in the class of TEMPORALS can be considered objective.

SDRT and RST do not systematically distinguish objective and subjective counterparts of relations that are otherwise very similar. For both frameworks, however, specific end labels can be regarded as objective, whereas others can be seen as clearly subjective. SDRT's NARRATION and PRECONDITION, for example, are both objective in CCR's terms, because they concern relations that describe real world events that are ordered in time (be it chronologically or anti-chronologically), and COMMENTARY can be labeled as a subjective relation, because one segment in this relation "provides an opinion or evaluation" of the content associated with the other segment (Reese et al. 2007: 15). Other relations in SDRT are underspecified for source of coherence. This holds, for instance, for CONTRAST, which covers both objective and subjective negative relations, and CONSEQUENCE, which covers objective as well as subjective conditional relations.

Similarly, RST contains several relation labels that are underspecified for source of coherence. Typical instances are various types of negative relations (e.g., CONTRAST, CONCESSION, and ANTITHESIS), and LIST relations, which can contain a list of either objective facts or subjective arguments. Still, RST does use

Notably, all 26 relations marked by *therefore* also receive a RESULT label, even though this connective can be used in both objective and subjective relations (Sweetser 1990: 78–79). This strengthens our belief that RESULT actually covers objective as well as subjective relations, despite the fact that PDTB's definition of this label hints at only objective relations.

subjectivity as a defining feature for specific end labels. This can be seen, for example, in the definition of EVALUATION:

> In an evaluation relationship, one span assesses the situation presented in the other span of the relationship on a scale of good to bad. An evaluation can be an appraisal, estimation, rating, interpretation, or assessment of a situation. The evaluation can be the viewpoint of the writer or another agent in the text. (Carlson and Marcu 2001: 57)

The definition of INTERPRETATION even mentions the word *subjective*: "In interpretation relations, one side of the relation gives a different perspective on the situation presented in the other side. It is subjective, presenting the personal opinion of the writer or of a third party" (Carlson and Marcu 2001: 60). Applying the same line of reasoning, RST's CONCLUSION and EVIDENCE can also be classified as subjective relations.

In addition, several RST-definitions refer to features that allow us to univocally classify them as objective relations. For example, in a CONSEQUENCE relation, "the situation presented in one span is a consequence of the situation presented in the other span. The reader/writer intentions are irrelevant to determining whether such a relation holds" (Carlson and Marcu 2001: 52). Because real world causality is at stake here, CONSEQUENCE can be seen as an objective relation. This is also the case for the RST labels CAUSE, RESULT, CAUSE-RESULT, and REASON. Remarkably, EXPLANATION-ARGUMENTATIVE relations should also be classified as objective relations according to RST's definition, despite the fact that their label suggests otherwise. Other clear examples of objective relations are RST's temporal relations.

In conclusion, classifying PDTB-labels in terms of CCR's source of coherence is not as easy as the PDTB distinction between pragmatic and non-pragmatic labels seems to suggest. Furthermore, SDRT and RST do not systematically distinguish objective and subjective counterparts of relations that are otherwise very similar, but given the definitions provided within these frameworks, several relation labels can be assigned either an objective or a subjective value. Overall, we have labeled about one fifth of all the relation labels in PDTB, RST, and SDRT as underspecified for source of coherence. Additional manual annotation would be required to classify specific instances of these relations in terms of source of coherence.

3.4 Implication order

The fourth dimension defined as part of CCR is implication order. CCR defines this implication order in terms of the mapping of P and Q onto S_1 and S_2 (see

Section 2.1.5). Examples (40) and (41) are positive causal relations with basic order, since P in S_1 is the cause for Q in S_2; the positive causal relation in (42) has a non-basic order, since Q in S_1 is caused by P in S_2.

(40) [We had sold only 35 tickets]$_{S1}$ so [we cancelled the event.]$_{S2}$

(41) Because [we had sold only 35 tickets,]$_{S1}$ [we cancelled the event.]$_{S2}$

(42) [We cancelled the event]$_{S1}$ because [we had sold only 35 tickets.]$_{S2}$

SDRT's definition of implication order is compatible with the one used in CCR. Examples (40) and (41) would for instance be labeled RESULT by SDRT, and (42) would be labeled EXPLANATION. However, SDRT also has a few relation labels for which implication order is not specified; for instance, CONTRAST and ALTERNATION. As we discussed in Section 3.2, relations with these labels are underspecified for basic operation. As a result, they contain relations for which implication order is relevant (causal and conditional), as well as relations for which this order is not relevant (additive). In Appendix A, these relation labels are marked as underspecified for implication order.

The mapping of the CCR dimension implication order onto RST and PDTB relation labels is not so straightforward. RST and PDTB both distinguish between similar relations with opposite orders in their relation inventory (on the basis of the nucleus-satellite, respectively, Arg1-Arg2 order), but their notions of order differ from implication order as employed by CCR. These notions will be discussed in detail in Section 3.6.

Note however also that the basic idea underlying implication order could also be applied to temporal relations, substituting the criterion that is defined in terms of the implication between P and Q by a criterion which is defined based on temporal sequence between P and Q. Evers-Vermeul et al. (2017) introduced this notion into CCR, under the label of temporality; given its close relationship to implication order, we will discuss temporality next.

3.5 Temporality

The final dimension is temporality, which distinguishes between relations for which temporal order is not relevant (non-temporal relations) and relations displaying a chronological, anti-chronological or synchronous temporal relation. The temporality dimension is relevant to relations with either an additive or a causal basic operation, but only additive relations

with segments order in time correspond to what are typically referred to as "Temporal relations." That is, Temporal relations in other frameworks can be decomposed as additive relations with a temporal order using the dimensions. In the other frameworks, relations that are causal are not considered to be Temporal relations, but their segments can be ordered in time, and hence can also be assigned the value "temporal" along CCR's dimension.

All frameworks give the temporal nature of the relation a prominent position by setting temporal relations apart (e.g., SDRT's Narration, PDTB's class of Temporals, and RST's class of Temporals) and/or distinguishing different types of temporal relations. For example, SDRT discriminates Preconditions – with a sequential order – from Background relations, some of which are synchronous in nature. And PDTB discriminates between Precedence and Succession relations, just like RST sets the mononuclear relation Temporal-before apart from Temporal-after and Temporal-same-time, and makes a distinction between the multinuclear relations Sequence, Inverted-Sequence and Temporal-same-time.

Although most labels can easily be assigned either a "temporal" or a "non-temporal" value for the temporality dimension, we came across a few labels that seem to contain both temporal and non-temporal relations. Background and Circumstance in RST, and Background in SDRT have comparable relation definitions: one of the segments provides extra information about the other segment. The additional distinction that RST makes between Background and Circumstance is that in Circumstance relations the information or context is more clearly specified. As such, the relation between the segments is "stronger" in Circumstance relations than in Background relations. An additional distinction is that the events are "somewhat co-temporal" in Circumstance relations, whereas in Background relations, the events "occur at distinctly different times" (Carlson and Marcu 2001: 47). Background and Circumstance in RST and Background in SDRT include only additive relations, but while temporal order is specified in some relations, other relations include no information about temporal order. Both (43) and (44) are for instance examples provided for SDRT's Background.

(43) [He had been on duty 15–20 minutes,] [when Lewinsky arrived saying she had some paperwork she needed to bring to the President.] (Reese et al. 2007: 9)

(44) [Also, about 585 workers were laid off at a stamping plant near Detroit.] [That plant normally employs 2,800 hourly workers.] (Reese et al. 2007: 9)

The relation in (43) is specified for temporal order: Lewinsky's arrival is presented to coincide with the moment that the guard had been on duty for 15–20 min. Temporal order is not, however, specified in the relation in (44): the only relation between the segments in this relation is one of addition. Both BACKGROUND and CIRCUMSTANCE in RST also include additive relations in which temporal order is specified, and additive relations for which temporal order is not relevant. In the mapping, the value for temporality is therefore underspecified for these relations. Assigning either the value "temporal" or the value "non-temporal" to relations in these classes is therefore not a straightforward matter.

Temporal order, which is part of the temporality dimension, can take the values chronological (when the events in S_1 precedes the event in S_2), anti-chronological (when the event in S_2 precedes the event in S_1), or synchronous. To permit easy reading, we will restrict this discussion to chronological and anti-chronological relations, and not address synchronous relations. Examples (45) and (46) are positive additive relations with chronological temporal order, since P in S_1 precedes Q in S_2. (47) and (48), on the other hand, are examples of positive additive relations with anti-chronological temporal order, since Q in S_1 follows P in S_2.

(45) [I visited my grandmother]$_{S1}$ before [I went shopping.]$_{S2}$

(46) After [I visited my grandmother,]$_{S1}$ [I went shopping.]$_{S2}$

(47) Before [I went shopping,]$_{S1}$ [I visited my grandmother.]$_{S2}$

(48) [I went shopping]$_{S1}$ after [I visited my grandmother.]$_{S2}$

SDRT's implementation of temporal order is compatible with the one used in CCR. Examples (45) and (46) would for instance be labeled NARRATION by SDRT and (47) and (48) would be labeled PRECONDITION. The mapping of the CCR dimension order onto RST and PDTB relation labels is less straightforward than for SDRT. RST and PDTB both distinguish between similar relations with opposite orders in their relation inventory (for example, RST's TEMPORAL-BEFORE versus TEMPORAL-AFTER, and PDTB's PRECEDENCE versus SUCCESSION), but their notions of order differ from order as employed by CCR. Rather than focusing on the event order, they look at the nuclearity of the segments (RST) and the position of the connective (PDTB).

3.6 Distinguishing various types of order

Before we proceed to specifically discussing RST's and PDTB's notions of order, we take one step back and consider the examples of the temporal relations shown in (45)–(48). First, we need to ask whether we want to distinguish these four relations in a discourse relation framework.

Section 2.2 mentioned several studies that underscored the cognitive plausibility of the dimensions implication order and of temporality. However, there are good motivations for distinguishing also those relations that are in the same chronological order but differ in terms of what is focussed on in the text (distinguishing (45) from (47), and by analogy, (46) from (48)), because differences in focus affect discourse expectations: a text containing example (45) will more likely continue with an event at grandmother's house rather than at a store. For example, a subsequent sentence like "I got some tea" would be interpreted as the subject receiving tea from the grandmother. On the other hand, in example (46), it is more likely that the story has moved to the shopping frame, i.e., a subsequent utterance "I got some tea" means that the subject bought tea while shopping. Given these different discourse expectations elicited by these discourse relations, it seems reasonable to distinguish between these relations (and the same argument holds for (47) and (48)).

Having established why the four discourse relations (45)–(48) should be distinguished, how can we describe their differences? We could make four different categories, but this would obscure the similarities between the pairs of relations. So instead, we can describe the distinctions using the notion of dimensions. There are three different ways in which these four relations can be grouped:

(45), (46) vs. (47), (48)
(45), (47) vs. (46), (48)
(45), (48) vs. (46), (47)

The first grouping option, (45) and (46) vs. (47) and (48), corresponds to the grouping based on the temporal order dimension, which we already discussed in Section 3.5. It corresponds to the mapping of P and Q onto S_1 and S_2, and would therefore also correspond to the implication order in the case of causal relations. In other words, this grouping option describes whether the order of mention in the text is the same as the real temporal order/implication order of the segments.

The second possibility is to group (45) and (47) vs. (46) and (48). This represents cases where the focus is on visiting the grandmother, and where the shopping is presented as the additional event. This second grouping describes the relations based on which segment is most important, i.e., about

which the immediately following discourse is likely to be. This is the *nucleus*, if we follow RST terminology (see a more detailed explanation of this notion in Section 3.6.1). We see this as the RST-instantiation of a dimension capturing the aspect of *relative importance*: "nuclearity" in RST.

The third and final option is to group (45) and (48) vs. (46) and (47). In these cases, the nucleus of the relation occurs first in (45) and (48), while it occurs last in (46) and (47). We see this as the PDTB-instantiation of relative importance, and we will refer to this aspect as the Arg1-Arg2 order of the segments, in agreement with the terminology used in PDTB.

Note that each of the relations in (45)–(48) can be uniquely described by just using any two out of the three possible order notions. That is, if we know any two of the orders, we can uniquely identify which kind of relation is meant, and hence automatically infer what the value for the third type of order would be. Hence, we do not necessarily have to represent all three orders, but can do with just two of them.

As discussed in Section 3.5, CCR and SDRT use the first notion of order. We will see in Section 3.6.1, respectively, Section 3.6.2 that RST and PDTB encode both the second and the third order notion. The second type of order is expressed in the discourse relation labels chosen in these frameworks (e.g., distinguishing TEMPORAL-BEFORE from TEMPORAL-AFTER relations), while the second and third order notion is captured by observing the order in which the arguments labeled as nucleus and satellite in RST, or as Arg1 and Arg2 in PDTB occur. Considering this state of affairs, it is clear that frameworks have expressed the first and second notions of order in their relation labels, but not the third type of order. We will hence do the same for our mapping, distinguishing temporal and implication order (first notion) from relative importance (second notion).

