Crible, L. 2022. The syntax and semantics of coherence relations. From relative configurations to predictive signals. *International Journal of Corpus Linguistics* (online first).

The syntax and semantics of coherence relations

From relative configurations to predictive signals

Ludivine Crible

Ghent University

This corpus-based study investigates the inter-relation between discourse markers (DMs) and other contextual signals that contribute to the interpretation of coherence relations. The objectives are three-fold: i) to provide a comprehensive and systematic portrait of the syntax and semantics of a set of coherence relations in English; ii) to draw a distinction between mere tendencies of co-occurrence and strong predictive signals; iii) to identify factors that account for the variation of these signals, focusing on relation complexity, DM strength and genre preferences. The methodology combines systematic coding (description) and multivariate statistical modelling (prediction). While the effect of genre and relation complexity was found to be null or moderate, the presence of discourse signals systematically varies with the ambiguity of the DM in the relation: signals co-occur more with ambiguous DMs than with more informative ones.

Keywords: discourse markers, discourse signals, underspecification, usage-based, cognitive complexity

1. Introduction

To make sense of a text is to build a cognitive representation of the many coherence relations that hold between discourse segments (Sanders et al., 1992). These links include additive, causal, contrastive and conditional relations, among many others. Although coherence relations are sometimes left implicit, they are typically signalled by discourse markers (DMs; e.g. Knott & Dale, 1994).[1] These expressions, also called 'connectives', serve as processing instructions by "constrain[ing] the interpretation of the utterances that contain them by virtue of the inferential connections they express" (Blakemore, 1987: 105). There is ample experimental evidence of the role of DMs in text comprehension and processing (e.g. Millis & Just, 1994; Murray, 1997). DMs indeed add relational meaning to the connected textual segments in isolation, which is a constituting condition of coherence relations. Some of their properties directly reflect – and in fact, help organize – categories of coherence relations (Knott & Sanders, 1998). DMs hence have a central status in the study of coherence relations. Yet, the DM category is very diverse and some of its members are less-than-ideal signals of coherence relations because of their famous ambiguity: this is the case for ambiguous expressions such as *while* or *since* that encode two meanings; it is also – and to a greater extent – the case for the conjunction *and*, which can be used in many different coherence relations but barely encodes any information (e.g. Cain & Nash, 2011; Spooren, 1997). Some DMs also have non-DM, grammatical uses (e.g. *and* in *Jack and Jill*; *while* in *walk for a while*), which adds to their semantic ambiguity.

Furthermore, DMs are not the only signals that can support the interpretation of a coherence relation. This is not a new observation, as most frameworks consider DMs only as the prototypical, not the exclusive, signals for relations, and alternative lexicalizations were already included in the Penn Discourse TreeBank 2.0 (henceforth PDTB; Prasad et al., 2008). However, it is only recently that studies have paid specific attention to the impact of other features on discourse interpretation, such as negation (Webber, 2013) or complementizers (Rohde et al., 2017). Chief among them, the RST Signalling Corpus (Das et al., 2015) has taken up the daunting task of annotating non-DM signals relevant to the signalling of coherence relations, covering syntactic and semantic features of the connected segments. Their approach, developed in Das and Taboada (2018), yielded valuable findings regarding the association between relation types and signal types.

The present study aims to complement frameworks focusing on DMs and builds on more inclusive approaches to discourse signalling. It adopts a new and innovative methodology, which provides a comprehensive and systematic portrait of the interaction between DMs and their linguistic environment. More specifically, our aim is to draw a distinction between, on the one hand, relative configurations, which show tendencies of co-occurrence between relations and signals, and on the other, predictive signals, which are specific to a given relation and can therefore serve as strong processing cues (see Hoek et al., 2018).

Apart from the innovative methodology, this study also stands apart from previous accounts of discourse signals by including two further factors of variation. The first one is genre: while signals have mostly been studied in newspaper corpora, little is known of their variation across genres (Taboada, 2006; Liu, 2019). Three communicative settings are compared in this study: spoken conversations, chat conversations and written essays. The second factor brought to light in the present analysis is the ambiguity of the DM in the relation. As mentioned before, not all DMs are equal in their ability to express coherence relations, and this further layer of variation is here labelled as ‘marker strength’.

In sum, this study pursues a triple objective. Firstly, the precise semantic and syntactic configurations of a set of coherence relations will be portrayed, broken down by genre and DM strength. Secondly, it will further our understanding of the processes of discourse signalling by adopting a new statistical approach and by integrating additional factors in the equation. These results will be considered from a cognitive perspective, relating to factors such as the complexity of the coherence relations (Hoek et al., 2017) or the causality-by-default hypothesis (Sanders, 2005). Lastly, the study will refine the continuum from implicit to explicit relations, taking into account not only different types of signals but also the predictive power or strength of these signals. In doing so, I shift the focus from absence vs. presence of DMs to the co-occurrence of DMs with other signals, thus situating coherence relations at the interface between semantics, syntax and discourse.

