



Discours

Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics

30 | 2022
Varia

Descriptively Adequate *and* Cognitively Plausible? Validating Distinctions between Types of Coherence Relations

Merel C.J. Scholman, Vera Demberg and Ted J.M. Sanders



Electronic version

URL: <https://journals.openedition.org/discours/12075>

DOI: 10.4000/discours.12075

ISSN: 1963-1723

Publisher:

Laboratoire LATTICE, Presses universitaires de Caen

Electronic reference

Merel C.J. Scholman, Vera Demberg and Ted J.M. Sanders, "Descriptively Adequate *and* Cognitively Plausible? Validating Distinctions between Types of Coherence Relations", *Discours* [Online], 30 | 2022, Online since 23 September 2022, connection on 03 February 2023. URL: <http://journals.openedition.org/discours/12075> ; DOI: <https://doi.org/10.4000/discours.12075>



Creative Commons - Attribution 4.0 International - CC BY 4.0
<https://creativecommons.org/licenses/by/4.0/>

Descriptively Adequate *and* Cognitively Plausible? Validating Distinctions between Types of Coherence Relations

Merel C.J. Scholman

Saarland University

Vera Demberg

Saarland University

Ted J.M. Sanders

Utrecht University

.....
Merel C.J. Scholman, Vera Demberg, Ted J.M. Sanders, « Descriptively Adequate *and* Cognitively Plausible? Validating Distinctions between Types of Coherence Relations », *Discours* [En ligne], 30 | 2022, mis en ligne le 23 septembre 2022.

.....
URL : <http://journals.openedition.org/discours/12075>

.....
Titre du numéro : *Varia*

Coordination : Anne Le Draoulec & Josette Rebeyrolle

Date de réception de l'article : 14/10/2021

Date d'acceptation de l'article : 14/02/2022

Descriptively Adequate *and* Cognitively Plausible? Validating Distinctions between Types of Coherence Relations

Merel C.J. Scholman
Saarland University

Vera Demberg
Saarland University

Ted J.M. Sanders
Utrecht University

.....
A central issue in linguistics concerns the relationship between theories and evidence in data. We investigate this issue in the field of discourse coherence, and particularly the study of coherence relations such as causal and contrastive. Proposed inventories of coherence relations differ greatly in the type and number of proposed relations. Such proposals are often validated by focusing on either the descriptive adequacy (researcher's intuitions on textual interpretations) or the cognitive plausibility of distinctions (empirical research on cognition). We argue that both are important, and note that the concept of cognitive plausibility is in need of a concrete definition and quantifiable operationalization. This contribution focuses on how the criterion of cognitive plausibility can be operationalized and presents a systematic validation approach to evaluate discourse frameworks. This is done by detailing how various sources of evidence can be used to support or falsify distinctions between coherence relational labels. Finally, we present methodological issues regarding verification and falsification that are of importance to all discourse researchers studying the relationship between theory and data.

Keywords: discourse, coherence relations, cognitive plausibility, descriptive adequacy, discourse theories

1. Validation of theories

- 1 A central issue in linguistics concerns the relationship between theories and evidence in data. The *Generalization Commitment*, proposed for the cognitive linguistics field, emphasizes that we seek general principles in our theoretical descriptions of linguistic phenomena (Lakoff, 1991). The *Cognitive Commitment* emphasizes the importance of incorporating a wide range of data from other disciplines, such as developmental psychology and psycholinguistics, into our theoretical description of language. Both principles stress the commitment to engage in scientific research and facilitate a debate on the relationship between theories and empirical evidence (Gibbs, 1996).
- 2 We apply the methodological debate on the relationship between data and theory to the disagreement that exists in the field of discourse coherence: which *coherence*

relation types should be distinguished? Coherence relations are semantic links between clauses and sentences (Hobbs, 1979; Sanders et al., 1992), as illustrated in Example [1].

[1] The spelling bee participant was ecstatic after she correctly spelled “vivisepulture”.

3 This sentence expresses two events that are linked in a temporal relationship, marked by the connective *after*. However, comprehenders are likely to interpret these events as being causally related: the participant was ecstatic *because* she spelled a word correctly. Hence, comprehenders assign more meaning to this sentence than that conveyed by the individual clauses alone (Sanders et al., 1992). They do this by inferring coherence relations.

4 In the field of discourse coherence, it is generally assumed that coherence relations can be characterized into a fixed, limited number of types. On a coarse-grained level, most proposals agree on four main types: causal, additive, temporal, and adversative relations. However, various proposals for finer-grained relations exist and, as a result, the field lacks a uniform set of relations that is agreed upon by all (see Bunt & Prasad, 2016; Hovy & Maier, 1995; Sanders et al., 2018; for attempts to unify frameworks). The lack of agreement on a set of relations can, in part, be attributed to the existence of a variety of different criteria for introducing distinctions. One goal of the current contribution is to discuss what the best-principled criteria for evaluating these relations can be.

5 The development of a general theory of coherence relations would be of great benefit to the field: an empirically validated theory that can account for the discourse structure can provide more insight into the link between the structure of a text as a linguistic object, its cognitive representations and the processes of text production and understanding. Such a theory, located at the intersection of linguistics and psycholinguistics, could lead to significant progress in the field of discourse studies.

6 How can we then evaluate and validate distinctions made in existing theories? One criterion would be for researchers to provide linguistic examples and have experts agree on their categorization. This is related to the *descriptive adequacy* of an approach. However, adhering to only this criterion can make it difficult to falsify distinctions (Knott & Dale, 1994). Consider Example [2].

[2] There was a banana peel on the ground. John slipped on it.

7 Imagine that the authors of this paper were to agree that this is an example of a unique relational category within the class of causal relations, namely BANANA PEEL. If relational distinctions are validated by providing examples and having experts agree, this would mean that the BANANA PEEL relation should be considered a valid distinction. Absurd as this example might seem, it emphasizes the need for additional criteria for validating coherence theories. Without a more stringent validation process, disagreement within the field and proliferation of relation types are inevitable.

8 We therefore suggest that proposed distinctions should be validated using *cognitively plausible* evidence, including a wide range of data from various disciplines and modalities (corresponding to the Cognitive Commitment as well as earlier efforts to motivate theories of coherence, e.g., Crible & Degand, 2019; Degand, 1998; Evers-Vermeul et al., 2017; Foltz, 2003; Knott, 1996; Sanders et al., 1992; Zufferey & Degand, 2017). By evaluating distinctions made in coherence relation frameworks similarly to how cognitive theories would be evaluated (that is, subjecting them to stringent criteria and rigorous testing by others), we can improve our understanding of the phenomenon of discourse coherence. Evaluating and validating frameworks as theories should therefore not be viewed as a criticism of a particular framework but as an effort to contribute to a more general theory of coherence.

9 The current paper argues that two general criteria of validation are essential: *descriptive adequacy* and *cognitive plausibility*. We first introduce each criterion, providing information on what the requirement means in general and what it means for coherence theories in particular. The paper then goes on to make the following contributions:

- We argue that generalizable theories of coherence relations should be both descriptively adequate and cognitively plausible.
- The criterion of cognitive plausibility is concretized by outlining possible sources of evidence that can provide support for particular distinctions.
- An approach to validate coherence relation theories is proposed, focusing on converging evidence and replication of results using various sources.

10 The concrete objective of this paper is to contribute to future research efforts aimed at motivating a set of relational distinctions, as well as spark a debate about the cognitive reality of distinctions and the development of a comprehensive theory of coherence. Such a debate within the community will be instrumental in gaining a deeper understanding of the link between the structure of a text as a linguistic object, its cognitive representations, and the processes of text production and understanding.

1.1. Descriptive adequacy

11 Descriptive adequacy is a quality measure for the evaluation of linguistic theories. It was first described by Chomsky (1965), who considered theories to have descriptive adequacy if they can account for a finite corpus of data and for the linguistic intuitions of native speakers. Although Chomsky's theory was originally proposed to account for the adequacy of grammars, it can be applied to account for discourse theories as well. Brewer and Lichtenstein (1982) tested the descriptive adequacy of theories of discourse comprehension by asking participants to rate the naturalness of stories that adhered to one of the theories. Sanders et al. (1992) consider a theory to be descriptively adequate if text analysts can use it to describe the structure of all kinds of natural text. Similarly, Wolf and Gibson (2005) consider visualizations of discourse structures descriptively adequate when they can account for all possible interpretation structures.

1.1.1. *Descriptively adequate theories of coherence relations*

- 12 A descriptively adequate coherence theory helps characterize the internal structure of discourse – written and spoken texts – by plausible text structures (Taboada & Mann, 2006). Descriptively adequate relational categories function as useful tools to describe text structure (Sanders et al., 1993). They are descriptive constructs, developed by linguists based on the available texts. There are three facets of descriptive adequacy that a theory of coherence should adhere to. First, the theory must be able to describe all relations in the texts that it aims to analyze (Knott, 1996). This facet can be referred to as exhaustiveness of the relational taxonomy, or *observational completeness* (see also Nuyts, 1992). Second, the categories that are part of the theory should be distinguishable; that is, a theory should not propose relational categories of which no examples can be found in natural text. This facet can be referred to as discreteness of the taxonomy, or *observational adequacy* (see Nuyts, 1992). Finally, theories must distinguish only those relations that are relevant and necessary, and not more (e.g., a causal relational category entails the banana peel and slip relation) (see also Moeschler & Auchlin, 2005). This facet can be referred to as efficiency of the relational taxonomy, or *economic adequacy*.