Remember that each type of order can be inferred if the other two are known; we will exploit this property to estimate a value for temporal or implication order if annotations do not express this notion, by referring to the textual order of the segments. Finally, we would like to note that the notions described here do not only hold for temporals. They apply just as much to causals, conditionals, concessives, and certain additive relations.

3.6.1 Relative importance of the segments: Nucleus-satellite-order in RST

RST defines its relations in terms of the roles of the nucleus and the satellite. The nuclei are the parts of a text that are most important and salient, whereas the satellites provide less essential or background information. The nucleus-satellite

distinction to a large extent corresponds to the grammatical arrangement of hypotaxis, but given that Carlson and Marcu (2001) only instruct assigning nucleus and satellite status on the basis of the perceived "importance" of the segments, this is not guaranteed to be a one-to-one relation.

In terms of relation labels, RST would assign (45) and (47) the label TEMPORAL-BEFORE, since the event in the nucleus (visiting grandma) occurs before the event in the satellite (shopping). (46) and (48) are both instances of TEMPORAL-AFTER, since the event in the satellite (visiting grandma) occurs before the event in the nucleus (shopping). RST hence encodes the order notion which we refer to as relative importance, and therefore, many relational labels cannot be directly mapped to temporal CCR order.

As discussed above, CCR's temporal order can in most cases be assigned relatively easily by jointly considering the relation label and the order of nucleus and satellite. For example, TEMPORAL-BEFORE relations with a nucleus-satellite order are equivalent to chronological CCR relations, whereas TEMPORAL-AFTER relations with a satellite-nucleus order are the counterparts of anti-chronological CCR relations.

For some relations, a distinction between a nucleus and a satellite within the relation does not seem appropriate. In these cases, the second and third type of order are not well defined. In RST, such relations are annotated as multi-nuclear relations: both segments of a relation are judged equally important. Two examples of multinuclear relations are SEQUENCE, in which the order of the segments follows the order of events in the real world, as in (49), and INVERTED-SEQUENCE, in which the event in the first segment takes place after the event in the second segment, as in (50).

(49) [I visited my grandmother.] Then [I went shopping.]

(50) [I went shopping.] [I had just visited my grandmother.]

In multinuclear temporal relations, the RST labels describe the first notion of order, and hence resembles the order described by CCR and SDRT. The SEQUENCE label can be directly mapped onto CCR's chronological order, and INVERTED-SEQUENCE to CCR's anti-chronological order.

The relationship between CCR and RST is analogous for causal relations, which CCR describes in terms of implication order. This is illustrated using examples (40)-(42) discussed earlier. RST would label (40) RESULT, since the "situation presented in the satellite is the cause of the situation presented in the nucleus" (Carlson and Marcu 2001: 66). Examples (41) and (42), however, would be considered REASON relations, because in both relations, the nucleus is "an

action carried out by an animate agent" and the "[s]atellite is the reason for the nucleus" (p.65). Again, it is possible, for explicit RST relations, to establish CCR's implication order relatively easily by jointly considering the relation label and the ordering of nucleus and satellite in the relation. For example, mononuclear relations marked with *because* in the first segment could be mapped as causal relations with basic order, while *because*-relations with the connective attaching to the second segment would have non-basic order. An overview of the resulting mappings is provided in Table 1 and Appendix C.

Table 1: Mapping of RST relations to CCR's implication order and temporal order.

	Example	Implication order	Temporal order	RST	N-S order
(40)	P so Q.	Basic	Chronological	RESULT	Sat-Nuc
(41)	Because P, Q.	Basic	Chronological	REASON	Sat-Nuc
(42)	Q because P.	Non-basic	Anti-chronological	REASON	Nuc-Sat
(45)	P before Q.	N.A.	Chronological	TEMPORAL-BEFORE	Nuc-Sat
(46)	After P, Q.	N.A.	Chronological	TEMPORAL-AFTER	Sat-Nuc
(47)	Before Q, P.	N.A.	Anti-chronological	TEMPORAL-BEFORE	Sat-Nuc
(48)	Q after P.	N.A.	Anti-chronological	TEMPORAL-AFTER	Nuc-Sat
(49)	P. Then Q.	N.A.	Chronological	SEQUENCE	Nuc-Nuc
(50)	Q. I had just P.	N.A.	Anti-chronological	INVERTED-SEQUENCE	Nuc-Nuc

3.6.2 Relative importance of the segments: Arg1-Arg 2-order in PDTB

PDTB uses a similar system as RST. Instead of making nucleus-satellite distinctions, it relies on hypotaxis. The relational argument that includes the connective is labeled as Arg2, and the other one is labeled as Arg1.

For example, in both (45) and (47) the *before*-clause is considered to be Arg2, and the visit to grandma is the core of Arg1. End labels are then assigned depending on the temporal ordering of Arg1 and Arg2. The label SUCCESSION, for instance, applies when Arg2 precedes Arg1 in real time, regardless of the order in which the segments are presented in the text, and PRECEDENCE is used "when the connective indicates that the situation in Arg1 precedes the situation in Arg2" (PDTB Research Group 2008: 28). PDTB's labels hence also express the second type of ordering, which we named "nuclearity," grouping together examples (45) and (47) and distinguishing them from relations in (46) and (48).

In relations that would be considered to be multinuclear in RST (coordination, implicit relations), PDTB-labels the first argument as Arg1 and the second one as Arg2: this is because PDTB asks annotators to add a connective for

implicit relations and add it *between* the segments, resulting in an Arg1-Arg2 order.

The value for CCR's temporal order can be determined for PDTB relations if we jointly make use of the PDTB relation label and the textual order of Arg1 and Arg2, which can easily be extracted from the PDTB corpus. For example, temporal PRECEDENCE relations with Arg1-Arg2 order can be mapped as chronological temporal relations, whereas the same relations with Arg2-Arg1 order can be as assigned an anti-chronological temporal order.

The same logic again holds for implication order, i.e. CCR's implication order can be inferred by taking into account the PDTB relation label (REASON vs. RESULT), as well as the textual order of Arg1 and Arg2, as illustrated in Table 2. This information, which we added as an additional dimension to our classification in Appendix B, can be gathered from the PDTB corpus.

Table 2: Mapping of PDTB relations to CCR's implication and temporal order.

	Example	Implication order	Temporal order	PDTB	Arg1-Arg2 order
(40)	P so Q.	Basic	Chronological	RESULT	Arg1-Arg2
(41)	Because P, Q.	Basic	Chronological	REASON	Arg2-Arg1
(42)	Q because P.	Non-basic	Anti-chronological	REASON	Arg1-Arg2
(45)	P before Q.	N.A.	Chronological	PRECEDENCE	Arg1-Arg2
(46)	After P, Q.	N.A.	Chronological	SUCCESSION	Arg2-Arg1
(47)	Before Q, P.	N.A.	Anti-chronological	PRECEDENCE	Arg2-Arg1
(48)	Q after P.	N.A.	Anti-chronological	SUCCESSION	Arg1-Arg2
(49)	P. Then Q.	N.A.	Chronological	PRECEDENCE	Arg1-Arg2
(50)	Q. I had just P.	N.A.	Anti-chronological	SUCCESSION	Arg1-Arg2

We note that mapping in this case is particularly easy for PDTB's RESULT relation, which is used when "the connective indicates that the situation described in Arg2 is the effect brought about by the situation described in Arg1" (PDTB Research Group 2008: 29). (40) is an example of a typical PDTB RESULT relation: Arg2 contains the clause preceded by *so and* is caused by Arg1. This relation has basic order in terms of CCR. Since none of the connectives listed by the PDTB as signaling RESULT relations can be put in front of the two segments (i.e. in a constellation [Connective Arg2-Arg1]), PDTB's RESULT can be directly mapped onto CCR as a basic order relation.

In PDTB 3.0, the relational hierarchy has been adapted to encode the nuclearity (or, as they refer to it, *directionality*) of the segments in the relation label (Prasad et al. in preparation; Webber et al. 2016). For example, the relation

sense hierarchy distinguishes ARG1-AS-CONDITION from ARG2-AS-CONDITION, ARG1-AS-EXCEPT from ARG2-AS-EXCEPT, and ARG1-AS-DENIER from ARG2-AS-DENIER in CONCESSION relations. These distinctions are of course not made for multinuclear relations, such as CONTRAST and CONJUNCTION.

It should be noted that a few explicitly marked relations require caution with this mapping. As the position of the connective determines which segment is labeled as Arg2, additional information is required to determine how PDTB assigns Arg1-Arg2 to fragments in which a single relation is signaled by a pair of markers, each of which attaches to one of the segments (e.g., *on the one hand ... on the other hand ..., either ... or ...*). Positive conditional relations, for instance, can be signaled by *if* or *if ... then ...*. For relations marked only by *if*, the argument following *if* is labeled as Arg2 in PDTB (e.g., (51)), whereas for relations marked by *if ... then ...*, the argument following *if* is labeled as Arg1 (e.g., (52)).[8] When determining CCR's implication order of such relations, it is important to check whether the relation is signaled by a single or double marker. For such examples, the Arg1-Arg2 order could be converted into the PDTB order the relation would have if only one of the pair of two markers would be present, i.e., (52)'s order could be converted into Arg2-Arg1 order, as in (51).

(51) If [the debts are repaid,]$_{Arg2}$ [bonds can be sold in the US.]$_{Arg1}$

(52) If [the debts are repaid,]$_{Arg1}$, then [bonds can be sold in the US.]$_{Arg2}$

3.6.3 Summary of various orders

In sum, CCR's notion of temporal order can be found in SDRT. Temporal order can be determined for RST with the help of the relation label and the order of nucleus and satellite, and for PDTB with the help of the relation label and the order of the annotated Arg1 and Arg2 labels. An example mapping with all three frameworks is given in Table 3.

In practice, mapping underspecified CCR's temporal order to the underspecified dimension of relative importance (nucleus-satellite or vice versa) should be possible by exploiting the information regarding the chosen connective. For instance, when analyzing temporal relations originally annotated in CCR, the identity of the connective, e.g., *before* and information on hypotaxis can be used to infer nuclearity of a relation.

---

8 This can be derived from example (23) in PDTB Research Group (2008: 10).

Table 3: Mapping of temporal relations in four frameworks.

Example	CCR	SDRT	RST	N-S order	PDTB	A1-A2 order
P before Q.	Chronological	NARRATION	TEMPORAL-BEFORE	Nuc-Sat	PRECEDENCE	Arg1-Arg2
Before Q, P.	Anti-chron.	PRECONDITION	TEMPORAL-BEFORE	Sat-Nuc	PRECEDENCE	Arg2-Arg1
After P, Q.	Chronological	NARRATION	TEMPORAL-AFTER	Sat-Nuc	SUCCESSION	Arg2-Arg1
Q after P.	Anti-chron.	PRECONDITION	TEMPORAL-AFTER	Nuc-Sat	SUCCESSION	Arg1-Arg2
P. Then Q.	Chronological	NARRATION	SEQUENCE	Nuc-Nuc	PRECEDENCE	Arg1-Arg2
Q. I had just P.	Anti-chron.	PRECONDITION	INVERTED SEQ.	Nuc-Nuc	SUCCESSION	Arg1-Arg2

3.7 On the use of CCR dimensions to establish a mapping between frameworks

All in all, the categorizations along the five CCR dimensions can be used as a first step toward mapping end labels from different frameworks to one another, especially if we supplement them with two dimensions from the respective frameworks: PDTB's Arg1-Arg2 distinction and RST's Nucleus-Satellite distinction. This will be illustrated using a TEMPORAL-BEFORE relation in RST such as (45), repeated as (53) for convenience.

(53)　[I visited my grandmother] before [I went shopping.]

Imagine that we want to map this RST relation to a PDTB relation label, using the CCR dimensions. This relation would be analyzed in CCR as a positive, objective relation with a chronological temporal order to which implication order does not imply. If we would only use these CCR values, (53) would map to PDTB's PRECEDENCE (positive, objective, temporal order underspecified) and SUCCESSION (same values). However, this does not mean that, in this case, the mapping has to remain coarse-grained and we end up with two possible relation labels. We will simply need to take an additional step and compare the connective that marks the relation and that also provides information about the classification along PDTB's Arg1-Arg2 dimension: PDTB PRECEDENCE relations are typically marked by the connective *before*, and PDTB SUCCESSION relations are typically marked by *after*. We therefore need to take into account the

connective that is used in order to complete the mapping: (53), with its chronological order, contains *before* and displays an Arg1-Arg2 order, which leaves us only with the option to label it as a PRECEDENCE relation (compare Table 3).