2. Coherence relations and their signals

In this section, I will review previous relevant approaches to coherence relations, discourse markers and other discourse signals, before I develop the hypotheses of the study.

2.1 Coherence relations

Coherence relations are “an aspect of meaning of two or more discourse segments that cannot be described in terms of the meaning of the segments in isolation” (Sanders et al., 1992: 2). There are many frameworks that propose classifications of coherence relations, most of them in the form of inventories of labels that can apply between two clauses. Two of the most widespread taxonomies come from Rhetorical Structure Theory (henceforth RST, Mann & Thompson, 1988) and the PDTB 2.0 (Prasad et al., 2008). Although the number and labels of relations can vary greatly between frameworks, a common core of relations such as cause-consequence, contrast, concession, addition, condition or alternative is usually included. These relations apply to both spoken and written data (e.g. Tonelli et al., 2010) and are included in other frameworks more dedicated to the analysis of DMs in spoken discourse (e.g. Cuenca, 2013).

Multiple corpus-based and experimental findings converge in suggesting that some relations are more complex than others, as evidenced through a later age of acquisition (e.g. Evers-Vermeul & Sanders, 2009) or longer reading times (e.g. Sanders & Noordman, 2000). For instance, negative relations such as contrast or concession have been found to be more costly to process than positive relations such as addition (see the continuity hypothesis, Murray, 1997; Xu et al., 2015). By contrast, the difference between additive and causal relations is less clear-cut: although logically more complex, causals are processed faster and remembered better than additives (Sanders & Noordman, 2000), due to the natural ease with which humans process causality. This observation led Sanders (2005: Section 4) to suggest a "paradox of causal complexity" (see also Hoek et al., 2017).

2.2 The polyfunctionality of discourse markers

DMs are optional procedural expressions which "bracket units of talk" (Schiffrin, 1987: 31) and "integrate their host utterance into a developing mental model of the discourse in such a way as to make that utterance appear optimally coherent" (Hansen, 2006: 25). They are famous for their polyfunctionality, especially in dialogue where they perform multiple functions at once (Petukhova & Bunt, 2009). Some DMs are monosemous (*because*, *therefore*, *whereas*) but many can occur in multiple relations.

In particular, the conjunction *and* is highly flexible and has been attested in a large number of relations, including negative ones (see Pander Maat, 1999; Prasad et al., 2008), contrary to what other models may suggest (Sanders et al., 1992: 4). Spooren (1997) dedicates a production study on the underspecified use of Dutch *en* "and" to express temporal and causal relations and shows that these uses are more frequent in children and learners than in more proficient speakers, thus displaying poor recipient design. Cain and Nash (2011) confirm these findings and show that such underspecified uses of *and* lead to longer reading times compared to stronger DMs.

Asr and Demberg (2012) have developed a probabilistic theorem to measure the 'cue strength' of DMs in expressing certain coherence relations. The formula takes into account the frequency of the DM, of the relation, of their combination and of the total number of DMs in the corpus. The higher the resulting score, the stronger the DM, that is, the more specific and exclusive it is to the particular relation. Asr and Demberg (2020) recently showed that these distributional differences in DM meanings are reflected in offline and online measures: for closely related DMs expressing multiple relations (*but* vs. *although*; contrast vs. concession), the most frequently expressed relation for each DM is judged more coherent and read faster than the less frequent relation. The notion of DM strength relates to psycholinguistic accounts of language production, such as the Uniform Information Density hypothesis (Levy & Jaeger, 2007), which states that language users tend to avoid peaks and troughs in information density, that is, they do not explicitly mark some information which was already available, nor leave unmarked some information which was unexpected.

Finally, one should note that the use and polyfunctionality of DMs have been related to genre variation in a number of studies showing that speech and writing attract partly different types of DMs and relations or functions. Biber (2006), for instance, found that adverbials such as *however, therefore* or *for example* are more typical of written registers, where writers can plan and edit their texts at will, as opposed to the higher planning pressure of spontaneous speech. DMs related to interpersonal relationships are more frequent in spoken than written registers (Kunz & Lapshinova-Koltunski, 2015), whereas the distance between writers and readers is larger in writing (Clark, 1996). Genre also affects how DMs are perceived: Crible and Demberg (2020) found that the contrastive use of *and* is inacceptable in comments to online press articles but not in more conversational texts (chat discussions), thus showing an effect of formality. Genre therefore appears as a primary factor in the meaning variation of DMs.

2.3 Other discourse signals

Beyond DMs, many other linguistic devices contribute to the construal of coherence relations. Pander Maat (1999), for instance, investigated whether the marking of additive and comparative relations was influenced by the presence of a "similarity assumption" and the expression of differences in the two connected segments. His corpus study on the specific genre of stock-market reports indeed showed that "comparative relations (that is, relations invoking some similarity assumption) are typically marked by connectives or lexical markers; additive relations are not" (Pander Maat, 1999: 179). This study is highly innovative in that it examines the context of DMs, including "lexical markers". However, the role of context is here limited to specific semantic properties.