1.1.2. *Evidence for descriptive adequacy of a theory of coherence relations*

- 13 Adhering to observational completeness means that theories should be able to describe all relations in the discourse that it aims to analyze. Using the theory successfully to conduct a full analysis of texts could therefore be considered as evidence for its descriptive adequacy. Underlying this type of evidence are two types of agreement. The first is agreement between multiple experts on whether the distinctions that the theory proposes correspond to the actual constructs in the language. This is not to say that every linguist should agree that the proposed relational categories exist, but at least more than one expert should agree on the categories in order for the theory to be considered descriptively adequate. The second is agreement between annotators when applying the theory to a body of text: successfully using a theory to analyze a text entails sufficient inter-annotator agreement. In other words, we should strive to devise annotation schemes that are not only sophisticated (i.e., descriptively adequate) enough to be useful, but also reliable enough to be replicated (Artstein & Poesio, 2008). Spooren and Degand (2010) note that low interrater agreement suggests that the categories used in a theory are non-replicable, and therefore unfit as a basis for theory building. It could therefore be argued that inter-annotator agreement is a source of evidence for the descriptive reliability of a theory. However, given that this type of evidence is based on annotator performance, we consider annotation reliability to be evidence for the *cognitive plausibility* of a theory instead. This will be elaborated on in Section 3.

1.2. Cognitive plausibility

- 14 In cognitive science, the term “cognitive plausibility” has various meanings. In cognitive modeling, it refers to being able to replicate the observed behavior of

individuals, and in AI, the term refers to the inputs, outputs, and processes of systems being comparable to those of humans (Kennedy, 2009; Smith & Hancox, 2001). In language acquisition, the cognitive plausibility of a computational model is considered higher when it is directly related to the actual acquisition task (Phillips & Pearl, 2015). These fields differ from each other and from the field of discourse coherence in what their object of investigation is, but what they have in common is that their theories need to explain human performance in order to be cognitively plausible. This covers a number of dimensions, including a description of the adult cognitive systems and a description of language acquisition and the development of the cognitive system (Nuyts, 1992).

1.2.1. *Cognitively plausible theories of coherence relations*

- 15 Coherence relations are considered to be cognitive entities that comprehenders infer when they create a mental representation of the discourse. Cognitively plausible theories of coherence must be able to describe discourse as it is encoded in human cognition; that is, it should distinguish relational categories that actually play a role in the construction of a cognitive representation of text (Sanders et al., 1992; see also Taboada & Mann, 2006).

1.2.2. *Evidence for the cognitive plausibility of a theory of coherence*

- 16 There are various sources of evidence that can validate the existence of relational categories. Evidence can come from empirical studies that show that the proposed relational categories have different effects on comprehenders' cognitive processes (e.g., in the areas of acquisition, comprehension, production). However, in the field of discourse coherence, elaborations on what exactly constitutes cognitively plausible sources of evidence have remained elusive. In this paper, the different types of evidence that can be used to validate a theory's cognitive plausibility will be discussed in more detail, but first, we discuss the relation between the two criteria.

1.3. Descriptively adequate and cognitively plausible?

- 17 Descriptively adequate theories often distinguish between more fine-grained relation types so that they can describe all relevant aspects of text structures. This is exemplified by very fine-grained distinctions found in many proposals, such as the level of detail in distinctions between various CONDITION relations (see Appendix for an illustration of how fine-grained CONDITION distinctions can be). By contrast, cognitively plausible theories distinguish between relation types that affect cognition, as evidenced by studies of acquisition, processing, etc. As a result, they often do not consider fine-grained distinctions, but rather focus on more global ones that might affect cognition, such as causals, contrastives, and additives.

- 18 Although descriptive adequacy and cognitive plausibility are not contradictory, in practice, coherence relation approaches often focus on one more than the other. The two most well-known frameworks, the Penn Discourse Treebank (PDTB, Prasad et al., 2007) and the Rhetorical Structure Theory Discourse Treebank

(RST-DT, Carlson et al., 2003), place more emphasis on descriptive adequacy, each in their own way¹. Other coherence relation approaches or theories focus more strongly on cognitive plausibility (e.g., Hobbs, 1979; Kehler, 2002; Knott & Dale, 1994; Sanders et al., 1992). For example, Sanders et al. (1992, 1993) proposed the Cognitive approach to Coherence Relations (CCR), aiming for cognitive plausibility. They explicitly did not aim at providing a “complete descriptively adequate taxonomy of coherence relations” (Sanders et al., 1992: 4). In a recent update to the taxonomy, however, they propose to include *DISJUNCTION* as a new distinction, based on arguments targeting both descriptive adequacy and cognitive plausibility (Hoek et al., 2019).

19 At this point, it is important to note the bias in the field towards describing written language, rather than (spontaneous) spoken language. In the few instances that these frameworks were applied to spoken language, serious adaptations had to be made. For example, Tonelli et al. (2010) added Speech-act relations to the PDTB taxonomy when they applied PDTB 2.0 to Italian spoken text, and Rehbein et al. (2016) noted the need for a *CONCLUSION* relation when they annotated English spoken text. These proposed adaptations could suggest that a focus on written language obscures a full understanding of discourse coherence. The insufficient treatment of spoken language in the area of discourse coherence is therefore considered problematic, given that a generalizable theory of coherence should be based on *text* in the broadest sense of the word – both written and spoken text. There have been important recent contributions that draw the community’s attention towards this matter (see, for example, Crible & Zufferey, 2015; Crible & Cuenca, 2017). Future studies will hopefully further address this bias, important as it is for both descriptive adequacy and cognitive plausibility.

20 Returning to the discussion of whether coherence relation approaches can be both descriptively adequate and cognitively plausible, it can be expected that approaches focus more on one criterion than the other, depending on what is relevant to their research goals. For example, a theory of discourse as a linguistic object will aim to describe the structure of the text, without taking into account the cognitive status of this structure. Such a theory can be useful for a variety of tasks, including automatic text summarization. The theory’s cognitive plausibility is irrelevant for text summarization since this usually does not relate directly to human processing tasks. The main focus is on extracting the most important information from the text. As a result, the theory can be suitable and sufficient for that particular goal,

1. The PDTB’s goal is to annotate relations in corpora. However, PDTB’s approach is related to cognitive plausibility as well: it makes distinctions based on the connectives or cue phrases found in the corpus. This approach is largely based on Knott’s (1996) work, which assumes that the existence of a cue phrase in a language can indicate the existence of a corresponding relational construct in cognition. The RST-DT is based on the original RST framework, which is intended to describe texts, rather than the processes of creating or reading and understanding them. However, its relational distinctions do relate to cognitive insights: relations are defined as constructs that are associated with the particular effects that a writer intends to achieve.

even though it does not account for the cognitive plausibility of the distinctions that it makes (but see Smith & Hancox, 2001).

- 21 However, when using descriptive adequacy as the only criterion for developing or validating a more general theory, it can be difficult to assess the appropriate level of detail (see also Knott, 1996). For example, as illustrated in Appendix, PDTB 2.0 distinguishes six subtypes of *CONDITION* relations and two more subtypes of *PRAGMATIC CONDITION* relations². By contrast, RST-DT distinguishes four types of *CONDITION* relations. Both theories seem feasible as a descriptive framework, and both have been used to annotate texts. They therefore meet the requirements for being considered descriptively adequate. Additional sources of evidence are thus needed to validate the distinctions. These sources speak to the cognitive plausibility of the distinctions.
- 22 Another issue with using descriptive adequacy as the sole requirement for theories concerns its explanatory power. Descriptively adequate approaches are useful tools to *describe* but not to *explain* discourse structure. Cognitively plausible approaches do have explanatory power, meaning they can make predictions about the object of investigation, and these predictions can be falsified in various fields of linguistics (see also Foltz, 2003; Knott, 1996; Sanders et al., 1993). Testing such empirical predictions can advance the field of discourse coherence.
- 23 For an empirically validated, generalizable theory of coherence relations, descriptive adequacy and cognitive plausibility are therefore both important: only theories that address both criteria can improve our understanding of the mental processes of discourse comprehension and production. Descriptively adequate approaches can function as a starting point for developing a general theory: they can provide an inventory of possible relational constructs, which researchers can then validate using cognitive evidence. Such cognitively plausible evidence for relational distinctions provides validation that these theories model the actual constructs that are present in our linguistic system (see also Scholman et al., 2016; Smith & Hancox, 2001; Spooren & Degand, 2010; for similar pleas to ensure cognitive plausibility of coherence theories and annotation frameworks).
- 24 Descriptive adequacy and cognitive plausibility are both necessary and sufficient for describing coherence relations, but they have a different status in the field: descriptive adequacy is defined clearly and most approaches follow this naturally, whereas cognitive plausibility is less well-defined and less often adhered to. The remainder of this paper therefore focuses on defining the cognitive plausibility criterion more clearly. In the next section, the criterion of cognitive plausibility is made tangible by outlining possible sources of evidence that can verify or falsify the cognitive plausibility of distinctions between labels or classes.

2. Relational labels that are written in SMALL CAPS refer to specific framework labels. Labels written in normal font refer to general relation types, irrespective of a specific framework.

2. Criteria for cognitive plausibility

- 25 The literature generally states that cognitively plausible theories have to be responsive to empirical results from cognitive research, thereby also providing support for the framework's structure and assumptions. However, the specific types of empirical results or types of cognitive research are not elaborated on. We list possible sources of evidence and provide a case study of evaluating certain distinctions using the available evidence.

2.1. Possible sources of evidence

- 26 The possible sources of evidence that are nominated are production, representation and comprehension, processing, acquisition, linguistic evidence, annotation, and crosslinguistic evidence. These sources operate on different levels: the first four sources speak to psycholinguistic evidence, the latter three to evidence on the linguistic level. Some give more direct insight into the plausibility of relational distinctions (such as processing studies), others provide indirect insight into distinctions by studying connectives (such as linguistic evidence or age of acquisition studies). Some of the sources in this list prescribe methodologies (e.g., production, representation), others build on these sources by using the same methodologies but looking at different age groups (acquisition) or languages (crosslinguistic evidence). Moreover, some sources allow for hypothesis testing in order to try to falsify the distinctions (e.g., processing studies), whereas others can indicate possible issues with distinctions, but are less suitable for falsifying or verifying hypotheses (e.g., annotation studies).