4 Additional features

In Section 3, we have shown how a classification of the end labels distinguished by PDTB, RST, and SDRT in terms of the five CCR dimensions form a good starting point if we want these frameworks to be able to "talk to one another." For SDRT, for instance, this results in the following cross-classification (see Appendix A, B and C for a complete overview of such categorizations).[9] Table 4 indicates that NARRATION can uniquely be labeled as a positive, additive, objective relation with a basic order, and that CONTRAST covers all negative relations (apart from the negative ALTERNATION relations), irrespective of their values on polarity, basic operation, and source of coherence.

Table 4: SDRT end labels categorized along the CCR dimensions.

SDRT	Polarity	Basic operation	Implication order	SoC
Narration	Positive	Additive	N.A.	Objective
Precondition	Positive	Additive	N.A.	Objective
Background	Positive	Additive	N.A.	Objective
Parallel	Positive	Additive	N.A.	Objective
Continuation	Positive	Additive	N.A.	Obj./subj.
Commentary	Positive	Additive	N.A.	Subjective
Result	Positive	Causal	Basic	Objective
Explanation	Positive	Causal	Non-basic	Objective
Consequence	Positive	Causal	Basic/non-b.	Obj./subj.
Elaboration	Positive	Add./causal	Basic/non-b.	Obj./subj.
Contrast	Negative	Add./causal	Basic/non-b.	Obj./subj.
Alternation	Pos./negative	Add./causal	Basic/non-b.	Obj./subj.

However, the CCR dimensions are not sufficient for mapping all relation labels distinguished by PDTB, RST and SDRT onto one another, as there are many instances where two or more labels would map onto the same set of values along the five dimensions. For example, Table 4 shows that within the category of positive additive objective relations to which order does not apply, SDRT

---

9 In Table 4, we have disregarded the temporal order dimension for brevity of the discussion.

distinguishes BACKGROUND relations from PARALLEL relations. And if we look at the relations that are underspecified for source of coherence, one more positive additive relation should be considered: CONTINUATION. Restricting the mapping to the distinctions made by CCR would mean that this additional distinction would get lost in the mapping. In this section, we therefore discuss additional features that allow us to capture all of the distinctions made by the other frameworks, thereby ensuring that each end label can be classified by a unique combination of features. In our selection of additional features, we have taken a bottom-up approach, meaning that we have used features that are already present in the annotation frameworks at hand.

4.1 Further distinctions within the class of additive relations

As the example of the variety of positive additive relations within SDRT shows, further distinctions can and should be made in addition to the dimensions included in CCR in order to arrive at a fine-grained mapping. In this subsection, we will discuss further distinctions within the class of additive relations, focusing on different types of specifications (Section 4.1.1), list relations (Section 4.1.2), and alternatives (Section 4.1.3).

4.1.1 Specificity

A recurring feature in the relation inventories of PDTB, RST, and SDRT is the specificity of the content of the segments in a relation. In PDTB, for instance, SPECIFICATION, GENERALIZATION, and EQUIVALENCE are all subtypes of the class RESTATEMENT. In SPECIFICATION relations, Arg2 is more specific than Arg1; in GENERALIZATION relations, Arg1 is more specific than Arg2; and in EQUIVALENCE relations, Arg1 and Arg2 are both equally specific. The specificity of the segments also plays a role in the definition of PDTB's INSTANTIATION, which applies to relations where "Arg1 evokes a set and Arg2 describes it in further detail" (PDTB Research Group 2008: 34). Similarly, it is a defining feature in SDRT's ELABORATION relations, in which the second segment provides more detail about the first segment, e.g., an example, a sub-event, or the manner in which the first segment occurred. Finally, the specificity of the segments features in several RST relations, for instance EXAMPLE, in which the satellite provides an example for the nucleus, DEFINITION, in which the satellite provides a definition of the nucleus, and the other ELABORATION relations, except for ELABORATION-ADDITIONAL.

Although the specificity of the segments can be found as a feature in the relation definitions of all three frameworks, the extent to which this feature is further specified by segment-specific properties differs (compare RST's Elaboration set-member vs. PDTB's Instantiation). In Appendix A–C we have annotated which relation labels involve some kind of specificity. For cases where a direct correspondence can be identified, the label indicates this correspondence (i.e., "specificity-example" for RST Example and PDTB-Instantiation relations, and "specificity-equivalent" for RST Restatement and PDTB Equivalence relations).

4.1.2 Lists

Several frameworks preserve specific end labels for additive relations that contain two or more arguments that can be listed. In both RST and PDTB 2.0, such relations are labeled as Lists. These frameworks, however, differ in their operationalization of this type of relation. RST simply requires that the elements in the relation can be listed, and should not be in a comparison, contrast or other stronger type of multinuclear relation (Carlson and Marcu 2001: 61). Characteristic of List relations in PDTB 2.0 is that the list needs to be defined in the preceding discourse (PDTB Research Group 2008: 37).[10] Example (54) illustrates this: the List relation is signaled by the reference to "elements" in the preceding sentence.

(54) But to Mitsubishi Estate, the acquisition has just the elements that should win support from both sides. [First of all, it is a friendly acquisition in which Rockefeller sought out Mitsubishi Estate and asked it to buy a majority share.] [Secondly, the two companies found a similarity in their business and development philosophies and intend to cooperate in a range of activities from real estate to telecommunications.] (wsj_0277)

---

10 Another difference between Lists in RST and PDTB is that List relations in PDTB consist of just two segments (or arguments, in PDTB terminology), just like all other PDTB relations, whereas List relations in RST can contain two or more segments. Still, PDTB acknowledges the fact that "certain types of constructions could be possibly viewed as structures with more than two arguments, such as Lists" (PDTB Research Group 2008: 1, footnote 4). In PDTB, lists of more than two elements "are composed of multiple binary structures, that is, every new successive element of the List forms one of the arguments of a 'new' relation, with all prior elements together forming the second argument" (PDTB Research Group 2008: 1, footnote 4).

Unlike RST and PDTB 2.0, SDRT and PDTB 3.0 do not contain a separate end label for list-like relations. In SDRT, list relations such as (54) can be found under the label ELABORATION, which also includes non-list relations such as (55) (referred to as INSTANTIATION by PDTB, and as EXAMPLE by RST).

(55) [Experts say such long hours for attendants pose a safety risk.] [For instance, tired flight attendants might not react quickly enough during an emergency evacuation.] (wsj_0730)

The SDRT end label CONTINUATION also contains list relations; in a series of three segments this label applies to the relation between S_2 and S_3 when these segments both relate in the same way to S_1. For example, in (56), according to Reese et al. (2007: 17) S_2 and S_3 are related to S_1 via BACKGROUND and to each other via CONTINUATION.

(56) [American officials felt talks had reached a point where mediation would be helpful.]$_{S1}$ [Negotiations with the pilots have been going on for 11 months;]$_{S2}$ [talks with flight attendants began six months ago.]$_{S3}$ (Reese et al. 2007: 17)

In PDTB 3.0, positive additive list relations are clustered with non-list relations as CONJUNCTIONS. In other words, introduction of a new feature that sets apart list relations from non-list relations, implies that ELABORATION, CONTINUATION, and CONJUNCTION will be underspecified for this List feature. In order to be able to identify the list relations in SDRT and PDTB 3.0, ELABORATION and CONJUNCTION relations will have to be checked to determine whether they contain a marker that signals a list relation (e.g., *also, firstly/secondly, moreover*).

4.1.3 Alternatives

Another important feature in the relation inventory of all three frameworks seems to be ALTERNATIVE. This feature helps distinguish additive relations in which the two segments are presented as alternatives – DISJUNCTION relations as in (57) – from additive relations in which this is not the case.

(57) [Call it a fad.] [Or call it the wave of the future.] (wsj_0623)

Using the ALTERNATIVE feature, we are able to separate CONJUNCTIVE, DISJUNCTIVE, CHOSEN ALTERNATIVE in PDTB and DISJUNCTION in RST from the

other additive relations with which they were clustered. The ALTERNATIVE feature also applies to SDRT relations with the label ALTERNATION that are not negative conditional relations.

PDTB 2.0 includes several relations that can be described with the ALTERNATIVE feature. Unlike RST and SDRT, it distinguishes between inclusive and exclusive disjunctions: CONJUNCTIVES as in (58), and DISJUNCTIVES like (59) and CHOSEN ALTERNATIVES such as (60). In CHOSEN ALTERNATIVE relations, it is specified which of the two alternatives holds (i.e., has been "chosen"), and which one does not.

(58) [Today's Fidelity ad goes a step further, encouraging investors to stay in the market] [or even to plunge in with Fidelity.] (wsj_2201)

(59) [Those looking for real-estate bargains in distressed metropolitan areas should lock in leases] [or buy now.] (wsj_2444)

(60) [Under current rules, (…) it isn't allowed to share in the continuing proceeds when the reruns are sold to local stations.] Instead, [ABC will have to sell off the rights for a one-time fee.] (wsj_2451)

Although an additional feature of ALTERNATIVE is needed to set these three PDTB relations apart from other types of additive relations, no additional feature is needed to cover the distinction between inclusive and exclusive disjunctions, because this distinction is already captured by the polarity dimension. In inclusive disjunctions, the two segments are presented as alternatives, but can hold at the same time. Inclusive disjunctions are therefore positive additive relations: P & Q. In exclusive disjunctions, on the other hand, the two segments cannot hold at the same time: either P is true, or Q. Since exclusive disjunctions always involve a negative counterpart of P or Q (i.e., P & not-Q, or not-P & Q), they are negative additive relations. While PDTB makes a distinction between inclusive and exclusive disjunctions, RST and SDRT do not. This is why the additive alternative relations in RST and SDRT have been marked as underspecified for polarity.

Determining the polarity of RST's DISJUNCTION and SDRT's ALTERNATION on the basis of linguistic markers will be rather complicated, since both inclusive and exclusive disjunctions are commonly signaled by *or*. Although there are several clues that can be used to identify whether a relation is an example of inclusive or exclusive disjunction (e.g., *either … or* marks exclusive disjunctions; *or even* inclusive disjunctions), additional manual annotation will often be required. Separating disjunctions from negative conditional relations in SDRT's

ALTERNATION, on the other hand, would probably be easy to do, since negative conditionals are often marked by *unless*.

4.2 Further distinctions within the class of causal relations

In Section 4.1 we have illustrated how further distinction within the class of additive relations can help out in working toward a one-to-one mapping between end labels from different frameworks. For the class of causal relations, such additional features are necessary as well, since the CCR combination positive causal applies to a variety of labels from other frameworks. Even if conditional relations are set apart from non-conditional causal relations, more fine-grained distinctions are necessary to arrive at, for example, correct PDTB classifications. In this subsection, we therefore discuss further distinctions within the class of causal relations, focusing on different types of conditionals (Section 4.2.1), and causal relations (Section 4.2.2).

4.2.1 Different types of conditionals

If we want to map a conditional relation in SDRT (e.g., a CONSEQUENCE relation) to a conditional relation in PDTB, mapping via the CCR dimensions does not suffice. CONSEQUENCE is a positive causal-conditional relation that is under-specified for order and source of coherence. In PDTB 2.0, this combination of CCR values would result in a list of six types of objective conditional relations (comprising PDTB's class of CONDITION), and two more conditional relations with a subjective value (forming PDTB's class of PRAGMATIC CONDITION).[11] A similar issue arises with RST: there are several conditional relations that have the same basic operation (CONDITION, HYPOTHETICAL, CONTINGENCY, and OTHERWISE), but cannot (easily) be distinguished on the basis of their values on the other CCR dimensions. This is mainly because they are underspecified on one (CONTINGENCY) or two dimensions (CONDITION). Again, additional features are needed to facilitate the mapping from conditional relations in one frame-work to another. We refer to Appendix D and the respective annotation manuals for suggestions.

---

11 Note that this translation problem is reduced in PDTB 3.0: The new scheme proposes the types CONDITION and CONDITION + *Speech act*, and their negative counterparts.

4.2.2 Goal-oriented relations

Even after the different types of conditionals are set apart, a relatively large number of positive causal RST labels remain. And although source of coherence helps setting subjective EVIDENCE and CONCLUSION relations apart from clearly objective cases such as REASON and all relation labels in the class of CAUSE, the combination of the five dimensions and additional features discussed so far does not suffice to distinguish these relations from the positive causal RST relations PURPOSE, ENABLEMENT, MEANS and PROBLEM-SOLUTION. The latter four relation types can all be characterized as goal-oriented, that is, one of their segments concerns an intentional, goal-directed action by an agent. These types of relations are frequently marked by *in order to* or *so that*. For example, (61) illustrates a relation in which an intentional action – hiring Prudential-Bache Securities Inc. – is undertaken to achieve the intended goal of restructuring the NBI and improving its balance sheet.