Other case studies, mainly experimental, have tested the influence of given contextual features on the use and interpretation of coherence relations, such as negation (Webber, 2013), complementizers (Rohde et al., 2017), pronouns (Mak et al., 2013) or implicit causality verbs (Koornneef & Sanders, 2013). They all converge in showing strong links between DMs and their context, in the form of expectations or constraints generated by these linguistic features. In a similar vein, annotations of coherence relations in the PDTB 2.0 include lexical markers called 'alternative lexicalizations' (henceforth AltLex), which, in the absence of DMs, express the meaning of the relation (e.g. *The result is that* or *The reason for this is*; Prasad et al., 2008). In the PDTB 3.0 (Prasad et al., 2018), another type of AltLex, 'AltLex-C', has been included, covering syntactic constructions such as *such NP that* for result relations or *so [too] <aux> NP* for similarity. In the PDTB 3.0, AltLexes are annotated even in the presence of an explicit DM, and their Appendix H shows that in most cases, the DM is *and* (Prasad et al., 2018: 79).

The RST Signalling Corpus (Das et al., 2015) has recently extended this line of work by annotating *any* feature responsible for or relevant to the interpretation of the coherence relation. Their method, developed in Das and Taboada (2018), identifies signals which conceptually relate to the annotated relation. For instance, in their corpus, an 'Elaboration' relation might either be indicated by a DM such as *in particular*, a co-referential chain, a synonymy relation, a specific syntactic construction or clause type, punctuation marks, numerical terms or even genre-specific patterns. The authors report on proportions of signalling and signal types across different relations, showing that only a few relations remain completely unsignalled in their corpus. The extent of the analysis and the number of features covered is impressive. However, the method remains inherently subjective: it is the analyst who decides which feature signals the relation.

Das and Taboada (2019) further examined the co-occurrence between DMs and signals in the same newspaper corpus, testing whether this co-occurrence was mainly motivated by the type of DM or by the type of relation. They make a number of interesting observations regarding the ambiguity of DMs such as *and* and their tendency to co-occur with signals, although monosemous DMs such as *moreover* or *for instance* also frequently co-occur with signals in their data. In addition, they found that additive relations (List, Contrast, Elaboration) are more frequently marked by multiple signals than causal relations (Condition, Contingency), using Sanders et al.'s (1992) classification in cognitive primitives. Their distribution analysis remains largely exploratory and descriptive, and calls for a systematic, statistical account of the co-occurrence phenomenon.

The RST Signalling Corpus provides highly valuable grounds for the present analysis. The latter differs, however, by the more objective, statistical methodology and the slightly different focus of analysis. As such, I follow Hoek et al.'s (2018) recommendations to distinguish between types or degrees of discourse signals. Inspired by a parallel corpus study, the authors suggest three groups of signals depending on the strength of association and mutual dependency between the signal and the relation: 'division of labor' signals, which are redundant with DMs and suffice to express the relation (e.g. negation in substitution relations); 'agreement' signals, which are conceptually related to the relation (e.g. negation in contrastive relations); 'general collocation' signals, merely based on the frequency of co-occurrence but which do not involve conceptual similarity (e.g. negation in result relations). While these distinctions may be hard to apply systematically, they raise awareness of the fact that not all signals are equally strong or equally specific to a given relation. For instance, subject reference (i.e. the connected segments share the same subject) is often assigned as a signal for Elaboration relations in the RST corpus, but such a pattern can be found in many other relations, so that it is not very informative in itself. A statistical approach, disentangling mere frequency-based configurations from predictive signals, as undertaken in this study, will, on the contrary, differentiate between tendencies of association and actual signals, relating these different strategies to relation types and DM types.

Finally, the present approach addresses the gap left by previous studies on the genre variation of discourse signals. Most previous studies focus on one (written) genre, with the exception of Liu (2019): they found that signals (including DMs) were most frequent in how-to guides and least frequent in news articles, and that some signals are genre-specific, even though these mostly consist in DMs or other lexical elements that relate to the nature of the genre (e.g. *hypothesized* for academic articles, *warnings* in how-to guides) rather than to the nature of the relation. Differences between speech and writing and between various subgenres may play a role in the distribution of signals. As mentioned above for DMs, the distance between author and addressee is larger in writing than in speech (Clark, 1996), which might result in the need to bridge this distance through explicit linguistic marking. By contrast, in conversation, the listener can benefit from other cues such as prosody or gestures. In addition, the higher cognitive demands of spontaneous language (no preparation, pressure to speak in due time) might favor strategies of speaker economy (i.e. say no more than necessary), as opposed to the more informative principle of hearer economy (Horn, 1984). Finally, speech is traditionally described as fragmented, while writing is more integrated (Chafe, 1982), tendencies which could be reflected in the syntax and cohesiveness of coherence relations. In terms of discourse signalling, it can therefore be expected that genre will play a role in the distribution of signals, as the explicitness of the coherence relations and the quantity of information available for their retrieval may relate to contextual factors such as degree of formality and planning.