- 27 The list of possible methodologies proposed for each source is not exhaustive; it is merely meant to give an indication. The methods that are discussed are exemplary for the methods that are commonly used in discourse coherence research, and an example is given for many methods to illustrate the phenomena that can be investigated. These examples merely function as pointers for those who are interested; discussing every example in detail is outside of the scope of this paper.

2.1.1. Production

- 28 Production data consist of written and spoken discourse. We distinguish between corpus data and elicited data/experimental studies. Corpus data are strongly related to descriptive adequacy: categories are distinguished based on the identification of corresponding relational constructs in written or spoken texts. However, it can also be used to provide cognitively plausible evidence: researchers can use this data to test hypotheses based on distributional data. For example, Andersson and Spenader (2014) investigate the occurrences of RESULT relations (Example [3]) and PURPOSE relations (Example [4]) with and without their typical marker *so* in written corpora.

[3] Mary often borrows her sister's clothes, so her sister gets upset with her.

[4] Mary often borrows her sister's clothes, so she can look younger and more attractive.

29 They find that PURPOSE relations always occur with *so*, whereas RESULT relations are often left unmarked. This distributional variability indicates that there is a significant difference between the two types of relations. The data is therefore taken to support the distinction between PURPOSE and RESULT relations³.

30 Examples of elicited production studies are item continuation and connective insertion studies. In item continuation experiments, participants are presented with prompts and asked to continue the prompt in a natural manner. This type of study is useful to test whether readers are more likely to infer a specific type of relation depending on different contexts. For example, Scholman et al. (2020) presented participants with items varying in the presence of a list signal (e.g., *several*, *multiple*) in the context, and asked them to continue the story. The results showed that comprehenders were sensitive to such a list signal (although differences between individuals occurred): participants provided a continuation reflecting a list coherence relation only when the context contained the list signal. This supports the distinction of a distinct list relation.

31 In connective insertion experiments, participants are presented with the segments of a relation and asked to insert a connective. Sanders et al. (1992) use such a task to investigate whether readers are able to infer the coherence relations between sentences. The results showed that participants were able to choose the connective corresponding to the original connective relatively well. Most disagreements between the inserted connective and original connective occurred for subjective relations. A relation is classified as objective if both segments describe situations in the real world; relations are subjective if speakers are actively engaged in the construction of the relation (compare objective cause-consequence to subjective claim-argument). Sanders et al.'s (1992) finding that participants could not accurately distinguish between these objective and subjective relations indicates that the distinction warrants more investigation. Issues related to the absence of evidence will be discussed in Section 3.

2.1.2. Representation and comprehension (offline)

32 Offline measures provide insight into the representation of the discourse after the comprehension process is completed⁴. Methodologies tapping into these representations include recall, comprehension questions, and judgment tasks (e.g., acceptability or plausibility ratings). Myers et al. (1987) studied participants' recall to investigate whether causal links affect the mental representation of the discourse.

3. It should be kept in mind that production data has a circular nature: distributional data of relations annotated in corpora are based on a particular analysis in terms of descriptive categories. Consequently, they do not provide independent evidence for or against the analysis.

4. Representation and comprehension can also be considered as being related to descriptive adequacy, since they relate to the linguistic intuitions and interpretations of readers/speakers. However, we consider studies that test such intuitions and interpretations to provide evidence for the cognitive plausibility of distinctions, since they speak to the *mental representation* of the discourse (Sanders et al., 1992).

The results showed that causally related sentences such as [5] were recalled better than non-causally related sentences such as [6], indicating that causally related information is stored better.

[5] Cathy felt very dizzy and fainted at her work. She was carried unconscious to a hospital.

[6] Cathy had worked overtime to finish a project. She was carried unconscious to a hospital.

33 Murray (1997) used a judgment task to investigate whether adversative connectives such as *however* contribute differently to the discourse than causal and additive connectives (e.g., *consequently* or *moreover*, respectively). They asked participants to rate on a five-point scale whether sentences “followed sensibly” from the preceding sentence. The results showed that sentences with incorrectly placed adversative connectives led to greater disruption in the perceived coherence than incorrectly placed additive and causal connectives.

34 Another method providing insight into the interpretation of relational types is the sorting task (Miller, 1969), for which participants sort relations based on their similarity to each other. Sanders et al. (1993) found that participants did not create separate clusters for *objective* and *subjective* relations⁵, in particular within the set of contrastive relations. Similar to the evidence discussed in the previous subsection, this speaks against the cognitive plausibility of this specific distinction.

2.1.3. Processing (online)

35 Various paradigms can be used to capture the online comprehension processes, the most popular ones being the visual world paradigm, self-paced reading, and eye-tracking-while-reading. For example, Wei et al. (2019) used the visual world paradigm to investigate whether comprehenders are sensitive to the subjectivity profile of Dutch and Chinese connectives⁶. Participants listened to causal relations marked by a subjective or objective connective while looking at two images: one depicted an event, the other depicted a person with a speech bubble containing a picture of the event. The results showed that participants looked at the image of the person more often after hearing a subjective connective than an objective connective. This indicates that the subjectivity of a connective has an immediate influence on readers’ expectations of the relation. The results of this study therefore support the cognitive plausibility of the distinction between objective and subjective relations, contrary to the results discussed in previous sections. Issues related to the relation between evidence and counter-evidence will be discussed in the next section.

5. In Sanders et al. (1993), this distinction is referred to as *semantic* and *pragmatic*.

6. In Dutch, the causal connective *omdat* (“because”) is typically used to mark objective relations, whereas the causal connective *want* (also “because”) is typically used to mark subjective relations.

36 Crible (2021) used a self-paced reading design to investigate the effect of negation on the processing of CONCESSION and RESULT. The results showed that CONCESSION is more difficult to process than RESULT, confirming findings from earlier studies (e.g., Bloom et al., 1980; Morera et al., 2017; Townsend, 1983) and further providing support for the distinction between positive causal and negative causal relations. Interestingly, this difference disappeared when the verb of the first relational argument was negative, which shows that an alternative discourse cue such as negation can modulate discourse-level processing and thereby impact result and concession differently.

37 Measuring reading times or the proportion of looks are a rather natural way of measuring comprehension: the interpretation process is not disturbed by the experimental task. However, reading times can sometimes be ambiguous regarding the types of processes they index; for example, longer reading times do not tell us what type of difficulty causes it. In a visual world study, Köhne-Fuetterer et al. (2021) investigated the processing of causal and concessive connectives (*therefore* and *nevertheless*, as illustrated in Example [7]).

- [7] Tim and Kim wonder whether they prefer to go dancing or to watch the new movie.
Kim likes dancing a lot.
- a. *Therefore*, they go to the newly renovated *disco* to enjoy themselves.
 - b. *Nevertheless*, they go to the newly renovated *cinema* to enjoy themselves.

38 They found that readers were able to quickly process and integrate causal but not concessive connectives in the discourse. These results may be due to an increased difficulty of processing concessives, or of integrating them with upcoming content. Neurological measures such as ERP (event-related potential) provide a clearer view of what processes occur during reading. Köhne-Fuetterer et al. (2021) therefore complemented the visual world results with two ERP studies. The data showed that, when readers encountered a concessive connective instead of a causal connective, they updated their mental representation from an expected causal relation (as in Example [7a]) to an unexpected concessive relation (as in Example [7b]). This specific updating process supports the explanation of the visual world study that their results are due to an increased difficulty of processing concession compared to causality.

2.1.4. Acquisition

39 Children acquire connectives and coherence relations in a relatively fixed order (e.g., Bloom et al., 1980; Evers-Vermeul & Sanders, 2009). The system that underlies this order is expected to correspond to a cognitively plausible categorization of relations (Sanders et al., 1992). A number of different paradigms have been used to study child language acquisition. Perhaps the simplest method for studying the acquisition of discourse skills is by recording children's spontaneous speech. Such recordings can be found in corpora of longitudinal child language data, such as the CHILDES (Child Language Data Exchange System) corpus (MacWhinney, 2014). These corpora have provided a fruitful source of production evidence.

40 Production methods such as elicited-production experiments also provide researchers with a method to investigate discourse coherence skills in children. Generally, the task consists of asking a child to respond to some sort of question or stimulus. Evers-Vermeul and Sanders (2011) used a directive task to investigate different domains of causality in children's speech: children were asked to instruct a puppet where to put stickers and to motivate their instruction. The results showed that children aged 2 were able to use causal connectives in an objective context, but that the subjective domain was acquired later⁷.

41 Another source of acquisition evidence comes from comprehension and processing methods, such as act-out, picture-pointing, or visual-world studies. Cain and Nash (2011) studied young readers' knowledge of temporal, causal, and adversative connectives using a cloze task, coherence judgment task and online reading task and found that children were able to understand the meaning of all three types of connectives. Knoepke et al. (2017) also used a coherence judgment task to investigate children's comprehension of causal sentences (Example [8]) and adversative (or negative causal) sentences (Example [9]).

[8] Lena stayed in the sun for too long. Therefore, she got a sunburn.

[9] Sandra was not tired. Nevertheless, she went to bed.

42 The results showed that both children aged 6 to 10 and adults were more accurate in judging the coherence of causal relations compared to adversative relations. The study therefore provides evidence that causal relations are less complex than adversative relations.