(61)   [NBI also said] [it has hired Prudential-Bache Securities Inc. as its financial adviser and investment banker] [to help it restructure financially and improve its balance sheet.] (wsj_0647)

In RST, relations such as (61) receive the label PURPOSE. Although neither SDRT nor PDTB has a designated label for PURPOSE relations, both approaches mention the relation type in their manuals as a subtype of another type of causal relation. SDRT considers *because the subject wanted to* an appropriate paraphrase of RST PURPOSE relations (cf. Reese et al. 2007: 12–13). This implies that (62), with the inter-clausal *because*, is considered an appropriate non-basic paraphrase of (61), which according to the SDRT guidelines should be annotated as an EXPLANATION relation.

(62)   NBI has hired PBSI (...) <u>because it wanted</u> PBSI <u>to</u> help it restructure financially...

In PDTB, RST PURPOSE relations are considered to be a subtype of RESULT relations, in which Arg2 is the result of Arg1. For example, in (63), the *so that* clause, Arg2, "expresses the situation that is expected to hold as the result of Arg1," the preceding clause, and "that the situation specified in Arg2 may or may not hold true at a subsequent time, even if Arg1 does" (PDTB Research Group 2008: 39). RST agrees when it says that "the situation presented in the satellite of a purpose relation is only putative, i.e., it is yet to be achieved" (Carlson and Marcu 2001: 64).

In cases such as (64), in which the nucleus instead of the satellite is unrealized, RST uses the label Enablement. Here, the goal-oriented "action presented in the satellite increases the chances of the situation in the nucleus being realized" (Carlson and Marcu 2001: 57).

(63) Northeast said it would refile its request and still hopes for [an expedited review by the FERC] [so that it could complete the purchase by next summer (...).] (wsj_0013)

(64) [The administration of federal credit should closely parallel private lending practices, including the development of a loan loss reserve and regular outside audits.] [Establishing these practices would permit earlier identification of emerging financial crises, provide better information for loan sales and budgeting decisions, and reduce fraud.] (wsj_1131)

Goal-orientation is also characteristic of RST's Means relations such as (65), in which the "satellite specifies a method, mechanism, instrument, channel or conduit for accomplishing some goal," telling "how something was or is to be accomplished" (Carlson and Marcu 2001: 62). In SDRT, this relation belongs to the category of Explanations.

(65) [They were able to raise the necessary funds] [by cashing in all their stock options.] (Reese et al. 2007: 13)

In RST Purpose relations, the goal segment describes a situation that is or has been aimed for. This segment can be rephrased as a negatively evaluated situation that is to be avoided (Hoey 1983; Sanders et al. 1993), in which case we are dealing with RST's Problem-Solution relation. This is exemplified in (66), a negative counterpart of (61).

(66) [NBI had an unfavorable balance sheet.] [That's why it hired Prudential-Bache Securities Inc. as its financial adviser and investment banker.]

Interestingly, these relations are often expressed in the form of a conditional: *If you have an unfavorable balance sheet, hire us!* In this form, PDTB considers them Implicit Assertions.

To sum up, a feature that indicates goal-orientedness could be used to set apart RST's Purpose, Enablement, Means, and Problem-Solution relations from other types of positive causal relations. Further segment-specific features

are needed to make an even more fine-grained distinction within this cluster of goal-oriented relations.

4.3 Other distinctions

In the previous sections, we have discussed additional features that were used by more than one framework to make further specifications within their relation inventory. Although these features, in addition to the dimensions, can account for most of the relation labels of RST, SDRT, and PDTB, and thus facilitate the mapping of relations between these frameworks, they are not sufficient to provide all relations with a unique set of features. Several more fine-grained distinctions are used across the frameworks to distinguish between relations. Most of these distinctions – such as PROPORTION and PREFERENCE in RST, OPPOSITION in PDBT, and COMMENTARY in SDRT – are unique to one framework and are formulated in terms of the propositional content of the segments. We refer to Appendix D and the respective annotation manuals for definitions of these features that help distinguish between relations that still cluster together after application of the dimensions and the features addressed in Sections 4.1 and 4.2.

5 How well do actual annotations map onto one another?

In order to assess the correspondence between frameworks, we have compared existing annotations for PDTB and RST. Both frameworks have been applied to Wall Street Journal texts, and annotations overlap for 385 newspaper articles in sections 6, 11, 13, 19, and 23. Since these annotations are on the same texts, they represent an ideal ground of comparison. While the theoretical mappings between PDTB, RST and SDRT relations in terms of CCR dimensions and other features were determined on the basis of definitions and examples in the annotation guidelines (see Sections 3 and 4), the mapping study presented here aims at showing how well annotations map onto one another in practice. A full report on this "mapping in practice" with more details on the exact methodology and more in-depth discussion of the results can be found in Demberg et al. (2017); here we focus on the findings most relevant to our study.

5.1 Alignment procedure

A first challenge in arriving at a mapping lies in determining which relation labels should be compared to one another: PDTB and RST differ in their ways of selecting the arguments of a coherence relation and describing the coherence structure of the entire text. Overall, we find many more RST relations than PDTB relations in the same text, because RST builds a fully connected tree per paragraph, that is, there are typically many high-level coherence relations combining different parts of a paragraph, also in the absence of connectives, while such longer, high-level relations in PDTB can only occur if they are marked with a connective. Furthermore, RST annotations are sometimes more fine-grained – between smaller units – because PDTB 2.0 does not annotate implicit relations within a sentence.

Since PDTB has fewer relations overall, our alignment procedure focused on finding the best RST relation equivalent for each PDTB relation. The alignment procedure is an optimization process that identifies the best corresponding RST Elementary Discourse Unit for each PDTB relation argument. We only selected those relations for which the mapping observes the strong nuclearity hypothesis of RST-DT (see Marcu 2000), which states that when an RST relation is postulated to hold between two spans of text, it should also hold between the nuclei of these two spans. We furthermore excluded PDTB's NoRel and EntRel relations, as well as RST's Same-Unit and Attribution relations. For relations annotated with more than one label in PDTB, we selected the PDTB relation label that most closely corresponded to the RST label. In cases where PDTB did not distinguish specific end labels but only annotated sub-class labels (e.g., not Opposition or Juxtaposition, but Contrast), we included the latter labels in our mapping.

In total, we were able to successfully map 76.6% of PDTB relations from the jointly annotated part of the corpus to corresponding RST labels, a total of 4,559 relations. These relations adhered to the strong nuclearity hypothesis, but the corresponding arguments of these relations are not necessarily of the same length due to differences in segmentation. The remaining 23.4% of relations (those that were not successfully mapped) include more complicated examples of discontinuous RST spans, or annotations where – due to violations of the strong nuclearity principle – we cannot be sure that PDTB and RST annotators identified corresponding arguments in their analysis.

5.2 Mapping through the lens of CCR

For three CCR dimensions – polarity, basic operation and source of coherence – as well as one additional feature (list), we will now present how well the PDTB

and RST annotations coincide. An analysis in terms of CCR's implication and temporal order is complicated by the fact that neither PDTB nor RST directly annotates for order; therefore, we have disregarded this dimension here. For ease of discussion, we have classified temporal relations as a subtype of additive relations (comparable to the approach in the original CCR proposal in Sanders et al. 1992), instead of conducting a separate analysis on the temporality dimension. We also attempted to incorporate the feature *alternative* in our analyses, but the number of annotated ALTERNATIVE relations in RST was too small to draw any reliable conclusions.

5.2.1 Polarity

Table 5 displays the percentage of agreement in the mapping of PDTB and RST relations in terms of the CCR dimension polarity. In 78.5% of all mapped coherence relations, the PDTB and RST annotations agreed on polarity, and for a further 12% of instances the dimension polarity was marked as underspecified according to one of the frameworks. This means that more than 90% of relation annotations were consistent with each other in terms of their value on polarity.

Table 5: Percentage of agreement (and number of instances) in PDTB-RST mapping in terms of polarity.

				RST
		Positive	**Negative**	**Underspecified**
PDTB	Positive	**64.4 (2,916)**	1.6 (71)	10.5 (475)
	Negative	8.2 (372)	**14.0 (636)**	1.1 (50)
	Underspecified	0.1 (4)	0.0 (2)	0.1 (3)

Note. Bold numbers indicate agreement between frameworks.

The disagreements on polarity included 8.2% of cases that were annotated as a positive relation in RST, while the accompanying PDTB label was negative. These were mostly cases where RST annotated ELABORATION-ADDITIONAL, COMPARISON, LIST, EXPLANATION-ARGUMENTATIVE or INTERPRETATION, while PDTB annotated a subtype of COMPARISON, most often CONTRAST. Taking a closer look at these cases, we found that these were to a very large extent – two third of the cases – implicit relations. The remaining third of relations with such an inconsistent annotation display *but, while* and *however* as the most frequent connectives, markers that clearly hint at a negative relation.

Another 1.6% of the instances were annotated as positive relations in PDTB but as negative ones in RST. Most of these cases were explicit relations annotated as CONJUNCTIONS in PDTB, while RST annotated a CONTRAST or ANTITHESIS relation. These relations were most often marked with *also* or *and*. Additionally, some of the explicit relations were annotated as TEMPORAL in PDTB but as CONCESSION in RST (marked with *if, after* or *while*). This indicates that some annotation mismatches likely occurred due to differences in the interpretation of these connectives.

5.2.2 Basic operation

Within the dimension basic operation, CCR distinguishes three possible values: additive, causal, and conditional. We classified temporality as a fourth subtype for this analysis. As Table 6 shows, the agreement on exact basic operation values was 62%, and an additional 16%, mostly RST relations, had the label *underspecified*.

Table 6: Percentage of agreement (and number of instances) in PDTB-RST mapping in terms of basic operation.

						RST
		Additive	Causal	Conditional	Temporal	Undersp.
PDTB	Additive	**43.1 (1,950)**	9.5 (428)	0.6 (29)	1.1 (49)	8.0 (361)
	Causal	7.0 (319)	**11.6 (524)**	0.1 (3)	0.1 (4)	3.2 (145)
	Conditional	0.1 (5)	0.2 (7)	**3.1 (140)**	0.0 (1)	0.5 (23)
	Temporal	1.8 (82)	1.2 (53)	0.5 (24)	**4.3 (197)**	4.0 (182)
	Underspecified	0.0 (1)	0.0 (1)	–	–	0.0 (1)

Note. Bold numbers indicate agreement between frameworks.

Most of the remaining 22% of disagreements on the basic operation can be attributed to the distinction between additive and causal relations, which is particularly often confused in the case of implicit relations. For causal relations, we can observe that the annotators of the two frameworks almost always agreed on the basic operation of explicitly marked causal relations, while there is a lot of disagreement on implicit causals, many of which are simply annotated as ELABORATION-ADDITIONAL in RST. We also see that a substantial number of explicit causals in PDTB are annotated as RST CIRCUMSTANCE, which according to our mapping based on the annotation guidelines should be an additive or

temporal relation. Closer inspection of these cases reveals that they are mostly marked by the connective *as*, which is ambiguous between a temporal and a causal interpretation. Similarly, we find that many PDTB CONDITIONALS are also annotated as RST CIRCUMSTANCE, which again is unexpected in terms of causal vs. additive basic operation.

Furthermore, PDTB has a stronger tendency to annotate coherence relations as temporal than RST. RST only very rarely labels coherence relations that are not marked with an explicit connective as temporal. Many implicit PDTB temporals are annotated as additive or causal in RST.

Our mapping, based on the definitions and annotation guidelines, suggested that both RST and PDTB distinguish negative additives (PDTB CONTRAST and RST CONTRAST) from negative causals (PDTB CONCESSION and RST CONCESSION), but in the actual mapping of the annotations we find that these distinctions do not map well onto one another in all cases. As shown in Table 7, the mapping of negative relations is successful for some subtypes: as expected, negative causal PDTB EXPECTATION relations, which belong to the sub-class CONCESSION, are usually annotated in RST as negative causal CONCESSION relations, and PDTB's negative additive OPPOSITION and JUXTAPOSITION relations are usually annotated in RST as negative additive CONTRAST relations. However, PDTB's additive CONTRAST and causal CONTRA-EXPECTATION relations are almost equally distributed among the three negative RST classes (i.e., ANTITHESIS, CONTRAST and CONCESSION), which indicates that the annotators of the respective framework often disagreed on the basic operation of negative relations. It should, however, be noted that the distinction between contrast and concession is known to be notoriously difficult and has often led to low inter-annotator agreement even within a single framework. In addition, PDTB JUXTAPOSITION relations are quite often annotated as RST

Table 7: Number of instances per combination of negative relations in PDTB and RST.

					RST
		ANTITHESIS	CONTRAST	CONCESSION	COMPARISON
PDTB	CONTRA-EXPECTATION	**19**	23	**21**	2
	EXPECTATION	**18**	3	**31**	0
	CONTRAST	**154**	**126**	91	16
	JUXTAPOSITION	**22**	**49**	9	30
	OPPOSITION	**10**	**24**	3	6

Notes. Only labels where $n > 10$ are included. Bold numbers indicate agreement between frameworks on polarity and basic operation of these labels.