2.4 Interaction between relations and signals: hypotheses

Against this backdrop, the present study examines the semantic and syntactic configurations of coherence relations in light of three potential factors of variation: the cognitive complexity of the relation, the strength of the DM and the formality of the genre. Each factor and its related hypothesis is developed below.

It is established that some relations are more cognitively complex than others, namely negative relations, whereas no strong difference was found between additives and causals. As a result, I expect complex relations (contrast, concession) to primarily occur in recurrent configurations. These configurations are assumed to contribute to the coherence relation by further constraining its interpretation, as opposed to simpler relations (addition, specification, consequence), which do not require extra marking. I also expect complex relations to be more frequently reinforced by predictive signals than simpler relations, where the DM might not be accompanied by any other discourse signals.

Within particular coherence relations, DMs also differ in terms of their polyfunctionality and resulting score of DM strength. Following the Uniform Information Density hypothesis (Levy & Jaeger, 2007), and in line with Das and Taboada's (2019) observations, it is hypothesized that the weaker the DM, the more frequently it will be compensated by predictive signals, in order to provide sufficient information in all contexts. By contrast, stronger (i.e. less polyfunctional) DMs can efficiently signal a relation on their own and therefore do not require such compensation.

Finally, different expectations in terms of planning pressure and formality across the three genres under scrutiny should be reflected in the distribution of discourse signals. I expect a cline from the spoken informal genre (discussion) to the written formal genre (essay), with the written informal genre (chat) in an intermediate position. More specifically, strong signals and specific configurations are likely to occur more frequently in essays, where writers tend to be maximally informative and addressee-oriented.

3. Method

This section presents the corpus used in the study and the annotation methods for discourse markers and other discourse signals.

3.1 Corpus data

The hypotheses presented in the previous section were tested on three samples extracted from the Loyola Computer-Mediated Communication corpus (Goldstein-Stewart et al., 2008). This corpus is a collection of texts and recordings produced by 21 American university students, talking about the same six topics (e.g. gay marriage, privacy rights) in six different communicative settings: discussion, interview, chat, email, blog, essay. Every participant wrote or spoke about every topic in every communicative setting. The corpus design is thus highly controlled and allows for stable comparisons between genres and speakers.

I selected three genres from this corpus, namely discussion, chat and essay, because they form a cline of formality and conditions of production: discussion is a spontaneous, informal spoken task with a relatively high cognitive pressure; chat conversations are also somewhat spontaneous and informal but the written modality gives more time to participants to prepare and edit their messages, thus reducing the planning pressure; essays are prepared and formal. The corpus size can be found in Table 1.

Table 1. Data used in the study

Discussion	89,515 words
Chat	72,466 words
Essay	64,864 words
Total	226,845 words

3.2 Discourse marker annotation

The following subsections elaborate on the identification of discourse markers, the functional taxonomy and the analysis of discourse signals.

3.2.1 *Identification of discourse markers*

DM tokens were manually identified in a comprehensive, bottom-up way following operational criteria of syntactic optionality, high degree of fixedness or grammaticalization, discourse-level scope (thus excluding phrasal uses such as *Jack and Jill went up the hill*) and procedural meaning (Crible, 2017: 252). This formal-functional definition targets expressions that perform functions related to discourse structure (coherence relations, topic change) and interaction management (speaker-hearer relationship). It includes, in effect, conjunctions (e.g. *and, but, because, although*), adverbs (e.g. *so, however, specifically*), verb phrases (e.g. *you see, I mean*), prepositional phrases (e.g. *in other words, on the one hand*) and interjections (e.g. *well, oh*). A total of 109 English DM types were identified in the data, when they met the above criteria. The segments connected by the DMs (when applicable) were not explicitly identified, as this annotation is DM-based (as opposed to segmentation frameworks such as RST). Segments typically correspond to clauses.

3.2.2 *Functional taxonomy*

Once identified, the functions of the DM tokens were manually disambiguated, following Crible and Degand's (2019) taxonomy of discourse functions. This system distinguishes between four domains of use and 15 functions. Only the function labels will be discussed here. They mostly correspond to coherence relations: addition, alternative, cause, concession, condition, consequence, contrast, hedging, monitoring, specification, temporal, agreeing, disagreeing, topic, quoting. They are defined and illustrated in Crible and Degand (2019). This disambiguation is context-bound and can make use of any information available, including syntax, prosody, etc. As a result, the function of a given DM can vary across contexts. The authors of the taxonomy report on inter-annotator agreement measures reaching 80.36% ($\kappa = 0.655$) at the function level. The partial subjectivity of this annotation task was limited as much as possible by operational guidelines and regular consistency checks. In this study, I focus on five discourse functions:

i. addition, where the introduced segment brings related discourse-new information
ii. specification, where the introduced segment brings an example or a detail
iii. consequence, where the introduced segment brings the result of the first one
iv. concession, where the introduced segment denies some expectation of the other one
v. and contrast, where the introduced segment highlights a difference with the other

According to Sanders et al.'s (1992) primitives, concession and contrast are negative relations, hence cognitively complex, while addition, specification and consequence are positive relations, hence cognitively simple (see Section 2.1). These relations were selected because they have all been found to be expressed by the conjunction *and*, as well as by more specific DMs. In our attempt to study signalling configurations across various degrees of DM strength, I started from the functional spectrum of *and* and added to our selection other DMs which can also express the same relations. This selection resulted in a sample of 1,957 tokens and 18 types distributed as shown in Table 2.