2.1.5. Linguistic evidence

43 Linguistic evidence is related to descriptive adequacy: it is possible, and even likely, that researchers make use of the available linguistic elements in their corpus to develop an inventory of relation types, as PDTB has done. Nevertheless, linguistic evidence is considered a cognitively plausible source, in accordance with Knott and Dale (1994: 44), who state that "if people actually use a particular set of relations when constructing and interpreting text, it is likely that the language they speak contains the resources to signal those particular relations explicitly". The marking of relations can therefore provide insight into the cognitive constructs of coherence relation types that comprehenders can compose and comprehend. This source of evidence is different from production evidence, which studies distributions and occurrences in a dataset. Linguistic evidence refers to the availability of various linguistic elements in the language system.

7. Evers-Vermeul and Sanders (2011) used the term *content* for the objective domain and *epistemic* and *speech act* for the subjective domain.

- 44 Relations can be expressed or marked in different ways, leading to different sources of linguistic evidence. The most commonly considered and explicit source is the set of connectives (e.g., *because*, *however*) and cue phrases (e.g., *for this reason*, *to sum up*, also referred to as secondary connectives, Rysová & Rysová, 2018). Connective inventories have been developed for various languages, including French (Roze et al., 2012), German (Stede & Umbach, 1998), Portuguese (Mendes et al., 2018), and Czech (Mírovský et al., 2017), among others (see also Stede et al., 2019).
- 45 Knott and Dale (1994) used substitutability tests (which tap into people's intuitions about whether one connective or cue phrase can replace another in a given context) for English connectives to create a framework of coherence relations (see also Knott & Sanders, 1998). But connectives and cue phrases can also be used to validate existing theories: should a framework distinguish several subtypes of a relation, which are all typically marked by the same connective or cue phrase with the same distributional properties (e.g., different types of CONDITION relations all marked typically and consistently by *if*), this could indicate that the subtypes are not cognitively plausible (see, e.g., Zufferey & Degand, 2017). Similarly, if a relational category is typically marked by a variety of connectives and cue phrases, the category might be too coarse-grained and could be further distinguished into subcategories (as noted elsewhere, e.g., Degand, 1998). Yung et al. (2021) used substitutability tests to study whether the relational distinctions in PDTB 3.0 are fine-grained enough. Their results showed that the relation inventory would need to be extended in order to capture temporal aspects in relations marked by *before* vs. *until*: both are classified as the same type of relation (SUCCESSION) but the results clearly showed that participants did not consider these connectives to be interchangeable. This is taken as evidence that their classification is too coarse-grained (but see Rysová & Rysová, 2018).
- 46 Connectives and cue phrases are generally considered the strongest markers of coherence relations. However, the encoding of a relation can take place on different levels; not only on the connective level (Stukker & Sanders, 2012). It is possible that certain relations are in fact encoded differently (e.g., in syntactic features). For example, relating to the distinction of objective vs. subjective relations, Levshina and Degand (2017) show that a set of semantic and syntactic features such as modality, tense (past vs. non-past) and the presence of evaluative adjectives, are reliable predictors of the more subjective and objective uses of *because*. This demonstrates that the subjectivity distinction can indeed be anchored in the immediate linguistic context rather than in the connective itself. A line of research that has recently gained more attention is uncovering non-connective cues that signal a large variety of relations; for example, studies have shown that verb tense is a marker for temporal relations (Grisot, 2018; Grisot & Blochowiak, 2021), negation for concession relations (Crible, 2021), negation and modals for CHOSEN ALTERNATIVE relations (Asr & Demberg, 2015; Webber, 2013), complementizers for consequence (Rohde et al., 2017), and implicit causality verbs for causal relations (Bott & Solstad, 2014; Kehler et al., 2008). The most comprehensive collection of signals is the RST-DT Signaling Corpus (Das

& Taboada, 2018), in which coherence relations have been annotated for all possible signals: reference, semantic, lexical, syntactic, and graphical features (see also Hoek et al., 2018). Future studies will hopefully provide more insight into the link between these markers and specific relational categories.

2.1.6. Performance of annotation

47 In order to study the distribution and linguistic realization of coherence relations, researchers use discourse-annotated corpora. These corpora are commonly manually annotated by trained, expert annotators. Performance of annotation could provide interesting, albeit indirect, insights into the plausibility of categorizations. This type of evidence is related to the distinction between *competence* (what language users can do in principle) and *performance* (what they actually do). After all, relational categories might be distinguishable in theory, but they also need to be able to be reliably distinguished in practice⁸.

48 One source of annotation evidence is how easily items can be annotated according to this distinction. Taboada and Mann (2006) refer to this as *observability*; the possibility of distinguishing one relation from another. Crible and Degand (2019b) evaluated inter-annotator agreement between coders using Crible and Degand's (2019a) taxonomy for spoken discourse markers to identify distinctions that might be superfluous or need further operationalization. They conducted a quantitative and qualitative analysis of inter-annotator (dis)agreements, which provided the grounds for a number of revisions to the original taxonomy. Zufferey and Degand (2017) set out to empirically validate the PDTB annotation framework by using it in a crosslinguistic annotation study, thereby proposing a number of revisions to the original framework. Similarly, PDTB 3.0 (Webber et al., 2019) made changes to the inventory of PDTB 2.0 based on annotation evidence. For example, annotation practice showed that it is hard to distinguish reliably between CONTRAST.OPPOSITION and CONTRAST.JUXTAPOSITION, illustrated in [10] and [11].

[10] Private construction spending was down but government building activity was up.

[11] Sales were roughly flat in the 1989 model year compared with a year earlier though industry sales fell.

49 The PDTB describes JUXTAPOSITION relations as presenting alternatives, whereas OPPOSITION relations present extremes of a gradable scale. As a result of the low inter-annotator agreement for these relations, PDTB 3.0 does not distinguish between these subtypes anymore.

50 In sum, the annotation process (the ease with which a distinction can be made) and outcome (amount of agreement between annotators) can be used as evidence that

8. This type of evidence could be considered to relate to the descriptive adequacy of a theory, because it is based on annotated data. However, we consider it evidence regarding cognitive plausibility, since it speaks to performance of annotators.

specific categorizations are not entirely valid or reliable. It is, however, important to keep in mind that achieving good interrater agreement does not necessarily ensure validity: two annotators may agree on a classification due to annotation biases or prejudices, while that classification may be wrong (Artstein & Poesio, 2008).

2.1.7. *Crosslinguistic evidence*

51 Finally, studies investigating discourse coherence in multiple languages can provide valuable insight into the plausibility of a category, given that crosslinguistic evidence increases the confidence we can have in our findings (Klavan & Divjak, 2016). First, however, we should consider a more fundamental question: are relational categories necessarily universal and language-independent? Although general classes of relations are argued to hold across languages, fine-grained relational distinctions seem to differ between languages. Consider the PDTB, which has been applied to different languages, such as Chinese, Czech, Hindi, Arabic, Italian, Turkish, and French. Some of these efforts have made changes to the original relational inventory by merging certain labels that are difficult to distinguish and proposing new labels (e.g., Al-Saif & Markert, 2010; Zhou & Xue, 2015; Zufferey & Degand, 2017).

52 The adaptations of frameworks to different languages can have two reasons. First, it is possible that certain fine-grained relational distinctions identified for one language might not be relevant in other languages⁹. This would mean that relational categories can be language-specific, in which case more research is needed to fully understand which relational distinctions are affected, and why they are found to be relevant in one language but not another. Second, differences can be caused by varying expert intuitions, rather than by actual linguistic differences. This implies that relational categories are not language-specific, and crosslinguistic evidence can be used to identify which distinctions warrant more investigation.

53 Future empirical crosslinguistic studies might provide more insight into this issue. Most studies to date have investigated the meanings and functions of discourse connectives in various languages (Gonzalez, 2005; Zufferey & Degand, 2017), translations of connectives and relations between connective lexicons for different languages (e.g., Bourgonje et al., 2017; Hoek & Zufferey, 2015; Knott & Sanders, 1998), and annotations of multilingual parallel corpora (e.g., Cunha & Iruskieta, 2010; Zeyrek et al., 2020; Zufferey & Degand, 2017). Less is known about different relational distinctions and the effects of relational distributions on processing and acquisition in different languages (but see Köhne-Fuetterer et al., 2021; Mak et al., 2013; Sun & Zhang, 2018; Zufferey et al., 2018). Studying crosslinguistic differences can provide insight into the feasibility of a universal cognitive theory of coherence relations.

9. Consider findings from colour research: some languages distinguish between certain colour concepts that other languages do not distinguish (see, e.g., Kay & Regier, 2006; Roberson et al., 2005). Of course, colour research is not located in the same domain as discourse coherence research, but it illustrates that languages and cultures do not necessarily conceptualize the same distinctions.

54 For now, we consider crosslinguistic evidence to have a reduced status: it can provide possible additional support for the cognitive plausibility of a distinction, but it is not considered to be absolutely necessary support. It can therefore not provide counter-evidence for the cognitive plausibility of a distinction. The reason is that language is shaped by experience and language shapes cognition, and so crosslinguistic differences may occur. As a result, striving for universal cognitive theories might even be counter-productive: a theory that can be applied to all languages might not be complete in any language, therefore rendering an insufficient account of discourse coherence in general. Indeed, Zufferey and Degand (2017) note that a taxonomy designed for multilingual purposes cannot aim for a total coverage of specificities in every language, due to the unique features of individual languages.

55 In sum, several sources of evidence are nominated for justifying relational distinctions based on the criterion of cognitive plausibility. Of course, these different sources do not provide the same type of insights. However, the different insights that can be gained using these methods are exactly what makes the combination of evidence stemming from various methodologies informative regarding the plausibility of distinctions.