COMPARISON relations, which according to the guidelines are specifically defined as not being contrastive. In other words, negative additive relations in PDTB receive a positive additive label in RST here. The mapping hence indicates that a second look at these examples might be instructive to assess how these relations should best be characterized.

5.2.3 Source of coherence

Source of coherence is underspecified in one or both frameworks in 80% of the annotated coherence relations (see Table 8). Among the relations that have a specified source of coherence according to both frameworks, there is 93% agreement on the objective relations. This also means that there are overall very few relations that were labeled as being subjective. Because of this high number of underspecifications and low number of subjective cases, source of coherence cannot be mapped well between this pair of frameworks, as we already predicted on the basis of the definitions in the annotation guidelines.

Table 8: Percentage of agreement (and number of instances) in PDTB-RST mapping in terms of source of coherence.

		RST		
		Objective	**Subjective**	**Underspecified**
PDTB	Objective	**17.5 (792)**	1.0 (44)	9.7 (440)
	Subjective	0.3 (13)	**0.1 (3)**	0.8 (36)
	Underspecified	15.8 (715)	5.7 (258)	49.2 (2,228)

Note. Bold numbers indicate agreement between frameworks.

5.2.4 List

A closer look at the mapping of LIST relations shows an effect that we already predicted on the basis of the annotation guidelines: PDTB's definition for the LIST relation is much stricter than RST's definition. LIST relations in the PDTB, with their narrow definition of lists having to be defined in the previous context, are most often annotated as RST LIST: 71%, with most of the remaining instances being annotated as ELABORATION-ADDITIONAL. By contrast, RST's LIST relations have a much wider distribution of PDTB relations – only 12% of RST LISTS are also PDTB LISTS. These cases are mostly annotated as CONJUNCTION or CONTRAST relations in PDTB.

5.3 Conclusion on the PDTB-RST mapping

The mapping of existing RST and PDTB annotations brought up a number of important points regarding our mapping based on the definitions and annotation guidelines. Although the mapping seems relatively successful in terms of the CCR dimensions polarity and basic operation, the overall number of predicted correspondences is certainly lower than what one would have hoped. Most of these mismatches are not related to the quality of the mapping proposal put forward in this article, but seem to be due to different operationalizations within the annotation frameworks (e.g., of LIST relations), or to different interpretations by the annotators. This holds particularly true for implicit relations, relations marked with ambiguous connectives such as *while, when* or *as*, and negative relations. Therefore, an added value of this "mapping in practice" is that it raises awareness of the types of annotation differences that occur if different frameworks are applied to the same set of data – any mapping can only be as successful as the actual correspondence in the data.

Still, the mapping in practice indicates a few areas that might need revision, or at least additional analyses for our suggested mappings. To be more specific, we found that RST CIRCUMSTANCE relations, although not causal according to the RST manual, are often mapped onto PDTB causals. Similarly, a decomposition of RST COMPARISON relations leads to them being mapped onto PDTB JUXTAPOSITION relations, even though these relations – according to the definitions in the respective annotation manuals – do not match in terms of polarity. Also, the mapping in terms of source of coherence was not very informative, because of the high number of relations that are underspecified for this dimension. Finally, the starting point of this mapping in practice, the alignment of RST and PDTB relations, raised interesting theoretical questions regarding the definition of what counts as a coherence relation in different frameworks.

6 Discussion

In order to make further progress in the study of language use at the discourse level, it is imperative to make optimal use of existing discourse-annotated corpora. So far, very different annotation schemes have been used, blocking the exchange of data from various corpora. Every framework uses different relation names, and the underlying definitions vary. In this paper, we have investigated how three often used and seemingly different discourse annotation frameworks can be related. We have shown how relation labels from one

framework can be "translated" into annotation tags from other frameworks based on a set of dimensions, which can be used as an interface that will allow researchers to find identical or at least closely related relations within a set of annotated corpora. This interface is open for extension to other frameworks.

The unifying dimensions taken from the Cognitive approach to Coherence Relations proved to be a useful tool for this interface. The different frameworks we have considered here also make distinctions that are different or more fine-grained than the ones made by the original CCR framework. In order to be able to represent any distinctions which are more detailed than the ones used in the original CCR framework, we have extended this initial list of dimensions with additional features. Representing all distinctions made by the frameworks guarantees the most specific mapping possible between all schemes. While there is of course no mapping scheme that can overcome the differences in granularity that exist between frameworks (i.e., a mapping scheme cannot add information that is not present in the original annotated resources), the mapping approach suggested here makes explicit to users which aspects of a coherence relation are conflated in some annotation schemes, and which additional distinctions are made. For instance, PDTB displays a high-level of granularity by distinguishing six types of CONDITIONALS; at the other end of the granularity spectrum, SDRT's CONTRAST relation covers virtually all negative relations.

The existence of such "umbrella labels" in PDTB, RST and SDRT, in which certain distinctions are not made at all, or are not made systematically within a class of relations, causes underspecification along the CCR dimensions during mapping. As a result, using the interface to translate a relation label from one framework to another does not always lead to one, but sometimes to several candidate relation labels in another framework. However, the number of labels a researcher has to choose from in order to reach a final mapping is very restricted, given the fact that most relation labels are specified for the other dimensions. For some cases – especially the ones related to order and directionality – we have presented suggestions for resolving those mappings to the correct classes, which in some cases can even be done automatically by relying on signals in the text or the annotation (e.g., order of Arg1 and Arg2 in PDTB).

The results of the mapping test in Section 5 uncovered various differences in RST's and PDTB's annotation of the same set of newspaper fragments. Interpreting these results, we were able to show the benefit of using an all-to-decomposing-features mapping over an all-to-smallest-common mapping and a one-to-one mapping. The existing data do not allow for an all-to-smallest-common mapping because there simply is no smallest-common distinction that the two schemes systematically agree on across all of the instances.

We also believe that our approach has considerable advantages over generating pair-wise one-to-one mappings. One reason is that a new mapping would have to be generated for each pair of coherence relation schemes. Furthermore, we find that we gain more insight from the feature-based mapping. A one-to-one mapping on the basis of the RST and PDTB manuals would generate an overview in which each RST relation label is related to a PDTB equivalent. Comparing this theoretical one-to-one mapping to the actual annotations of the Wall Street Journal fragments would result in a percentage of correctly predicted matches and a percentage of mismatches between the PDTB and the RST labels, but these results would be very hard to interpret. Apart from looking at the PDTB and RST classes that the various relation labels belong to, it would be difficult to pinpoint the exact sources of mismatches. By contrast, breaking each relation label down in terms of dimensions and features, as in the all-to-decomposing-features approach, allows for a meaningful interpretation of various types of mismatches, as we have shown in Section 5.

The close comparison of the three widely used annotation schemes and the "mapping in practice" we presented in Section 5 have resulted in important observations of a more general, methodological and theoretical interest. Below, we will discuss these issues in turn.

6.1 Methodological issues

First, we have come across important differences in segmentation: frameworks differ both in what they consider a minimal discourse unit and in what constitutes a maximum unit. This became especially apparent during the application of the interface to the corpus of newspaper texts that are annotated by both PDTB and RST: about 25% of the PDTB relations could not be mapped onto RST equivalents because of large differences in segmentation (see Section 5). The number of RST relations that cannot be mapped to PDTB-labels is likely much higher, because the two frameworks differ in the minimal units they define and in the "completeness" of the annotation.

A good example of differences in segmentation granularity are so-called ATTRIBUTION relations. In RST and SDRT, a constituent that explicitly mentions a source or spokesperson (e.g., *according to Bill, I believe, a spokesman said*) is considered a separate segment that is involved in a coherence relation of ATTRIBUTION with the remainder of the proposition, the content of what is said or believed. PDBT and CCR do not identify such constituents as separate discourse segments, and therefore do not acknowledge ATTRIBUTION as a coherence relation. In our interface, we therefore refrained from classifying such

relations in terms of the unifying dimensions. None of the frameworks seem to have an upper boundary when it comes to segment size. However, in practice, RST annotates many more relations that have large segments – since it aims to annotate the entire text – while for instance PDTB annotates the minimal units and in general ends up annotating very global relations only if they are explicitly signaled.

The observed differences in segmentation pose a serious methodological problem that needs to be addressed in the field of discourse annotation: if we do not agree about the segments between which a relation should hold, how can we agree on the relation interpretation? Various elements for this discussion are in place (see, among many others, Hoek et al. 2017a; Mann and Thompson 1988; Schilperoord and Verhagen 1998), and agreeing on segmentation would mean a large step forward to our goal of improving communication between frameworks, or a completely unified approach to discourse annotation.

Another methodological issue concerns inter-annotator agreement in discourse annotation. One of the explanations put forward for the discrepancies between PDTB and RST labels for the newspaper fragments is the frequently observed phenomenon that there is always some disagreement in discourse annotation. There are various reasons for this disagreement, and recently, various solutions have been discussed (cf. Spooren and Degand 2010; van Enschot et al. submitted).

For example, one way of obtaining higher inter-annotator agreement scores is to split this process into several smaller steps. The unifying dimensions put forward in this paper are expected to be useful as intermediate annotation steps toward choosing a specific end label from any of the frameworks. Scholman et al. (2016) have shown how a flowchart-like approach with similar distinctions worked quite well in an annotation experiment with a group of inexperienced annotators. Furthermore, in case of disagreement between annotators, the dimensions allow us to systematically investigate on which aspect of the relation they disagree, so that the source of confusion or disagreement can be identified. This should be beneficial for further agreement between annotators.

Still, there will always be cases in which it is simply hard to distinguish between two or even more readings of a fragment, as natural discourse often allows for more than one interpretation. We also know that systematic ambiguities exist, and in such instances annotators may use a strategy to stimulate similar annotations in comparable cases. For instance, the sequence *George came in; Bill started to laugh* is a simple example of a well-known ambiguity between causal and temporal readings. In such cases, one solution is that analysts use a heuristic, such as a specification bias: "whenever there is the choice, choose the most specific or informative relation label," leading to a

causal rather than a temporal relation label in this case. In concrete corpus studies, heuristic strategies like these have been successfully applied, for instance in the area of child language acquisition (cf. Evers-Vermeul 2009, 2011; Spooren and Sanders 2008; van Veen 2011). The dimensions and features listed in this paper could serve as a basis for formulating such interpretation biases, and thereby help develop better heuristics for the practice of corpus annotation. Again, however, note that it is crucial for different annotation schemes to make those heuristics explicit, as they affect annotation decisions and different heuristics being used by different frameworks can lead to less consistency if joining those resources later.

Another methodological improvement can be made by applying substitution and paraphrase tests during the annotation process (Knott and Sanders 1998; Sanders 1997). In a substitution test, the original connective is substituted by another connective that is known for signaling a certain type of relation, while the meaning of the original relation is preserved. If there is no original connective present, the proposed connective is merely mentally inserted. In a paraphrase test, the annotator is instructed to choose one of two or more given paraphrases that best suits the coherence relation expressed in the text. We noticed that the framework manuals and their discussion of annotated examples hardly propagate the use of such text-linguistic tests, even though such tests have proved their value in studies on connectives in language use in various languages and across genres and media (see among others, Degand and Maat 2003; Knott and Dale 1994; Li et al. 2013; Pit 2007; Spooren and Sanders 2008; Stukker et al. 2008; Zufferey 2012). A notable exception is PDTB's process for annotating implicit relations, where annotators are instructed to insert a connective that, according to them, accurately captures the relation. Because annotators are completely free in their choice of connective, this latter approach might generate a bias. Paraphrase and substitution tests in which the annotator has to select a connective from a restricted set, for both implicit and explicit relations, might lack such a bias. All dimensions proposed in this paper allow for useful paraphrases, and many for substitution tests, which can vary per language.

6.2 Theoretical issues

In this paper, we have described the goal of unifying existing discourse annotations, and we have proposed a solution to reach this goal by creating an interface that can make existing annotation systems "talk to each other." The basic

motivation is to make the best possible use of all those manual annotations that have already been made. For future annotation efforts, we would advise to take into account the lessons from previous annotation efforts, specifically in light of the comparison between annotations that have been made, and to take a clear stance – in terms of annotation guidelines – on those aspects that have shown here to lead to difficulties in mapping.

At the same time, this study has provided general insights that may contribute to more fundamental discussions on discourse annotation. First, apart from the issue of segmentation raised in the previous section, there is the fundamental question of what constitutes a coherence relation? Do we stick to the frequently applied key criterion of a relational surplus and require coherence relations to add meaning to the combination of their segments? Or should the list of "coherence relations" also include relation types that are defined by segment-specific features (e.g., RST's PROPORTION, PDTB's UNREAL PAST)? In line with the approach taken in this paper, it might be fruitful to combine both answers: work from categories defined by the relational surplus, and fill this in with the more detailed relation labels. That way, the whole inventory of detailed relations is organized in families.