Table 2. DMs included in the study and their functions

DM	**ADD**	**SPE**	**CSQ**	**CCS**	**CTR**	**Total**
and	301	271	50	78	39	**739**
but	13			150	179	**342**
so			159			**159**
however				93	33	**126**
actually		31		53		**84**
while				51	17	**68**
also	62					**62**
then			60			**60**
though				60		**60**
even if				52		**52**
although				45		**45**
whereas				1	32	**33**
therefore			24			**24**
for example		24				**24**
plus	24					**24**
on the other hand				18	3	**21**
yet				19		**19**
thus			15			**15**
Total	**400**	**326**	**308**	**620**	**303**	**1,957**

These DMs were sampled from a larger database, where all tokens have been functionally disambiguated. In this original corpus, there were 2,483 tokens of *and*, 1,903 of *but*, 844 of *so*, 389 of *then*, 223 of *though*, in addition to the other DMs in the table (and other DMs not mentioned in this study). Given the magnitude of the signalling annotation (see Section 3.2.3), I only analyzed a random sample of the above-mentioned DMs: for *and*, only 301 additive uses were annotated for signals, all its other uses (e.g. in consequence relations) were analyzed; all contrastive and additive uses of *but* were analyzed and only a sample of concessive uses of *but* were included, 50 in each genre; 53 tokens of *so* in each genre were analyzed; 20 tokens in each genre were analyzed for *then* and *though*.

To rank these DMs on a scale of marking strength, I applied Asr and Demberg's (2012) formula and log-transformed the scores in order to obtain a less skewed interval variable to be used in the statistical analysis.[2] These numerical scores allow us to distinguish between DMs (e.g. contrastive *but* has a log-transformed score of 0.56 vs. the score for *whereas* is 1.57) and within DMs, that is, the strength of a DM varies with the different relations it can express (e.g. *and* has a score of 0.68 for the relation of addition but only of -0.79 for concession). The smaller the score, the less frequent the function for a given DM. These scores are presented in Table 3, where empty cells indicate that the DM does not express the particular relation. They are also shown in Figure 1 for better visualization.

Table 3. Log-transformed scores of cue strength by DM and relation in the sample

	Addition	**Specification**	**Consequence**	**Concession**	**Contrast**
and	0.678518	0.352183	-0.72125	-0.79588	-0.22915
plus	0.764923				
also	0.764176				
for example		1.31597			
actually		0.820201		0.454845	
so			0.960946		
then			0.870404		
therefore			0.96895		
thus			0.969416		
but	-1.39794			0.649335	0.558709
however				0.564666	0.998695
while				0.459392	0.866287
whereas					1.567144
though				0.68842	
although				0.696356	
even if				0.696356	
yet				0.696356	
on the other hand				0.62941	0.735599

INSERT FIGURE 1 HERE

Figure 1. Scores of cue strength for each DM

3.2.3 *Discourse signals analysis*

A number of contextual features were further annotated on this sample of 1,957 DMs, in order to describe their syntactic and semantic configuration as well as other discourse signals. The list of features was elaborated from the literature, especially Das and Taboada (2018), and from testing phases on pilot data, adding, removing or re-categorizing features when relevant. It was designed to be applied to all types of DMs, not just the ones included in the present study. For instance, disfluencies or pauses are potential signals in the context of reformulative DMs; epistemic expressions can occur in subjective relations. This does not mean that a disfluency (or any feature) is considered as a clear signal for any relation, but that they are potentially relevant, in a bottom-up approach. The resulting taxonomy comprises the 25 features or potential signals shown in Table 4.

Table 4. Contextual features annotated as potential discourse signals

Group	**Signal type**	**Values**
Adjacent	Adjacent DM	*the DM token* (e.g. and *so*)
	Other DM in the unit	*the DM token* (e.g. and the cat was *however* mad)
	Adjacent pause	silent or filled pause
	Punctuation	dashes, parentheses, commas
	Disfluency	interruption, repetition, repair
	Response particle	*yes*, *no*
Sentence	Mood of the host	declarative, interrogative, imperative, exclamative
	Polarity of each unit	positive, negative + combinations
	Polarity difference	same, different (across units)
	Verb tense of each u.	present, past, future, conditional + combinations
	Tense difference	same, different (across units)
	Subject referent	same, different (across units)
	Unit type of the host	full, relative, completive, non-finite, elision, gerund
Syntax	Construction	parallelism, SV inversion, cleft, presentational, dislocation
Semantics	Semantic relation	synonymy, antonymy, metonymy, hyperonymy, hyponymy, equivalence, comparison
	AltLex	expression encoding the meaning of the relation
	Evaluative language	expression of stance (e.g. *wonderful*)
	Epistemic language	expression of reasoning (e.g. *probably*)
	Speech-act	expression of speaking (e.g. *ask*)
	Deictics	time and place references (e.g. *here*, *yesterday*, *2020*)