2.2. Applying the criteria: a case study

Source	Evidence
Production	✓/✗
Representation/Comprehension	✓
Processing	✓
Acquisition	NA
Linguistic system	✓
Annotation	✗
Crosslinguistic system	✓

Table 1 – An overview of the evidence from various sources regarding the cognitive plausibility of CONTRAST and CONCESSION relations.
 ✓: evidence available; ✗: counter-evidence available; NA: no evidence available

56 We now turn to an illustration of how coherence relation distinctions can be evaluated using the further defined criterion of cognitive plausibility. Table 1 presents an overview of the evidence that has been found to support (or contradict) the cognitive plausibility of adversative relations, in particular the distinction between CONTRAST (i.e., semantic opposition) and CONCESSION (i.e., negative causal) relations. CONTRAST relations are characterized by relating two segments containing information, as illustrated in Example [12]. CONCESSION relations are characterized by one segment creating an expectation and another segment denying this expectation (Robaldo & Miltsakaki, 2014; Prasad et al., 2007), as in [13].

[12] Noah likes to go swimming, whereas Toby prefers to play at the playground.

[13] We wanted to go to the zoo, but it was raining.

57 The CONTRAST-CONCESSION distinction is chosen because both evidence and counter-evidence have been found, and some methodologies have not yet been applied to investigate the plausibility, as shown in Table 1.

- *Evidence*: Several studies have identified or investigated unique linguistic markers for these relations in various languages (for example, *whereas* and *by contrast* are typical English contrastive connectives, and *even though* and *nevertheless* are typical concessive connectives; see, e.g., Knott, 1996; Prasad et al., 2007). Moreover, comprehenders have been found to be able to distinguish between CONTRAST and CONCESSION relatively well (Sanders et al., 1992)¹⁰. Finally, readers show a processing difference between CONTRAST and CONCESSION relations (Asr & Demberg, 2020).
- *Counter-evidence*: Expert annotators have consistently found that agreement on the classification of relations in these two categories is difficult to reach (Webber et al., 2019; Zufferey & Degand, 2017, see also Demberg et al., 2019). This could indicate that the distinction is not a functional one. Moreover, a connective insertion study revealed that comprehenders are able to distinguish CONCESSION relations relatively well, but not CONTRAST relations (Scholman & Demberg, 2017). This indicates particular difficulty with the CONTRAST distinction.
- *No evidence available*: A literature review revealed no sources on the acquisition of this distinction. Future studies will hopefully address this gap in order to provide more support for the justification of the distinction.

58 The small case study raises several issues regarding what counts as enough evidence or counter-evidence to validate the relational class.

3. Verification and falsification

59 One goal of using the more clearly defined criterion of cognitive plausibility is to be able to validate proposed relational distinctions and theories as a whole. This goal presupposes that those distinctions can be falsified. The current section explores the notion of falsification.

60 Theories that aim to explain language as it is represented in the mind can be held accountable for the hypotheses they generate. Accountability relates to two principles: reliability and validity. *Reliability* is a measure of stability or consistency: can experts agree on a description of a linguistic object using the theory? And do

10. The notion of CONTRAST used in this paper is similar to Sanders et al.'s (1992) negative additive relations and CONCESSION is similar to negative causal relations.

experts agree with their own earlier annotations? To determine the *validity* of a theory, one can ask: how accurately does the theory represent the construct? Testing the validity of a theory is related to falsification: researchers should not look for evidence that supports or confirms the theory, but rather look for evidence that the theory is not valid (Popper, 1959). This means that it should be possible to refute the hypotheses that can be developed based on theories. In the next subsections, we focus on issues related to falsification and verification.

3.1. Falsification: what counts as evidence that a distinction is not valid?

61 When is a distinction falsified? One answer would be to adhere to a checklist: a lack of evidence in studies using one of the methodologies indicates that the distinction is not valid. In other words, if one were to commit fully to cognitive plausibility, an ideal framework would be one that only includes relational categories for which cognitive evidence has been found in all sources, and no counter-evidence has been found in any of the sources.

62 However, this approach runs certain risks. Non-significant or null results can always occur, even when examining true effects (Harms & Lakens, 2018). The absence of evidence is therefore not necessarily the same as evidence of absence. If studies have failed to produce evidence supporting a particular type of relation, this could be due to many different factors. Of course, it could be that this particular category is not cognitively plausible. But it could also be due to power issues or methodological incompatibility. A strict falsification approach therefore runs the risk of a Type II error, and thus we consider a strict falsification approach to be counter-productive for theories of coherence relations. Rather, we argue that verification can play a bigger role in validating coherence theories, as will be discussed in the next subsection.

63 Nevertheless, we need a more defined idea of what counts as counter-evidence. We argue that null results from one source can be considered as counter-evidence for the cognitive plausibility of the distinction. But this alone is not enough to count as evidence that a distinction is not cognitively plausible. The distinction should still be falsified by evidence from multiple sources, using different methodologies and materials. Crucially, we propose that failure to find any effect in all four psycholinguistic domains (production, representation, processing and acquisition) is sufficient evidence for non-validity of the distinction.

3.2. Verification: what counts as enough evidence?

64 Ideally, the distinctions made by theories are validated by available evidence from all sources. However, it is likely that a theory cannot account for all behavioral phenomena within its range. To address this, Nuyts (1992) suggests speaking of the *probability* of a theory being cognitively plausible: the more data that a theory accounts for, the higher the probability that it is an adequate characterization of the phenomenon under investigation. From this point of view, any evidence of a distinction being cognitively plausible is a good start, but there really is no way to say that there is enough evidence. Perhaps “sufficient” is a better term.

65 What, then, counts as sufficient evidence? Support for a particular distinction should come from multiple sources, ideally at least from each major psycholinguistic modality. Evidence from only a single source can be considered indicative of the existence of the category, and will need to be supplemented with evidence from other sources, using different methodologies, participants and materials. This approach emphasizes the importance of replicability and can rule out the risk of false positives.

66 This idea is consistent with other proposals arguing in favor of providing converging evidence from multiple methods before considering a theory valid. Arguing for a general theory of inferences, Magliano and Graesser (1991) advocate a “three-pronged method” that coordinates (i) predictions generated by theories or hypotheses, (ii) data from think aloud protocols, and (iii) behavioral measures that assess processing time. In a similar vein, but recommending different methodologies, Sanders and Evers-Vermeul (2019) advocate a combination of (i) corpus studies on language use, (ii) experimental studies on discourse processing and representation, and (iii) corpus-based and experimental studies on language acquisition. They argue for this combination of converging methodologies because each methodology has its own merits and drawbacks, and they can therefore complement each other. Similar approaches have been taken in work that evaluates or motivates distinctions (e.g., Evers-Vermeul et al., 2017; Hoek et al., 2019; Zufferey & Popescu-Belis, 2017).

67 A remaining issue concerns those distinctions for which both evidence and counter-evidence are found, such as objective vs. subjective relations or the distinction between CONTRAST and CONCESSION. Perhaps the issue is with the descriptive adequacy of such distinctions. For example, the distinction between CONTRAST and CONCESSION might not be optimally categorized yet. Theories have provided different ways to distinguish these relations from each other (Izutsu, 2008; Prasad et al., 2007; Carlson et al., 2003; Pander Maat, 1998), and they have all provided linguistic examples as validation for their categorizations. Robaldo and Miltsakaki (2014) argued that previous classifications do not account for the full range of variants of CONCESSION relations, and they proposed a different categorization based on the source of expectation (e.g., *correlation* and *implicature* are two different sources). They then report a significant improvement on the annotation of previous disagreements on CONCESSION-CONTRAST relations in the PDTB. This indicates that the new classification might be a more plausible one, although this, of course, warrants more investigation. Distinctions for which there is conflicting evidence might require more investigation to determine whether the classification is the most optimal one. This is the type of debate that the community would benefit from, and that could advance the field of discourse coherence.

4. Conclusion

68 We have discussed different criteria to validate theories of discourse coherence: descriptive adequacy and cognitive plausibility. Both criteria are based on the notion that distinctions made in theories should correspond to and account for linguistic data. Theories and the relational categories that they distinguish can therefore be both

descriptively adequate and cognitively plausible. However, in practice, many frameworks – as well as their corresponding corpora – have a focus on descriptive adequacy. Their relational inventories differ from each other in the number and types of relations, the labels, and the underlying definitions of labels. This severely limits the interoperability of frameworks and data resources available in the field (Sanders et al., 2018).

69 Crucially, we have argued that a text-analytic model that aims to have explanatory power should adhere to both criteria. Only theories that address both criteria can improve our understanding of the mental processes of discourse comprehension and production. Cognitively plausible coherence relation constructs are at the basis of linguistic phenomena, and they are the cause of the observations made in descriptively adequate approaches. Descriptively adequate approaches that capture those phenomena are more generalizable. In other words, distinctions that are both descriptively adequate and cognitively plausible should be more durable and relevant, because they are grounded in how linguistic processes actually work.

70 The current contribution emphasized the need for converging evidence in order to verify our theories of coherence. Of course, obtaining evidence for all relational distinctions using all suggested sources is difficult to achieve, if only for the sheer number of resources needed for such an undertaking. We do not claim that theories should not be used or trusted without providing cognitively plausible evidence using all of these sources. Rather, the goal is to start a discussion on issues related to validation and to work towards validating more distinctions. Certain distinctions, mainly on a coarse-grained level, have been verified in numerous studies. For example, a large body of literature has shown that comprehenders differ in their acquisition, representation and processing of additive vs. causal relations (see Bloom et al., 1980; Knott & Dale, 1994; Sanders et al., 1992) and positive vs. adversative relations (see Köhne-Fuetterer et al., 2021; Knoepke et al., 2017; Evers-Vermeul & Sanders, 2009). Nevertheless, many other distinctions have not yet been validated, and we see this as a fruitful direction for future studies. This holds in particular for more fine-grained distinctions within classes, for example between various types of additive, negative, and temporal relations.