Second, can the goals of descriptive adequacy – to give a full account of all relations in a corpus – and cognitive validity be combined? In line with the previous answer, we would claim this is possible when the global categories are based on theoretical as well as cognitive considerations.

Third, we should ensure that discourse annotation can be applied to other sources than written text. Clearly, it would be very attractive if discourse markers in spoken discourse (Crible 2017; Cuenca and Marin 2012), or language use on social media could be annotated with a compatible annotation system, so that comparing language use in different media and contexts becomes easier.

This is only a first list of the fundamental issues. In the meantime, we should be aware of the fact that practical reasons will often play a role in determining the exact labels that are used during annotation efforts. Researchers may have specific research questions that cannot be answered by a particular framework, or certain distinctions can be left out because that level of detail is not required for the research goal at hand.

Discussions on the best annotation systems for coherence relations have been on-going for several decades (Bunt and Prasad 2016; Hovy and Maier 1995). Any attempt to develop a final system for discourse annotation can benefit from the conclusions we have just outlined. Of course, developing a new framework would also imply that the existing annotated data would have to be reannotated,

because otherwise these could no longer be used. We therefore believe that the solution proposed in the current paper, using unifying dimensions (UniDim) to translate relation labels, is the most viable method to create coherence in the field and to bring together the existing frameworks. The UniDim approach allows for existing annotations to be reused, and allows researchers to keep using the framework that is best suited for answering their specific research questions. Researchers can easily add additional features, or disregard certain dimensions if these do not apply to the researchers' specific research questions.

In conclusion, we have proposed to make optimal use of existing annotations and to bring together existing annotation frameworks. To that end, we have created a mapping of three of the major annotation frameworks (PDTB, RST and SDRT) by decomposing them into unifying dimensions. We have also shown how the mapping works "in practice," by mapping PDTB and RST labels on the same set of text fragments. This has allowed us to identify cases for which the mapping works well, but also issues that complicate the mapping, such as segmentation and inter-annotator agreement. We hope to have shown the benefits of using a UniDim-account, which allows researchers to make even better use of the existing annotated corpora at the discourse level, thereby further developing this exciting field.

Appendix A: Classification of SDRT labels in terms of CCR dimensions and additional features.

SDRT	Polarity	Basic op.	Impl. order	SoC	Temporality	Add. features
Alternation	pos/neg	add/cau	any	obj/sub	(chron/anti)	(conditional), (alternative)
Background	pos	add	N.A.	obj	(chron/anti/syn)	
Commentary	pos	add	N.A.	sub	N.A.	
Consequence	pos	cau	bas/non-b	obj/sub	chron/anti	conditional
Continuation	pos	add	N.A.	obj/sub	N.A.	list
Contrast	neg	add/cau	any	obj/sub	(chron/anti)	
Elaboration	pos	add/cau	any	obj/sub	(chron/anti)	specificity, (list)
Explanation	pos	cau	non-b	obj	anti	(goal)
Narration	pos	add	N.A.	obj	chron	
Parallel	pos	add	N.A.	obj	N.A.	
Precondition	pos	add	N.A.	obj	anti	
Result	pos	cau	bas	obj	chron	

Notes.

– pos = positive, neg = negative, add = additive, cau = causal, bas = basic, non-b = non-basic, obj = objective, sub = subjective; chron = chronological, anti = anti-chronological, syn = synchronous, N.A. = does not apply, goal = goal-oriented relation.

– Classifications between brackets only apply to part of the relations falling under the respective category.

– We refrained from categorizing Source and Attribution in terms of the CCR dimensions, which are not considered to be coherence relations in the CCR approach because they concern a combination of a communication verb and its complement instead of a combination of two discourse segments.

Appendix B: Classification of PDTB-labels in terms of CCR dimensions and additional features.

Class_Subtype	End label	A1-A2	Pol.	Basic Op.	Impl. Order	SoC	Temp.	Add. features
Temporal								
Synchronous			pos	add	N.A.	obj	sync	
Asynchronous	Precedence	A1-A2	pos	add	N.A.	obj	chron	
	Precedence	A2-A1	pos	add	N.A.	obj	anti	
	Succession	A1-A2	pos	add	N.A.	obj	anti	
	Succession	A2-A1	pos	add	N.A.	obj	chron	
Contingency								
Cause	Reason	A1-A2	pos	cau	non-b	obj	anti	
	Reason	A2-A1	pos	cau	bas	obj	chron	
	Result	A1-A2	pos	cau	bas	obj	chron	(goal)
Pragm. cause	Justification	A1-A2	pos	cau	non-b	sub	N.A.	
	Justification	A2-A1	pos	cau	bas	sub	N.A.	
Condition	Hypothetical	A1-A2	pos	cau	non-b	obj/sub	anti/N.A.	conditional
	Hypothetical	A2-A1	pos	cau	bas	obj/sub	chron/N.A.	conditional
	General	A1-A2	pos	cau	non-b	obj	antic	conditional
	General	A2-A1	pos	cau	bas	obj	chron	conditional
	Unreal pres.	A1-A2	pos	cau	non-b	obj/sub	anti/N.A.	conditional
	Unreal pres.	A2-A1	pos	cau	bas	obj	chron	conditional
	Unreal past	A1-A2	pos	cau	non-b	obj/sub	anti/N.A.	conditional
	Unreal past	A2-A1	pos	cau	bas	obj	chron	conditional
	Factual pres.	A1-A2	pos	cau	non-b	sub	N.A.	conditional
	Factual pres.	A2-A1	pos	cau	bas	sub	N.A.	conditional
	Factual past	A1-A2	pos	cau	non-b	sub	N.A.	conditional
	Factual past	A2-A1	pos	cau	bas	sub	N.A.	conditional
Pragm. cond.	Relevance	A1-A2	pos	cau	non-b	sub	N.A.	conditional
	Relevance	A2-A1	pos	cau	bas	sub	N.A.	conditional
	Impl. assert.	A1-A2	pos	cau	non-b	sub	N.A.	conditional
	Impl. assert.	A2-A1	pos	cau	bas	sub	N.A.	conditional
Comparison								
Contrast	Juxtaposition		neg	add	N.A.	obj	any	
	Opposition		neg	add	N.A.	obj	any	
Pragm. contr.			neg	add	N.A.	subj	N.A.	
Concession	Expectation	A1-A2	neg	cau	non-b	obj/sub	anti/N.A.	
	Expectation	A2-A1	neg	cau	bas	obj/sub	chron/N.A.	
	Contra-expec.	A1-A2	neg	cau	bas	obj/sub	chron/N.A.	
Expansion								
Conjunction			pos	add	N.A.	obj/sub	N.A.	
Instantiation			pos	add	N.A.	obj/sub	N.A.	spec.-example

(*continued*)

Class_Subtype	End label	A1-A2	Pol.	Basic Op.	Impl. Order	SoC	Temp.	Add. features
Restatement	Specification		pos	add	N.A.	obj/sub	N.A.	specificity
	Equivalence		pos	add	N.A.	obj/sub	N.A.	spec.-equiv.
	Generalization		pos	add	N.A.	obj/sub	N.A.	specificity
Alternative	Conjunctive		pos	add	N.A.	obj/sub	N.A.	alternative
	Disjunctive		neg	add	N.A.	obj/sub	N.A.	alternative
	Disjunctive	A1-A2	neg	cau	bas	obj/sub	chron	conditional
	Disjunctive	A2-A1	neg	cau	non-b	obj/sub	anti	conditional
	Chosen alt.		neg	add	N.A.	obj/sub	N.A.	alternative
Exception			neg	add	N.A.	obj/sub	N.A.	
List			pos	add	N.A.	obj/sub	syn/chron/ N.A.	(list)

Notes.

–   A = Arg, pos = positive, neg = negative, add = additive, cau = causal, bas = basic, non-b = non-basic, obj = objective, sub = subjective; chron = chronological, anti = anti-chronological, syn = synchronous, N.A. = does not apply, equiv = equivalent.

–   No classification has been provided for Pragmatic Concession, which was listed as one of PDTB's labels (see Figure 1 in PDTB Research Group 2008: 27), but for which no definitions and/or examples were given in the remainder of the manual.

–   For Result, no Arg2-Arg1 counterpart is provided, since none of the connectives listed by the PDTB as signaling Result relations can be put in a constellation "Connective Arg2-Arg1." Note, however, that a cue phrase such as *in order to*, of which there are no instances in this PDTB category, would allow for such a constellation, and hence for a non-basic order in a Result relations. Also, the PDTB definition implies that Result concerns objective relations, but see our discussion in Section 3.3.

–   When determining the order of the segments in a relation, caution is required if this relation is marked by a double marker (see Section 3.6.2 for further details).

Appendix C: Classification of RST-DT labels in terms of CCR dimensions and additional features.

Class	End label	Nucl.	N-S	Pol.	Basic Op.	Impl. order	SoC	Temp.	Add. features
Background	Background	Mono	N-S	pos/neg	add	N.A.	obj	anti/N.A.	
	Background	Mono	S-N	pos/neg	add	N.A.	obj	chron/N.A.	
	Circumstance	Mono		pos/neg	add	N.A.	obj	syn/N.A.	
Cause	Cause	Mono	N-S	pos	cau	bas	obj	chron	
	Cause	Mono	S-N	pos	cau	non-b	obj	anti	
	Cause-result	Multi		pos	cau	bas/non-b	obj	chron/anti	
	Result	Mono	N-S	pos	cau	non-b	obj	anti	
	Result	Mono	S-N	pos	cau	bas	obj	chron	
	Consequence-n	Mono	N-S	pos	cau	non-b	obj	anti	
	Consequence-n	Mono	S-N	pos	cau	bas	obj	chron	
	Consequence-s	Mono	N-S	pos	cau	bas	obj	chron	
	Consequence-s	Mono	S-N	pos	cau	non-b	obj	anti	
	Consequence	Multi		pos	cau	bas/non-b	obj	chron/anti	
Comparison	Comparison	Both		pos	add	N.A.	obj/sub	N.A.	
	Preference	Mono		neg	add	N.A.	obj/sub	N.A.	
	Analogy	Both		pos	add	N.A.	sub	N.A.	
	Proportion	Multi		pos	add/cau	any	obj/sub	any	
Conditional	Condition	Mono	N-S	pos/neg	cau	non-b	obj/sub	anti/N.A.	conditional
	Condition	Mono	S-N	pos/neg	cau	bas	obj/sub	chron/N.A.	conditional
	Hypothetical	Mono	N-S	pos	cau	non-b	sub	N.A.	conditional
	Hypothetical	Mono	S-N	pos	cau	bas	sub	N.A.	conditional
	Contingency	Mono	N-S	pos/neg	cau	non-b	obj	anti	conditional
	Contingency	Mono	S-N	pos/neg	cau	bas	obj	chron	conditional
	Otherwise	Mono	N-S	neg	cau	bas	obj/sub	chron/N.A.	conditional
	Otherwise	Multi		neg	cau	bas	obj/sub	chron/N.A.	conditional
Contrast	Contrast	Multi		neg	add	N.A.	obj/sub	any	
	Concession	Mono	N-S	neg	cau	non-b	obj/sub	anti/N.A.	
	Concession	Mono	S-N	neg	cau	bas	obj/sub	chron/N.A.	
	Antithesis	Mono		neg	add/cau	any	obj/sub	any	
Elaboration	El.-additional	Mono		pos	add	N.A.	obj/sub	N.A.	
	El.-gen.-spec.	Mono		pos	add	N.A.	obj/sub	N.A.	specificity
	El.-part-whole	Mono		pos	add	N.A.	obj	N.A.	specificity
	El.-process-step	Mono		pos	add	N.A.	obj	N.A.	specificity
	El.-object-attr.	Mono		pos	add	N.A.	obj	N.A.	specificity
	El.-set-member	Mono		pos	add	N.A.	obj	N.A.	spec.-ex.
	Example	Mono		pos	add	N.A.	obj	N.A.	spec.-ex.
	Definition	Mono		pos	add	N.A.	obj	N.A.	specificity
Enablement	Purpose	Mono	N-S	pos	cau	bas	obj/sub	chron/N.A.	goal
	Purpose	Mono	S-N	pos	cau	non-b	obj/sub	anti/N.A.	goal
	Enablement	Mono	N-S	pos	cau	non-b	obj/sub	anti/N.A.	goal
	Enablement	Mono	S-N	pos	cau	bas	obj/sub	chron/N.A.	goal
Evaluation	Evaluation	Both		pos	add/cau	any	sub	N.A.	specificity
	Interpretation	Both		pos	add/cau	any	sub	N.A.	specificity
	Conclusion	Mono	N-S	pos	cau	bas	sub	N.A.	specificity

(*continued*)