	Proper nouns	names referring to places, groups or persons (e.g. *London*)
	Numerals	the unit contains numbers or cardinals (e.g. *first, twenty*)
	Demonstratives	including possessives (e.g. *this, their*)
	Pronouns	referential chain between the two segments (*Mary… she*)
	Repetition	exact lexical repetition in the two segments

As we can see, the features include aspects of co-occurrence, morpho-syntax and semantics of the DM and its host unit. Sentence features (polarity, subject referents, etc.) were all systematically coded for all DMs. All other features were annotated when they were present in the immediate context, leaving the value blank when they were absent, given that not all DMs occur in contexts where a semantic relation or a deictic expression can be observed, for instance. However, whenever such features were present, they were always annotated, regardless of whether they actually contribute or are relevant to the DM or the coherence relation. This is the main difference with Das and Taboada's (2018) more subjective approach, where features are only annotated if they are judged to be relevant to the relation. Only AltLex expressions (which include verbs and nouns) were identified when they were related to the meaning or function of the DM (e.g. the verb or noun *result* for a consequence relation).[3]

More specifically, with this method, all variables (except AltLex) are given a value (including a "null" value) independently from the meaning of the DM, thus providing a very systematic and comprehensive portrait of the DM configurations. Annotated examples are detailed below.

(1) I have nothing to hide and don't really care if they scan my calls

In Example (1), the DM *and* expresses a relation of addition. Its contextual features are the following:

i. Co-occurring: no adjacent DM, no other DM in the unit, no pause (e.g. comma), no specific punctuation or prosodic pattern, no sign of disfluency, no response particle in the unit

ii. Sentence: declarative mood, negative polarity, same polarity in both segments, present tense, same verb tense in both segments, same subject referent, subject elision

iii. Syntax: no construction

iv. Semantics: no semantic relation, no AltLex, etc. (no semantic features)

This example strikes by the absence of contextual features (co-occurring, syntactic or semantic), while the sentence features are characteristic of continuity (same polarity, tense and subject; elision of the subject).

(2) In some places this difference is really subtle, but in others the gaps are much more apparent.

In (2), *but* expresses a relation of contrast. Its configuration is very different from (1):

i. Co-occurring: no adjacent DM, no other DM in the unit, pause (comma), no specific punctuation or prosodic pattern, no sign of disfluency, no response particle in the unit
ii. Sentence: declarative mood, positive polarity, same polarity in both segments, present tense, same verb tense in both segments, different subject referents, full unit
iii. Syntax: parallelism
iv. Semantics: comparative relation (*more*) and synonymy (*difference*, *gap*), AltLex (*others*), evaluative language (*really subtle*), no other semantic features

The two examples differ in their polarity, subject referent, presence of a pause or comma, type of unit, construction and most notably in their semantic features, with two semantic relations and an alternative lexicalization, as well as an expression of stance in (2).

In sum, this method does not require the analyst to make a decision regarding the relevance of the features (again, except for AltLex). On the contrary, this relevance will be established through the statistical analysis of the data, showing associations between relation types and their recurrent configurations. The most predictive signals will be identified as relation-specific features, which are both conceptually related to the meaning of the relation (see Hoek et al.'s (2018) agreement category) and only significantly frequent with this relation. Any annotation endeavor involves the analyst's subjectivity, and no one is infallible, so some features may be overlooked due to human error. However, subjectivity and circularity are as limited as possible, especially in comparison to previous approaches, offering a robust and innovative approach to discourse signalling.

4. Results: From configurations to predictive signals of coherence relations

I report on corpus-based findings extracted from the sample of 1,957 DM tokens. These were fully analyzed in terms of sense disambiguation and contextual features (potential signals), as explained above.

4.1 Configurations vs. signals across relations and genres

This first results section will compare the syntax and semantics of the five coherence relations under scrutiny and test whether these configurations vary across genres. I will thus be able to test hypotheses regarding the effect of genre and that of the cognitive complexity of the coherence relation on the distribution of weak and strong discourse signals. I start with the proportion of the most frequent value for sentence features, presented in Table 5. The table also includes two features from other classes of signals (syntactic constructions and semantic relations), which show interesting patterns of association.