71 Several points remain open for discussion. First, how should we view distinctions with conflicting evidence (i.e., evidence and counter-evidence)? Our current proposal is to reconsider their descriptive adequacy: distinctions such as CONTRAST vs. CONCESSION might require further investigation to determine whether the current classification is the most optimal one. The question is whether this solution works for all distinctions with evidence and counter-evidence.

72 Second, the issue of crosslinguistic evidence deserves more consideration in the community. We have argued that a crosslinguistic, universal theory of discourse coherence is not desirable given the research that is currently available. We question whether it is feasible and valid to expect all relation types to be (equally) cognitively plausible in all languages, and rather expect differences in the types of fine-grained relation types that can be distinguished in various languages. This is evidenced,

for example, by the presence of connectives that make fine-grained distinctions in only some languages – consider subjective causal *want* and objective causal *omdat* in Dutch, vs. the general causal connective *because* in English. Additional research is needed to provide insight into the crosslinguistic generalizability of relational categories. Crosslinguistic evidence can therefore be used to further strengthen the support for a distinction, but should not be considered absolutely necessary support until such research is available.

73 In sum, we have argued that generalizable theories of coherence relations should be both descriptively adequate and cognitively plausible, and we have concretized the criterion of cognitive plausibility. We hope this proposal will contribute to future research efforts aimed at validating relational distinctions, as well as spark a debate about the evaluation of distinctions and the development of a comprehensive theory of coherence. This can lead the field to a deeper understanding of the link between the structure of a text as a linguistic object, its cognitive representations and the processes of text production and understanding.

References

- AL-SAIF, A. & MARKERT, K. 2010. The Leeds Arabic Discourse Treebank: Annotating Discourse Connectives for Arabic. In N. CALZOLARI, K. CHOUKRI, B. MAEGAARD, J. MARIANI, J. ODIJK, S. PIPERIDIS, M. ROSNER & D. TAPIAS (eds.), *Proceedings of the 7th International Conference on Language Resources and Evaluation – LREC 2010 (May 2010, Valletta, Malta)*. Paris: European Language Resources Association (ELRA): 2046-2053. Available online: http://www.lrec-conf.org/proceedings/lrec2010/pdf/479_Paper.pdf.
- ANDERSSON, M. & SPENADER, J. 2014. RESULT and PURPOSE Relations with and without “So”. *Lingua* 148: 1-27.
- ARTSTEIN, R. & POESIO, M. 2008. Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics* 34 (4): 555-596.
- ASR, F.T. & DEMBERG, V. 2015. Uniform Information Density at the Level of Discourse Relations: Negation Markers and Discourse Connective Omission. In M. PURVER, M. SADRADEH & M. STONE (eds.), *Proceedings of the 11th International Conference on Computational Semantics (15-17 April 2015, London, UK)*. Stroudsburg: Association for Computational Linguistics (ACL): 118-128. Available online: <https://aclanthology.org/W15-0117.pdf>.
- ASR, F.T. & DEMBERG, V. 2020. Interpretation of Discourse Connectives Is Probabilistic: Evidence from the Study of *But* and *Although*. *Discourse Processes* 57 (4): 376-399.
- BLOOM, L., LAHEY, M., HOOD, L., LIFTER, K. & FIESS, K. 1980. Complex Sentences: Acquisition of Syntactic Connectives and the Semantic Relations They Encode. *Journal of Child Language* 7 (2): 235-261.
- BOTT, O. & SOLSTAD, T. 2014. From Verbs to Discourse: A Novel Account of Implicit Causality. In B. HEMFORTH, B. MERTINS & C. FABRICIUS-HANSEN (eds.), *Psycholinguistic Approaches to Meaning and Understanding across Languages*. Cham: Springer: 213-251.

- BOURGONJE, P., GRISHINA, Y. & STEDE, M. 2017. Toward a Bilingual Lexical Database on Connectives: Exploiting a German/Italian Parallel Corpus. In R. BASILI, M. NISSIM & G. SATTA (eds.), *Proceedings of the 4th Italian Conference on Computational Linguistics – CLiC-it 2017 (Rome, Italy, 11-13 December 2017)*. CEUR Workshop Proceedings 2006. 1-6. Available online: <http://ceur-ws.org/Vol-2006/paper006.pdf>.
- BREWER, W.F. & LICHTENSTEIN, E.H. 1982. Stories Are to Entertain: A Structural-Affect Theory of Stories. *Journal of Pragmatics* 6 (5-6): 473-486.
- BUNT, H. & PRASAD, R. 2016. ISO DR-Core (ISO 24617-8): Core Concepts for the Annotation of Discourse Relations. In H. BUNT, V.V. PETUKHOVA, A. MALCHANAU, A. CHENGYU FANG & K. WIJNHOFEN (eds.), *Proceedings of the 12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation – ISA-12 (28 May 2016, Portorož, Slovenia)*. 45-54. Available online: <https://sigsem.uvt.nl/isa12/ISA12Proceedings.pdf>.
- CAIN, K. & NASH, H.M. 2011. The Influence of Connectives on Young Readers' Processing and Comprehension of Text. *Journal of Educational Psychology* 103 (2): 429-441.
- CARLSON, L. & MARCU, D. 2001. *Discourse Tagging Reference Manual*. ISI Technical Report ISI-TR-545. Marina del Rey: University of Southern California Information Sciences Institute. 87 p. Available online: <https://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf>.
- CARLSON, L., MARCU, D. & OKUROWSKI, M.E. 2003. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In J. VAN KUPPEVELT & R.W. SMITH (eds.), *Current and New Directions in Discourse and Dialogue*. Dordrecht – London: Kluwer Academic Publishers: 85-112.
- CHOMSKY, N. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- CRIBLE, L. 2021. Negation Cancels Discourse-Level Processing Differences: Evidence from Reading Times in Concession and Result Relations. *Journal of Psycholinguistic Research* 50 (6): 1283-1308.
- CRIBLE, L. & CUENCA, M.J. 2017. Discourse Markers in Speech: Characteristics and Challenges for Corpus Annotation. *Dialogue and Discourse* 8 (2): 149-166.
- CRIBLE, L. & DEGAND, L. 2019a. Domains and Functions: A Two-Dimensional Account of Discourse Markers. *Discours* 24: 1-35. Available online: <https://journals.openedition.org/discours/9997>.
- CRIBLE, L. & DEGAND, L. 2019b. Reliability vs. Granularity in Discourse Annotation: What Is the Trade-off? *Corpus Linguistics and Linguistic Theory* 15 (1): 71-99.
- CRIBLE, L. & ZUFFEREY, S. 2015. Using a Unified Taxonomy to Annotate Discourse Markers in Speech and Writing. In H. BUNT (ed.), *Proceedings of the 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation – ISA-11 (14 April 2015, London)*. Stroudsburg: Association for Computational Linguistics (ACL): 14-22. Available online: <https://aclanthology.org/W15-0202.pdf>.
- CUNHA, I. DA & IRUSKIETA, M. 2010. Comparing Rhetorical Structures in Different Languages: The Influence of Translation Strategies. *Discourse Studies* 12 (5): 563-598.
- DAS, D. & TABOADA, M. 2018. RST Signalling Corpus: A Corpus of Signals of Coherence Relations. *Language Resources and Evaluation* 52 (1): 149-184.

- DEGAND, L. 1998. On Classifying Connectives and Coherence Relations. In M. STEDE, L. WANNER & E. HOVY (eds.), *Proceedings of the 1998 ACL Workshop on Discourse Relations and Discourse Markers (15 August 1998, Montreal)*. Stroudsburg: Association for Computational Linguistics (ACL): 29-35. Available online: <https://aclanthology.org/W98-0305.pdf>.
- DEMBERG, V., SCHOLMAN, M.C.J. & ASR, F.T. 2019. How Compatible Are Our Discourse Annotation Frameworks? Insights from Mapping RST-DT and PDTB Annotations. *Dialogue and Discourse* 10 (1): 87-135.
- EVERS-VERMEUL, J., HOEK, J. & SCHOLMAN, M.C.J. 2017. On Temporality in Discourse Annotation: Theoretical and Practical Considerations. *Dialogue and Discourse* 8 (2): 1-20.
- EVERS-VERMEUL, J. & SANDERS, T.J.M. 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, J. & SANDERS, T.J.M. 2011. Discovering Domains – On the Acquisition of Causal Connectives. *Journal of Pragmatics* 43 (6): 1645-1662.
- FOLTZ, P.W. 2003. Quantitative Cognitive Models of Text and Discourse Processing. In A.C. GRAESSER, M.A. GERNSBACHER & S.R. GOLDMAN (eds.), *Handbook of discourse processes*. Mahwah – London: L. Erlbaum: 487-523.
- GIBBS, R.W. 1996. What's Cognitive about Cognitive Linguistics? In E.H. CASAD (ed.), *Cognitive Linguistics in the Redwoods: The Expansion of a New Paradigm in Linguistics*. Berlin – New York: De Gruyter: 27-53.
- GONZÁLEZ, M. 2005. Pragmatic Markers and Discourse Coherence Relations in English and Catalan Oral Narrative. *Discourse Studies* 7 (1): 53-86.
- GRISOT, C. 2018. Tense, Grammatical Aspect and Subjectivity: An Experimental Study Using Inter-Annotator Agreement Rates and Corpus-Based Data. *Corpus Pragmatics* 2 (1): 27-55.
- GRISOT, C. & BLOCHOWIAK, J. 2017. Temporal Connectives and Verbal Tenses as Processing Instructions: Evidence from French. *Pragmatics and Cognition* 24 (3): 404-440.
- GRISOT, C. & BLOCHOWIAK, J. 2021. Temporal Relations at the Sentence and Text Genre Level: The Role of Linguistic Cueing and Non-Linguistic Biases – An Annotation Study of a Bilingual Corpus. *Corpus Pragmatics* 5 (3): 379-419.
- HARMS, C. & LAKENS, D. 2018. Making “Null Effects” Informative: Statistical Techniques and Inferential Frameworks. *Journal of Clinical and Translational Research* 3 (Suppl. 2): 382-393.
- HOBBS, J.R. 1979. Coherence and Coreference. *Cognitive Science* 3 (1): 67-90.
- HOEK, J., EVERS-VERMEUL, J. & SANDERS, T.J.M. 2019. Using the Cognitive Approach to Coherence Relations for Discourse Annotation. *Dialogue and Discourse* 10 (2): 1-33.
- HOEK, J. & ZUFFEREY, S. 2015. Factors Influencing the Implication of Discourse Relations across Languages. In H. BUNT (ed.), *Proceedings of the 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation – ISA-11 (14 April 2015, London)*. Stroudsburg: Association for Computational Linguistics (ACL): 1-7. Available online: <https://aclanthology.org/W15-0205.pdf>.