Class	End label	Nucl.	N-S	Pol.	Basic Op.	Impl. order	SoC	Temp.	Add. features
	Conclusion	Mono	S-N	pos	cau	non-b	sub	N.A.	specificity
	Conclusion	Multi		pos	cau	bas/non-b	sub	N.A.	specificity
	Comment	Mono		pos	add	N.A.	sub	N.A.	specificity
Explanation	Evidence	Mono	N-S	pos	cau	non-b	sub	anti	
	Evidence	Mono	S-N	pos	cau	bas	sub	chron	
	Exp.-argument.	Mono	N-S	pos	cau	non-b	obj	anti	
	Exp.-argument.	Mono	S-N	pos	cau	bas	obj	chron	
	Reason	Mono	N-S	pos	cau	non-b	obj	anti	
	Reason	Mono	S-N	pos	cau	bas	obj	chron	
	Reason	Multi		pos	cau	bas/non-b	obj	chron/anti	
Joint	List	Multi		pos	add	N.A.	obj/sub	syn/chron/ N.A.	list
	Disjunction	Multi		pos/neg	add	N.A.	obj/sub	syn/N.A.	alternative
Summary	Summary	Mono		pos	add	N.A.	obj	N.A.	specificity
	Restatement	Mono		pos	add	N.A.	obj	N.A.	spec.-equiv.
Temporal	Temp.-before	Mono	N-S	pos	add	N.A.	obj	chron	
	Temp.-before	Mono	S-N	pos	add	N.A.	obj	anti	
	Temp.-after	Mono	N-S	pos	add	N.A.	obj	anti	
	Temp.-after	Mono	S-N	pos	add	N.A.	obj	chron	
	Tmp.-same- time	Both		pos	add	N.A.	obj	syn	
	Sequence	Multi		pos	add	N.A.	obj	chron	
	Inverted-seq.	Multi		pos	add	N.A.	obj	anti	
Manner-	Means	Mono	N-S	pos	cau	non-b	obj	anti	
Means	Means	Mono	S-N	pos	cau	bas	obj	chron	goal
Topic-	Problem-sol.-n	Mono	N-S	pos	cau	non-b	obj/sub	anti/N.A.	goal
Comment	Problem-sol.-n	Mono	S-N	pos	cau	bas	obj/sub	chron/N.A.	goal
	Problem-sol.-s	Mono	N-S	pos	cau	bas	obj/sub	chron/N.A.	goal
	Problem-sol.-s	Mono	S-N	pos	cau	non-b	obj/sub	anti/N.A.	goal
	Problem-sol.	Multi		pos	cau	bas/non-b	obj/sub	chron/anti/ N.A.	goal

Notes.

− N = nucleus, S = satellite, pos = positive, neg = negative, add = additive, cau = causal, bas = basic, non-b = non-basic, obj = objective, sub = subjective; chron = chronological, anti = anti-chronological, syn = synchronous, N.A. = does not apply, ex = example, equiv = equivalent.

− We refrained from classifying ATTRIBUTION relations, which in CCR are not considered to be coherence relations because they concern a combination of a communication verb and its complement instead of a combination of two discourse segments.

− We also disregarded relation labels introducing a classification that was orthogonal to another classification in terms of coherence. This was the case for the class of TOPIC CHANGE, which includes the labels TOPIC SHIFT and TOPIC DRIFT, because not only combinations of larger text spans but any combination of two discourse segments can be labeled in terms of topic shift vs. topic drift, on top of another classification in terms of type of coherence. Similarly, no CCR labeling was provided for the relation labels focusing on the illocutionary status of discourse segments (i.e., QUESTION-ANSWER, STATEMENT-RESPONSE, COMMENT-TOPIC, TOPIC-COMMENT, and RHETORICAL QUESTION). Apart from the type of speech act involved, the related segments can still be classified in terms of another coherence relation.

− MANNER relations involve segments that do not meet the clause criterion set by CCR (see Section 2.1). Therefore, these types of relations have not been classified in terms of the CCR dimensions.

− OTHERWISE relations always involve a relation in which *otherwise* is in S2, which is why no Satellite-Nucleus version is provided for this relation.

Appendix D: Framework-specific features and related questions.

Framework	Question	If yes	If no
SDRT	(Only for additive and subjective relations) **Commentary** – Is the judgment in S2 about the content of S1?	Commentary	Elaboration
PDTB	**Antonym** – Are the values ascribed to some shared property antonyms, or different values on a scale?	Opposition	Juxtaposition
PDTB	**Generalizability** – Does this conditional relation *always* hold?	General	Hypothetical
PDTB	**Tense** – Does the *if*-clause hold in the present (*yes*) or in the past (*no*)?	Factual present	Factual past
PDTB	**Tense** – Is the *if*-clause presented to not hold in the present (*yes*) or to not have held in the past (*no*)?	Unreal present	Unreal past
PDTB	**Stance*** – Does the conditional relation convey a negative stance or an expectation that the *if*-clause is not going to hold/did not hold?	Unreal past/present	Hypothetical
RST	**Proportionality** – Do the two segments express a proportionality or equivalence of tendency?	Proportion	Other positive relation
RST	**Preference –** Is a clear preference indicated for one of the situations, acts, events, etc., that are compared in the two segments?	Preference	Contrast
RST	**Analogy –** Does the relation contain an inference that if the two events, situations, etc., agree with each other in some respects, they will probably agree in other respects?	Analogy	Other pos. add. subjective relation
RST	**Expectation –** Is there a violated expectation in the relation?	Concession	Antithesis
RST	**Animate agent** – Does the relation involve a justification or explanation of the actions of an animate agent?	Reason	Explanation-argumentative

*This distinction is also related to tense, as negative stance can be expressed through temporal distancing (from the time of speaking > e.g. indirect quote, past tense narrative).

References

Afantenos, Stergos, Nicholas Asher, Farah Benamara, Myriam Bras, Cécile Fabre, Mai Ho-Dac, Anne Le Draoulec, Philippe Muller, Marie-Paule Pery-Woodley, Laurent Prevot, Josette Rebeyrolles, Ludovic Tanguy, Marianne Vergez-Couret & Laure Vieu. 2012. An empirical resource for discovering cognitive principles of discourse organisation: The Annodis Corpus. *Proceedings of the Eight International Conference on Language Resources and Evaluation*. Istanbul, Turkey.

Andersson, Marta. 2016. *The architecture of result relations: Corpus and experimental approaches to result coherence relations in English*. Stockholm: University of Stockholm dissertation.

Asher, Nicholas & Alex Lascarides. 2003. *Logics of conversation*, Cambridge: Cambridge University Press.

Asr, Fatemeh Torabi & Vera Demberg. 2012. Implicitness of discourse relations. *Proceedings of COLING*. Mumbai, India.

Baker, Linda. 1978. Processing temporal relationships in simple stories: Effects of input sequence. *Journal of Verbal Learning and Verbal Behavior* 17(5). 559–572.

Benamara, Farah & Maite Taboada. 2015. Mapping different rhetorical relation annotations: A proposal. *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics (SEM 2015)*. Denver, USA.

Bloom, Lois, Margaret Hood Lahey, Lois Lifter & Karin Fiess. 1980. Complex sentences: Acquisition of syntactic connectives and the semantic relations they encode. *Journal of Child Language* 7(2). 235–261.

Bosma, Wauther. 2005. Query-based summarization using Rhetorical Structure Theory. *LOT Occasional Series* 4. 29–44. http://doc.utwente.nl/65574/1/clin04.pdf (accessed 25 April 2018).

Bunt, Harry & Rashmi Prasad. 2016. Core concepts for the annotation of discourse relations. In *Proceedings 12th Joint ACL-ISO Workshop on Interoperable Semantics (ISA-12)*, 45–54. Portoroz, Slovenia/Paris: ELRA.

Canestrelli, Annoloes, Willem Mak & Ted Sanders. 2013. Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes* 28(9). 1394–1413.

Carlson, Lynn & Daniel Marcu. 2001. *Discourse tagging reference manual*. http://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf (accessed 25 April 2018).

Chiarcos, Christian. 2014. Towards interoperable discourse annotation: Discourse features in the Ontologies of Linguistic Annotation. *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (*LREC'14*). Iceland: Reykjavík. http://www.lrec-conf.org/proceedings/lrec2014/pdf/893_Paper.pdf (accessed 25 April 2018).

Clark, Herbert & Eve Clark. 1968. Semantic distinctions and memory for complex sentences. *The Quarterly Journal of Experimental Psychology* 20(2). 129–138.

Clark, Herbert & Eve Clark. 1977. *Psychology and language: An introduction to psycholinguistics*, New York: Harcourt Brace Jovanovich.

Crible, Ludivine. 2017. *Discourse markers and (dis)fluency across registers: A contrastive usage-based study in English and French*. Louvain-la-Neuve: Université Catholique de Louvain dissertation.

Cuenca, Maria-Josep & Maria-Josep Marín. 2012. Discourse markers and modality in spoken Catalan: The case of *(és) clar*. *Journal of Pragmatics* 44(15). 2211–2225.

Dancygier, Barbara & Eve Sweetser. 2005. *Mental spaces in grammar: Conditional constructions*, Cambridge: Cambridge University Press.

Das, Debopam & Maite Taboada. 2018. RST Signalling Corpus: A corpus of signals of coherence relations. *Language Resources & Evaluation* 52. 149–184.

Degand, Liesbeth & Henk Pander Maat. 2003. A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. In Arie Verhagen & Jeroen van de Weijer (eds.), *Usage Based Approaches to Dutch*, 175–199. Utrecht: LOT.

Demberg, Vera, Fatemeh Torabi Asr & Merel Scholman. 2017. How consistent are our discourse annotations? Insights from mapping RST-DT and PDTB annotations. *ArXiv e-prints April* arXiv: 1704.08893.

Evers-Vermeul, Jacqueline. 2005. *The development of Dutch connectives: Change and acquisition as windows on form-function relations*. Utrecht: University of Utrecht dissertation. http://www.lotpublications.nl/Documents/110_fulltext.pdf.

Evers-Vermeul, Jacqueline, Jet Hoek & Merel Scholman. 2017. On temporality in discourse annotation: Theoretical and practical considerations. *Dialogue & Discourse* 8(2). 1–20.

Evers-Vermeul, Jacqueline & Ted Sanders. 2009. The emergence of Dutch connectives: How cumulative cognitive complexity explains the order of acquisition. *Journal of Child Language* 36(4). 829–854.

Evers-Vermeul, Jacqueline & Ted Sanders. 2011. Discovering domains: On the acquisition of causal connectives. *Journal of Pragmatics* 43(6). 1645–1662.

Ford, Cecilia. 1993. *Grammar in interaction: Adverbial clauses in American English conversations*, Cambridge: Cambridge University Press.

Graesser, Arthur, Murray Singer & Tom Trabasso. 1994. Constructing inferences during narrative text comprehension. *Psychological Review* 101(3). 371–395.

Grosz, Barbara & Candace Sidner. 1986. Attention, intentions and the structure of discourse. *Computational Linguistics* 12(3). 175–204.

Halliday, Michael & Ruqaiya Hasan. 1976. *Cohesion in English*, London: Longman.

Hobbs, Jerry. 1979. Coherence and coreference. *Cognitive Science* 3. 67–90.

Hoek, Jet, Jacqueline Evers-Vermeul & Ted Sanders. 2017a. Segmenting discourse: Incorporating interpretation into segmentation? *Corpus Linguistics and Linguistic Theory* [online preview].

Hoek, Jet, Sandrine Zufferey, Jacqueline Evers-Vermeul & Ted Sanders. 2017b. Cognitive complexity and the linguistic marking of coherence relations: A parallel corpus study. *Journal of Pragmatics* 121. 113–131.

Hoey, Michael. 1983. *On the surface of discourse*, London: Allen & Unwin.

Hovy, Eduard & Elisabeth Maier. 1995. Parsimonious or profligate: How many and which discourse structure relations. *Unpublished manuscript*. http://www.isi.edu/natural-language/people/hovy/papers/93discproc.pdf (accessed 25 April 2018).

Iruskieta, Mikel, Maria Aranzabe, Arantza De Ilarraza, Itziar Gonzalez, Mikel Lersundi & de la Calle Oier. 2013. The RST Basque TreeBank: An online search interface to check rhetorical relations. *Proceedings of the 4th Workshop RST and Discourse Studies*. Fortaleza, Brazil.

Kehler, Andrew. 2002. *Coherence, reference, and the theory of grammar*, Stanford: CSLI Publications.

Knott, Alistair & Robert Dale. 1994. Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes* 18. 35–62.

Knott, Alistair & Ted Sanders. 1998. The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics* 30. 135–175.

Koehn, Philipp. 2009. *Statistical Machine Translation*, Cambridge: Cambridge University Press.

Lakoff, Robyn. 1971. If's, and's and but's about conjunction. In Charles Fillmore & D. Terrence Langendoen (eds.), *Studies in linguistic semantics*, 114–149. New York: Holt.