Table 5. Proportion of contextual features across relations

Features	**ADD**	**SPE**	**CSQ**	**CCS**	**CTR**
Declarative mood	94.88	96.73	77.93	94.19	94.39
Positive polarity	72.44	71.73	56.86	54.42	46.86
Same polarity	77.32	74.40	64.21	56.94	49.50
Same verb tense	69.02	63.69	51.84	60.65	67.33
Different subject	65.61	81.55	72.24	66.77	68.98
Full unit	80.73	94.64	92.64	93.71	81.19
No construction	92.68	70.83	95.99	97.10	86.80
No semantic relation	95.37	96.73	97.32	93.55	69.97

A number of observations can be made from this table. Overall, it appears that none of these features can be assigned to one specific relation, and differences are more relative than clear-cut. In other words, attempting to discriminate between coherence relations solely on the basis of such basic features is not very informative. Still, tendencies for preferred configurations do emerge for some relations. Starting with mood, sentences are mostly declarative (around 95%), albeit to a smaller extent in consequence relations (77.93%), where interrogatives and imperatives are also relatively frequent. Polarity shows a clear cline from positive additive relations (addition and specification) to negative relations (concession, contrast), where the connected segments tend to exhibit different polarities. This can be seen clearly from the mosaic plot in Figure 2, where color shadings represent the Pearson residuals.

INSERT FIGURE 2 HERE

Figure 2. Difference in polarity across relations

Positive and negative relations are clearly opposed on this graph, with consequence relations in a neutral position (not significantly different). However, even in contrastive and concessive relations, there is still a substantial proportion of sentences that are in the positive polarity (around 50%). Therefore, this feature, often claimed to be a clear sign of negative relations, is only partially predictive of the relation type and should rather be seen as a relative tendency.

Difference in verb tense between the connected segments is not particularly relevant, except once more for consequence, where about half occurrences display a difference in tense. This relates to the concept of consequence, whereby a situation in one given time period causes the situation in a different time period, typically in the future, as in (3), where the author is chatting about gender discrimination.

(3) it’s been around for thousands of years… **so** it’ll take a lot to change it

Connected segments in all relations tend to have different subject referents, although the proportion is higher for specification, where a more specific referent is often introduced to develop the previous topic. In terms of unit types, addition and contrast stand out with a lower proportion of full units: these two relation types tend to connect subordinate or dependent units, such as relative or complement clauses (e.g. *John said that he was happy* ***and*** *that he wanted to sing*), more frequently than the other relations in the sample. This higher degree of dependency can be expected for addition, which can connect units at a very local level, but it is more surprising for contrast and mostly corresponds to cases of elision, as in (4), where the predicate adjective “equal” is elided in the second segment.

(4) I think **while** in theory things are equal, in practice they are not

Turning to syntactic constructions, we see that they occur with a substantial proportion in specification relations, whereas around 90% of the other relation types do not present any specific construction. Specification presents a much larger number of cleft and presentational constructions (*this is the NP that*, *there is*) than all other relations, as in (5).

(5) It isn't a sacred institution when a couple fills out some paperwork and is granted a marriage license, however, **and** *that's* the only part of the government's involvement *that* has any legal ramifications

The presence of semantic relations such as synonymy or antonymy is quite rare, except in contrastive relations where they take up 30% of the cases. They mostly consist in antonyms as well as a few comparatives, as in (6).

(6) My brother joined the Marine Corps after the war started, and he has already returned from his second tour of duty there. **However**, many of our soldiers are not *as fortunate as* my brother.

The other semantic features included in the analysis are more varied and much rarer. Some show interesting tendencies. For instance, deictic expressions are very rare (around 1%) in all relations except in contrast (9.24%), where they may contribute to comparing two situations. Pronouns referring to an entity of the previous segment are, however, quite common (more than 20%) in all relations except contrast (9.9%), precisely because there is little in common between two connected segments in a contrastive relation. Lexical repetition in the two segments is present in about 15% in all relations, and demonstratives in around 6%. On the whole, these semantic features may contribute locally to the interpretation of some contexts, but they are too rare to form recurrent configurations.

Such descriptive tendencies are rich and show meaningful links between the conceptual nature of the relation and the linguistic context in which it occurs. From such a systematic method of analysis, we can extract typical configurations for each relation, which may serve as constraints on the interpretation of the DM in a purely frequency-based approach. However, the resulting configurations cannot be taken as strong predictors of the relation expressed by the DM, given that, as we have seen, most features are present to a certain extent in all relations and the differences are only gradual. For instance, negative polarity is not a sufficient cue for contrast because it also concerns 30% of relations such as addition or specification.

This line of reasoning goes against the approach in Das and Taboada (2018), who annotate very frequent and pervasive features as 'signals' for a given relation provided it is conceptually related to that relation, regardless of whether these features also occur in other relations. As they acknowledge themselves: "the signals are *compatible* with a relation, not necessarily indicators of the relation exclusively" (Das & Taboada, 2018: 765, original emphasis). For instance, they identify 'reference' signals (e.g. pronouns, demonstratives) as typically associated with the Elaboration relation (similar to our specification), while I have shown that such cohesive features are equally frequent in all or most relations in our sample.

This brings us to the main distinction that our statistical approach allows us to draw, between relative configurations on the one hand and relation-specific statistical predictors on the other. The former can be found in Table 6 and are illustrated below. They do not always correspond to the absolute most frequent feature but take into account relative differences with the other relations.