- HOEK, J., ZUFFEREY, S., EVERS-VERMEUL, J. & SANDERS, T.J.M. 2018. The Linguistic Marking of Coherence Relations: Interactions between Connectives and Segment-Internal Elements. *Pragmatics and Cognition* 25 (2): 276-309.
- HOVY, E.H. & MAIER, E. 1995. Parsimonious or Profligate: How Many and Which Discourse Structure Relations? Unpublished manuscript. 32 p. Available online: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.41.745&rep=rep1&type=pdf>.
- IZUTSU, M.N. 2008. Contrast, Concessive, and Corrective: Toward a Comprehensive Study of Opposition Relations. *Journal of Pragmatics* 40 (4): 646-675.
- KAY, P. & REGIER, T. 2006. Language, Thought and Color: Recent Developments. *Trends in Cognitive Sciences* 10 (2): 51-54.
- KEHLER, A. 2002. *Coherence, Reference, and the Theory of Grammar*. Stanford: CSLI Publications.
- KEHLER, A., KERTZ, L., ROHDE, H. & ELMAN, J.L. 2008. Coherence and Coreference Revisited. *Journal of Semantics* 25 (1): 1-44.
- KENNEDY, W. 2009. "Cognitive Plausibility" in Cognitive Modeling, Artificial Intelligence, and Social Simulation. In A. HOWES, D. PEEBLES & R.P. COOPER (eds.), *Proceedings of the 9th International Conference on Cognitive Modeling – ICCM2009 (24-26 July 2009, Manchester, UK)*. 454-455. Available online: <https://iccm-conference.neocities.org/2009/proceedings/cd/papers/235/paper235.pdf>.
- KLAVAN, J. & DIVJAK, D. 2016. The Cognitive Plausibility of Statistical Classification Models: Comparing Textual and Behavioral Evidence. *Folia Linguistica* 50 (2): 355-384.
- KNOEPKE, J., RICHTER, T., ISBERNER, M.B., NAUMANN, J., NEEB, Y. & WEINERT, S. 2017. Processing of Positive-Causal and Negative-Causal Coherence Relations in Primary School Children and Adults: A Test of the Cumulative Cognitive Complexity Approach in German. *Journal of Child Language* 44 (2): 297-328.
- KNOTT, A. 1996. *A Data-Driven Methodology for Motivating a Set of Coherence Relations*. PhD thesis. University of Edinburgh.
- KNOTT, A. & DALE, R. 1994. Using Linguistic Phenomena to Motivate a Set of Coherence Relations. *Discourse Processes* 18 (1): 35-62.
- KNOTT, A. & SANDERS, T.J.M. 1998. The Classification of Coherence Relations and Their Linguistic Markers: An Exploration of Two Languages. *Journal of Pragmatics* 30 (2): 135-175.
- KÖHNE-FUETTERER, J., DRENHAUS, H., DELOGU, F. & DEMBERG, V. 2021. The Online Processing of Causal and Concessive Discourse Connectives. *Linguistics* 59 (2): 417-448.
- LAKOFF, G. 1991. Cognitive versus Generative Linguistics: How Commitments Influence Results. *Language and Communication* 11 (1-2): 53-62.
- LEVSHINA, N. & DEGAND, L. 2017. Just Because: In Search of Objective Criteria of Subjectivity of Causal Connectives. *Dialogue and Discourse* 8 (1): 132-150.
- MACWHINNEY, B. 2014. *The CHILDES Project: Tools for Analyzing Talk*. New York: Psychology Press. Vol. 2: *The Database*.
- MAGLIANO, J.P. & GRAESSER, A.C. 1991. A Three-Pronged Method for Studying Inference Generation in Literary Text. *Poetics* 20 (3): 193-232.

- MAK, W.M., TRIBUSHININA, E. & ANDREIUSHINA, E. 2013. Semantics of Connectives Guides Referential Expectations in Discourse: An Eye-Tracking Study of Dutch and Russian. *Discourse Processes* 50 (8): 557-576.
- MANN, W.C. & THOMPSON, S.A. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text – Interdisciplinary Journal for the Study of Discourse* 8 (3): 243-281.
- MENDES, A., RIO, I. DEL, STEDE, M. & DOMBEK, F. 2018. A Lexicon of Discourse Markers for Portuguese – LMD-PT. In N. CALZOLARI, K. CHOUKRI, C. CIERI, T. DECLERCK, S. GOGGI, K. HASIDA, H. ISAHARA, B. MAEGAARD, J. MARIANI, H. MAZO, A. MORENO, J. ODIJK, S. PIPERIDIS & T. TOKUNAGA (eds.), *Proceedings of the 11th International Conference on Language Resources and Evaluation – LREC 2018 (7-12 May 2018, Miyazaki, Japan)*. Paris: European Language Resources Association (ELRA): 4379-4384. Available online: <https://aclanthology.org/L18-1693.pdf>.
- MILLER, G.A. 1969. A Psychological Method to Investigate Verbal Concepts. *Journal of Mathematical Psychology* 6 (2): 169-191.
- MÍROVSKÝ, J., SYNKOVÁ, P., RYSOVÁ, M. & POLÁKOVÁ, L. 2017. CzeDLex – A Lexicon of Czech Discourse Connectives. *The Prague Bulletin of Mathematical Linguistics* 109 (1): 61-91.
- MOESCHLER, J. & AUCHLIN, A. 2005. *Introduction à la linguistique contemporaine*. Paris: A. Colin [2nd ed.].
- MORERA, Y., LEÓN, J.A., ESCUDERO, I. & VEGA, M. DE 2017. Do Causal and Concessive Connectives Guide Emotional Expectancies in Comprehension? A Double-Task Paradigm Using Emotional Icons? *Discourse Processes* 54 (8): 583-598.
- MURRAY, J.D. 1997. Connectives and Narrative Text: The Role of Continuity. *Memory and Cognition* 25 (2): 227-236.
- MYERS, J.L., SHINJO, M. & DUFFY, S.A. 1987. Degree of Causal Relatedness and Memory. *Journal of Memory and Language* 26 (4): 453-465.
- NUYTS, J. 1992. *Aspects of a Cognitive-Pragmatic Theory of Language: On Cognition, Functionalism and Grammar*. Amsterdam – Philadelphia: J. Benjamins.
- PANDER MAAT, H. 1998. Classifying Negative Coherence Relations on the Basis of Linguistic Evidence. *Journal of Pragmatics* 30 (2): 177-204.
- PHILLIPS, L. & PEARL, L. 2015. The Utility of Cognitive Plausibility in Language Acquisition Modeling: Evidence from Word Segmentation. *Cognitive Science* 39 (8): 1824-1854.
- POPPER, K.R. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.
- PRASAD, R., MILTSAKAKI, E., DINESH, N., LEE, A., JOSHI, A., ROBALDO, L. & WEBBER, B.L. 2007. *The Penn Discourse Treebank 2.0 Annotation Manual*. IRCS Technical Reports Series 203. Philadelphia: University of Pennsylvania ScholarlyCommons. Available online: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1203&context=ircs_reports.
- REHBEIN, I., SCHOLMAN, M.C.J. & DEMBERG, V. 2016. Annotating Discourse Relations in Spoken Language: A Comparison of the PDTB and CCR Frameworks. In N. CALZOLARI, K. CHOUKRI, T. DECLERCK, S. GOGGI, M. GROBELNIK, B. MAEGAARD, J. MARIANI, H. MAZO, A. MORENO, J. ODIJK & S. PIPERIDIS (eds.), *Proceedings of the 10th International Conference on Language Resources and Evaluation – LREC 2016 (May 2016, Portorož, Slovenia)*. Paris: European Language Resources Association (ELRA): 1039-1046. Available online: <https://aclanthology.org/L16-1165.pdf>.