Li, Fang, Jacqueline Evers-Vermeul & Ted Sanders. 2013. Subjectivity and result marking in Mandarin: A corpus-based investigation. *Chinese Language and Discourse* 4(1). 74–119.

Li, Fang, Ted Sanders & Jacqueline Evers-Vermeul. 2016. On the subjectivity of Mandarin reason connectives: Robust profiles or genre-sensitivity? In Wilbert Spooren, Gerard Steen & Ninke Stukker (eds.), *Genre in language, discourse and cognition*, 13–49. Berlin: De Gruyter Mouton.

Louis, Annie, Aravind Joshi & Ani Nenkova. 2010. Discourse indicators for content selection in summarization. *Proceedings of SIGDIAL 2010: The 11th Annual Meeting of the Special Interest Group on Discourse and Dialogu*e, 147–156. Tokyo, Japan.

Mann, William & Sandra Thompson. 1986. Relational propositions in discourse. *Discourse Processes* 9(1). 57–90.

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

Marcu, Daniel. 2000. *The theory and practice of discourse parsing and summarization*, Cambridge: MIT Press.

Martin, James. 1992. *English text: System and structure*, Amsterdam & Philadelphia: John Benjamins.

Meyer, Bonnie. 1985. Prose analysis: Purposes, procedures, and problems. In Bruce Britton & John Black (eds.), *Understanding expository text*, 269–304. Hillsdale: Erlbaum.

Meyer, Thomas & Andrei Popescu-Belis. 2012. Using sense-labeled discourse connectives for statistical machine translation. *Proceedings of the EACL 2012 Joint ESIRMT-HyTra Workshop*, 129–138. Avignon, France.

Meyer, Thomas, Andrei Popescu-Belis, Sandrine Zufferey & Bruno Cartoni. 2011. Multilingual annotation and disambiguation of discourse connectives for machine translation. *Proceedings of the SIGDIAL 2011 Conference*, 194–203. Portland, Oregon.

Mithun, Shamima. 2010. Exploiting rhetorical relations in blog summarization. In Atefeh Varzindar & Vlado Kešelj (eds), *Advances in artificial intelligence*, 388–392. Berlin/ Heidelberg: Springer.

Noordman, Leo & Femke de Blijzer. 2000. On processing causal relations. In Elisabeth Couper-Kuhlen & Bernd Kortmann (eds.), *Cause, condition, concession, contrast: Cognitive and discourse perspectives*, 35–55. Berlin: Mouton de Gruyter.

Noordman, Leo & Wietske Vonk. 1998. Memory-based processing in understanding causal information. *Discourse Processes* 26(2–3). 191–212.

PDTB Research Group (2008). The Penn Discourse Treebank 2.0 Annotation Manual. *Technical Report IRCS-08-01*. Philadelphia: Institute for Research in Cognitive Science, University of Pennsylvania. https://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtb-annotation-manual.pdf.

Pit, Mirna. 2007. Cross-linguistic analyses of backward causal connectives in Dutch, German and French. *Languages in Contrast* 7. 53–82.

Poláková, Lucie, Jiri Mírovský, Anna Nedoluzhko, Pavlina Jínová, Sarka Zikánová & Eva Hajicová. 2013. Introducing the Prague Discourse Treebank 1.0. *Proceedings of the 6th International Joint Conference on Natural Language Processing*, 91–99. Nagoya, Japan.

Prasad, Rashmi, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi & Bonnie Webber. 2008. The Penn Discourse Treebank 2.0. *Proceedings of the 6th International Conference of Language Resources and Evaluation (LREC 2008)*. Marrakech: Morocco. https://www.seas.upenn.edu/~pdtb/papers/pdtb-lrec08.pdf.

Prasad, Rashmi, Bonnie Webber, Alan Lee & Aravind Joshi. in preparation. *Discourse Relations in the PDTB 3.0*.

Reese, Brian, Julie Hunter, Nicholas Asher, Pascal Denis & Jason Baldridge. 2007. Reference manual for the analysis and annotation of rhetorical structure (version 1.0). Technical report. Austin: University of Texas, Departments of Linguistics and Philosophy. http://timeml.org/jamesp/annotation_manual.pdf.

Sanders, Ted. 1997. Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context. *Discourse Processes* 24. 119–147.

Sanders, Ted & Leo Noordman. 2000. The role of coherence relations and their linguistic markers in text processing. *Discourse Processes* 29. 37–60.

Sanders, Ted & Merel Scholman. 2015. Annotating coherence relations in corpora of language use. Paper presented at the 14th International Pragmatics Conference (IPRA). Antwerp, Belgium.

Sanders, Ted & Wilbert Spooren. 1999. Communicative intentions and coherence relations. In Wolfram Bublitz, Uta Lenk & Eija Ventola (eds.), *Coherence in spoken and written discourse*, 235–250. Amsterdam & Philadelphia: John Benjamins.

Sanders, Ted & Wilbert Spooren. 2009a. Causal categories in discourse: Converging evidence from language use. In Ted Sanders & Eve Sweetser (Eds.), *Causal categories in discourse and cognition*, 205–246. Berlin: Mouton de Gruyter.

Sanders, Ted & Wilbert Spooren. 2009b. The cognition of discourse coherence. In Jan Renkema (ed.), *Discourse, of course*, 197–212. Amsterdam: John Benjamins.

Sanders, Ted, Wilbert Spooren & Leo Noordman. 1992. Toward a taxonomy of coherence relations. *Discourse Processes* 15. 1–35.

Sanders, Ted, Wilbert Spooren & Leo Noordman. 1993. Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics* 4(2). 93–133.

Sanders, Ted, Kirsten Vis & Daan Broeder. 2012. Project notes on the Dutch project DiscAn. *Eighth Joint ACL – ISO Workshop on Interoperable Semantic Annotation*. Pisa, Italy.

Schilperoord, Joost & Arie Verhagen. 1998. Conceptual dependency and the clausal structure of discourse. In Jean-Pierre Koenig (Ed.), *Discourse and cognition. Bridging the gap*, Stanford: CSLI publications.

Scholman, Merel, Jacqueline Evers-Vermeul & Ted Sanders. 2016. Categories of coherence relations in discourse annotation: Towards a reliable categorization of coherence relations. *Dialogue & Discourse* 7(2). 1–28.

Singer, Murray, Michael Halldorson, Jeffrey Lear & Peter Andrusiak. 1992. Validation of causal bridging inferences in discourse understanding. *Journal of Memory and Language* 31(4). 507–524.

Spooren, Wilbert. 1989. *Some aspects of the form and interpretation of global contrastive coherence relations*. Nijmegen: Radboud University Nijmegen dissertation.

Spooren, Wilbert. 1997. The processing of underspecified coherence relations. *Discourse Processes* 24(1). 149–168.

Spooren, Wilbert & Liesbeth Degand. 2010. Coding coherence relations: Reliability and validity. *Corpus Linguistics and Linguistic Theory* 6. 241–266.

Spooren, Wilbert & Ted Sanders. 2008. The acquisition order of coherence relations: On cognitive complexity in discourse. *Journal of Pragmatics* 40(12). 2003–2026.

Stukker, Ninke & Ted Sanders. 2012. Causal connectives in discourse: A cross-linguistic perspective. *Journal of Pragmatics* 44(2). 169–190.

Stukker, Ninke, Ted Sanders & Arie Verhagen. 2008. Causality in verbs and in discourse connectives. Converging evidence of cross-level parallels in Dutch linguistic categorization. *Journal of Pragmatics* 40(7). 1296–1322.

Sweetser, Eve. 1990. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*, Cambridge: Cambridge University Press.

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

Townsend, David. 1983. Thematic processing in sentences and texts. *Cognition* 13(2). 223–261.

Traxler, Matthew, Michael Bybee & Martin Pickering. 1997. Influence of connectives on language comprehension: Eye-tracking evidence for incremental interpretation. *The Quarterly Journal of Experimental Psychology* 50A(3). 481–497.

van den Broek, Paul. 1990. Causal inferences in the comprehension of narrative texts. In Arthur Graesser & G. H. Bower (eds.), *Psychology of learning and motivation: Inferences and text comprehension*, 175–196. New York, NY: Academic Press.

van Enschot, Renske, Antal van den Bosch, Christian Burgers, Liesbeth Degand, Florian Kunneman, Yvette Linders, Fons Maes & Wilbert Spooren. submitted. *Taming our wild data: On interrater reliability in discourse research*, Submitted for publication.

van Silfhout, Gerdineke, Jacqueline Evers-Vermeul & Ted Sanders. 2015. Connectives as processing signals: How students benefit in processing narrative and expository texts. *Discourse Processes* 52. 47–76.

van Veen, Rosie. 2011. *The acquisition of causal connectives: The role of parental input and cognitive complexity*. Utrecht: Utrecht University dissertation. http://www.lotpublications.nl/Documents/286_fulltext.pdf.

Verberne, Susan, Lou Boves, Peter-Arno Coppen & Nelleke Oostdijk. 2007. Discourse-based answering of *why*-questions: Employing RST structure for finding answers to *why*-questions. *Traitement Automatique des Langues* 47(2). 21–41.

Webber, Bonnie, Rashmi Prasad, Alan Lee & Aravind Joshi. 2016. A discourse-annotated corpus of conjoined VPs. In A. Friedrich & K. Tomanek (eds.), *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL*, 22–31. Berlin: ACL.

Xue, Nianwen, Hwee Ng, Sameer Pradhan, Rashmi Prasad, Christopher Bryant & Attapol Rutherford. 2015. The CoNLL–2015 shared task on shallow discourse parsing. *Proceedings of the 19th Conference on Computational Natural Language Learning: Shared Task*, 1–16. Beijing, China.

Zufferey, Sandrine. 2010. *Lexical pragmatics and theory of mind: The acquisition of connectives*, Amsterdam: John Benjamins.

Zufferey, Sandrine. 2012. 'Car, parce que, puisque' revisited. Three empirical studies on French causal connectives. *Journal of Pragmatics* 44(2). 138–153.

Zufferey, Sandrine, Willem Mak, Sara Verbrugge & Ted Sanders. 2018. Usage and processing of the French connectives 'car' and 'parce que'. *Journal of French Language Studies* 28(1). 85–112.

Bionotes

Ted J.M. Sanders

Ted J.M. Sanders is a full professor of language use and discourse studies at Utrecht University. Since his PhD (1992), he has worked on a cognitive approach to coherence relations (CCR). Believing in converging evidence, he combines (cognitive and text) linguistics with corpus studies in cross-linguistic comparison, as well as experimental psycholinguistics of acquisition, discourse processing and representation. Ted Sanders was a board member of the *Society for Text and Discourse*, serving as President from 2010 to 2013. He is currently Vice Dean of the Utrecht Faculty of Humanities, responsible for the Graduate School.

Vera Demberg

Vera Demberg (born 1981, PhD University of Edinburgh, 2010) is a professor of computer science and computational linguistics at Saarland University. Her research interests include discourse processing, and specifically coherence relations, which she has been working on using both experimental psycholinguistic and computational methods.

Jet Hoek

Jet Hoek is a PhD candidate in linguistics at Utrecht University (The Netherlands). Her research focuses on discourse structure and the linguistic marking of coherence relations, using both corpus-based and experimental methods.

Merel C.J. Scholman

Merel Scholman is a PhD candidate at Saarland University (Saarbrücken, Germany). She uses a combination of off-line and on-line methodologies – corpus-based studies, crowdsourced studies, and lab reading time experiments – to investigate questions related to the interpretation and processing of discourse. Her PhD project focuses on distinctions between coherence relations in theories, annotations, and interpretations.

Fatemeh Torabi Asr

Fatemeh Torabi Asr (born 1985, PhD Saarland University, 2015) is a software engineer with research interests at the intersection of computer science, linguistics, and psychology. She is currently a post-doctoral researcher at the Discourse Processing Lab in Simon Fraser University, Canada. Her works are majorly focused around computational analysis of discourse and semantic phenomena and applying machine learning to natural language processing tasks.

Sandrine Zufferey

Sandrine Zufferey (born 1978, PhD University of Geneva, 2007) is a professor of French linguistics at the University of Bern in Switzerland. Her current research focuses on the way native and non-native speakers use, understand and process discourse connectives, as well as on the way discourse relations are encoded across languages and in translations.

Jacqueline Evers-Vermeul

Jacqueline Evers-Vermeul, assistant professor at the Department of Languages, Literature and Communication at Utrecht University, studies the representation, production and comprehension of text-structural aspects of discourse. She uses a variety of methods to study adult connective use (synchronic and diachronic corpus-based analyses), students' understanding of school texts and examination questions (on- and off-line reading experiments among children and teenagers), and children's connective acquisition (visual-world experiments among 2- and 3-year-olds, growth-curve analyses of dense longitudinal corpora).