- ROBALDO, L. & MILTSAKAKI, E. 2014. Corpus-Driven Semantics of Concession: Where Do Expectations Come From? *Dialogue and Discourse* 5 (1): 1-36.
- ROBERSON, D., DAVIDOFF, J., DAVIES, I.R.L. & SHAPIRO, L.R. 2005. Color Categories: Evidence for the Cultural Relativity Hypothesis. *Cognitive Psychology* 50 (4): 378-411.
- ROHDE, H., TYLER, J. & CARLSON, K. 2017. Form and Function: Optional Complementizers Reduce Causal Inferences. *Glossa: A Journal of General Linguistics* 2 (art. 53): 1-27. Available online: <https://www.glossa-journal.org/article/4915/galley/12365/download/>.
- ROZE, C., DANLOS, L. & MULLER, P. 2012. LEXCONN: A French Lexicon of Discourse Connectives. *Discours* 10: 1-15. Available online: <https://journals.openedition.org/discours/8645>.
- RYSOVÁ, M. & RYSOVÁ, K. 2018. Primary and Secondary Discourse Connectives: Constraints and Preferences. *Journal of Pragmatics* 130: 16-32.
- SANDERS, T.J.M., DEMBERG, V., HOEK, J., SCHOLMAN, M.C.J., ASR, F.T., ZUFFEREY, S. & EVERS-VERMEUL, J. 2018. Unifying Dimensions in Coherence Relations: How Various Annotation Frameworks Are Related. *Corpus Linguistics and Linguistic Theory* (ahead of print): 1-71.
- SANDERS, T.J.M. & EVERS-VERMEUL, J. 2019. Subjectivity and Causality in Discourse and Cognition: Evidence from Corpus Analyses, Acquisition and Processing. In Ó. LOUREDA, I. RECIO FERNÁNDEZ, L. NADAL & A. CRUZ (eds.), *Empirical Studies of the Construction of Discourse*. Amsterdam – Philadelphia: J. Benjamins: 273-298.
- SANDERS, T.J.M., SPOOREN, W.P.M. & NOORDMAN, L.G.M. 1992. Toward a Taxonomy of Coherence Relations. *Discourse Processes* 15 (1): 1-35.
- SANDERS, T.J.M., SPOOREN, W.P.M. & NOORDMAN, L.G.M. 1993. Coherence Relations in a Cognitive Theory of Discourse Representation. *Cognitive Linguistics* 4 (2): 93-133.
- SCHOLMAN, M.C.J. & DEMBERG, V. 2017. Crowdsourcing Discourse Interpretations: On the Influence of Context and the Reliability of a Connective Insertion Task. In N. SCHNEIDER & N. XUE (eds.), *Proceedings of the 11th Linguistic Annotation Workshop – LAW XI (3 April 2017, Valencia, Spain)*. Stroudsburg: Association for Computational Linguistics (ACL): 24-33. Available online: <https://aclanthology.org/W17-0803.pdf>.
- SCHOLMAN, M.C.J., DEMBERG, V. & SANDERS, T.J.M. 2020. Individual Differences in Expecting Coherence Relations: Exploring the Variability in Sensitivity to Contextual Signals in Discourse. *Discourse Processes* 57 (10): 844-861.
- SCHOLMAN, M.C.J., EVERS-VERMEUL, J. & SANDERS, T.J.M. 2016. Categories of Coherence Relations in Discourse Annotation: Towards a Reliable Categorization of Coherence Relations. *Dialogue and Discourse* 7 (2): 1-28.
- SMITH, E. & HANCOX, P. 2001. Representation, Coherence and Inference. *Artificial Intelligence Review* 15 (4): 295-323.
- SPOOREN, W.P.M. & DEGAND, L. 2010. Coding Coherence Relations: Reliability and Validity. *Corpus Linguistics and Linguistic Theory* 6 (2): 241-266.
- STEDE, M., SCHEFFLER, T. & MENDES, A. 2019. Connective-Lex: A Web-Based Multilingual Lexical Resource for Connectives. *Discours* 24: 1-38. Available online: <https://journals.openedition.org/discours/10098>.

- STEDE, M. & UMBACH, C. 1998. DiMLex: A Lexicon of Discourse Markers for Text Generation and Understanding. In *COLING 1998 Volume 2: The 17th International Conference on Computational Linguistics*. Stroudsburg: Association for Computational Linguistics (ACL): 1238-1242. Available online: <https://aclanthology.org/C98-2197.pdf>.
- STUKKER, N. & SANDERS, T.J.M. 2012. Subjectivity and Prototype Structure in Causal Connectives: A Cross-Linguistic Perspective. *Journal of Pragmatics* 44 (2): 169-190.
- SUN, K. & ZHANG, L. 2018. Quantitative Aspects of PDTB-Style Discourse Relations across Languages. *Journal of Quantitative Linguistics* 25 (4): 342-371.
- TABOADA, M. & MANN, W.C. 2006. Rhetorical Structure Theory: Looking Back and Moving Ahead. *Discourse Studies* 8 (3): 423-459.
- TONELLI, S., RICCARDI, G., PRASAD, R. & JOSHI, A. 2010. Annotation of Discourse Relations for Conversational Spoken Dialogs. In N. CALZOLARI, K. CHOUKRI, B. MAEGAARD, J. MARIANI, J. ODIJK, S. PIPERIDIS, M. ROSNER & D. TAPIAS (eds.), *Proceedings of the 7th International Conference on Language Resources and Evaluation – LREC 2010*. Paris: European Language Resources Association (ELRA): 2084-2090. Available online: http://www.lrec-conf.org/proceedings/lrec2010/pdf/184_Paper.pdf.
- TOWNSEND, D.J. 1983. Thematic Processing in Sentences and Texts. *Cognition* 13 (2): 223-261.
- WEBBER, B. 2013. What Excludes an Alternative in Coherence Relations? In A. KOLLER & K. ERK (eds.), *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Long Papers*. Stroudsburg: Association for Computational Linguistics (ACL): 276-287. Available online: <https://aclanthology.org/W13-0124.pdf>.
- WEBBER, B., PRASAD, R., LEE, A. & JOSHI, A. 2019. *The Penn Discourse Treebank 3.0 Annotation Manual*. Philadelphia: University of Pennsylvania, Linguistic Data Consortium. 81 p. Available online: <https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf>.
- WEI, Y., MAK, W.M., EVERS-VERMEUL, J. & SANDERS, T.J.M. 2019. Causal Connectives as Indicators of Source Information: Evidence from the Visual World Paradigm. *Acta Psychologica* 198 (art. 102866): 1-13.
- WOLF, F. & GIBSON, E. 2005. Representing Discourse Coherence: A Corpus-Based Study. *Computational Linguistics* 31 (2): 249-288.
- YUNG, F., SCHOLMAN, M.C.J. & DEMBERG, V. 2021. A Practical Perspective on Connective Generation. In C. BRAUD, C. HARDMEIER, J.J. LI, A. LOUIS, M. STRUBE & A. ZELDES (eds.), *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*. Stroudsburg: Association for Computational Linguistics (ACL): 72-83. Available online: <https://aclanthology.org/2021.codi-main.7.pdf>.
- ZEYREK, D., MENDES, A., GRISHINA, Y., KURFALI, M., GIBBON, S. & OGRONCZUK, M. 2020. TED Multilingual Discourse Bank (TED-MDB): A Parallel Corpus Annotated in the PDTB Style. *Language Resources and Evaluation* 54: 587-613.
- ZHOU, Y. & XUE, N. 2015. The Chinese Discourse TreeBank: A Chinese Corpus Annotated with Discourse Relations. *Language Resources and Evaluation* 49: 397-431.
- ZUFFEREY, S. & DEGAND, L. 2017. Annotating the Meaning of Discourse Connectives in Multilingual Corpora. *Corpus Linguistics and Linguistic Theory* 13 (2): 399-422.

- ZUFFEREY, S., MAK, W.M., VERBRUGGE, S. & SANDERS, T.J.M. 2018. Usage and Processing of the French Causal Connectives “Car” and “Parce que”. *Journal of French Language Studies* 28 (1): 85-112.
- ZUFFEREY, S. & POPESCU-BELIS, A. 2017. Discourse Connectives: Theoretical Models and Empirical Validations in Humans and Computers. In J. BLOCHOWIAK, C. GRISOT, S. DURRLEMAN & C. LAENZLINGER (eds.), *Formal Models in the Study of Language: Applications in Interdisciplinary Contexts*. Cham: Springer: 375-390.

Appendix – Condition relations in existing frameworks

Tables 2 and 3 present an overview of the relational labels belonging to the Condition classes in the two most well-known discourse frameworks, the Penn Discourse Treebank 2.0 (PDTB; Prasad et al., 2007) and the Rhetorical Structure Theory Discourse Treebank (RST-DT; Carlson & Marcu, 2001). Both approaches adhere to descriptive adequacy, and both have been used successfully by many different researchers to annotate texts. However, as can be seen in Tables 2 and 3, they differ in their classification of fine-grained relation types.

A short description based on the manuals is provided in order to illustrate the fine-grained differences between the definitions. The respective manuals provide examples of each label.

Relation type	Subtype	Short description
Condition		
	Hypothetical	If the condition holds true, the consequence is caused to hold at some instant in all possible futures.
	General	Every time that the condition holds true, the consequence is also caused to be true.
	Factual present	The condition has either been presented as a fact in the prior discourse or is believed by somebody other than the speaker/writer.
	Factual past	Similar to Factual present but the condition describes a situation that is assumed to have taken place at a time in the past.
	Unreal present	The condition describes a condition that either does not hold at present or is considered unlikely to hold.
	Unreal past	The condition describes a situation that did not occur in the past and the consequence expresses what the consequence would have been if it had.
Pragmatic condition		
	Relevance	The condition provides the context in which the description of the situation in consequence is relevant.
	Implicit assertion	Applies in special rhetorical uses of <i>if</i> -constructions when the interpretation of the conditional construction is an implicit assertion.

Table 2 – Types of Condition relations in the PDTB 2 (Prasad et al., 2007)

Relation type	Subtype	Short description
Condition		
	Condition	The truth of the proposition is a consequence of the fulfillment of the condition.
	Hypothetical	The condition is not factual; the consequences would arise should the condition come true.
	Contingency	The satellite suggests an abstract notion of recurrence or habituality.
	Otherwise	Realizing the condition will prevent the realization of the consequences. (<i>Otherwise</i> typically marks this relation type.)

Table 3 – Types of Condition relations in the RST-DT (Carlson & Marcu, 2001)