Table 6. Typical configuration by relation

	Addition	**Specification**	**Consequence**	**Concession**	**Contrast**
Mood	declarative	declarative	interrogative, imperative	declarative	declarative
Polarity	positive	positive	positive	negative	negative
Verb tense	same	same	different	same	same
Subject	different	different	different	different	different
Unit	non-full	full	full	full	non-full
Construction	n/a	presentational	n/a	n/a	parallelism
Semantic	n/a	n/a	n/a	n/a	antonyms

(7) as long as it was proved there were no long term health effects **and** it was well regulated, I would be okay with pot being legalized [addition]

(8) there's two chemicals that really separate like males from females **and** that's testosteone and estrogen [specification]

(9) that is still a privacy issue **so** should we allow it? [consequence]

(10) There is a stereotype of women that they belong in the kitchen, meant for nothing else then taking care of the house and the family. **But** this is not at all true. [concession]

(11) some people like things the way they are **and** many don't [contrast]

In (7), the unit introduced by *and* is declarative and subordinated to the conjunction "as long as", the polarity (positive) and verb tense (past) is the same in both segments, the subject referent differs across segments, and there is no construction or semantic relation. In specification, (8), the features are similar except for the presentational constructions ("there's", "that's") and the unit introduced is independent. In the consequence example, (8), the host unit is interrogative and the verb tense is different (present vs. conditional). In concession, (9), the only notable feature is the negative polarity of the host unit, which differs from the first segment. Lastly in (10), contrast is signalled by the negative polarity of the host unit, the verb elision, the parallelism and the antonymy ("some" vs. "many").

All these features were then modelled and ranked by the conditional importance of variables. This is a tool used in random forests that reflects the impact of each potential predictor on the dependent variable (i.e. the coherence relation, in our case). I also included genre as a potential predictor, in case some relations were specific to one of the three genres in the corpus. Figure 3 shows the result of this first model.

INSERT FIGURE 3 HERE

Figure 3. Conditional importance of variables for coherence relations

We see that the most predictive feature is syntax, i.e. syntactic constructions such as presentational *This is the… that*. Semantic relations, type of unit, other semantic features and difference in polarity all contribute to the disambiguation to a similar extent. Mood, genre, subject difference and tense difference are less relevant: not only is their conditional importance smaller (below 0.02 on this graph), they are also much more pervasive features (e.g. declarative sentences are profuse; a given genre is shared by all relations in a text; different subject referents are common in all relations, see Table 6). I therefore ran a conditional inference tree on the four most predictive variables, excluding the miscellaneous "semantic" category which is too diverse and comprises many rare values, making the statistical model not interpretable. I obtained the classification presented in Figure 4.

INSERT FIGURE 4 HERE

Figure 4. Conditional inference tree for coherence relations

In classification trees, the height of bars at the bottom represents the frequency of each relation given the features specified in the nodes above; the higher the bar, the more frequent the relation in that configuration. From right to left, the tree makes a first distinction between specification (two boxes at the right extreme of the graph, nodes 20 and 21) and all other relations on the basis of syntactic constructions (cleft and presentational mostly), confirming the strong syntactic marking of this relation. Specification also relates to the semantic relations of equivalence, synonymy and hyponymy (node 17). The contrastive relation is then particularly identifiable by semantic relations (antonyms, comparatives or combinations thereof), as well as by specific constructions (parallelism, nodes 9 and 14, where it is as frequent as addition) in various configurations.

Addition associates with parallelism (node 14, along with contrast), with the *not only…but* construction (additive *but*, node 10) and with non-full units (complement clauses, elision, node 12). The only node which is relatively more specific to concession is node 8: full units with a difference in polarity between the segments. Finally, no feature is statistically related to consequence on the basis of this inference tree (it is never the most frequent relation in any node), suggesting that consequence can be inferred from many different configurations, with no specific signal attached to it.

In sum, this statistical analysis showed that specification, contrast and addition are preferentially marked by specific syntactic and/or semantic features, which can function as strong signals for the interpretation of the relation. By contrast, consequence and, to a smaller extent, concession occur in much more diverse configurations, with no particular predictive signal (apart from negation in concessive relations, with the reserve that I expressed above regarding the limits of polarity as a signal). If we zoom in on consequence relations, the only strong features in the context of the DMs are indicative words (here referred to as ‘AltLex’) that encode or strongly evoke the concept of consequence, such as the verb in (12).

(12) they’ve got to have a certain number of personnel and by hiring those defense contractors then that frees up the standard army soldiers /**so** it like <u>allows</u> them not to be there for so long

This lack of strong specific predictors for concession and consequence may be due to their shared causality component. Both relations are indeed causal: in consequence, the first segment is the cause for the second one (*Because of X, Y happened*); in concession, one segment creates an expectation which is denied (*Because of X, you might think Y, but in fact Z*). As evidenced in previous studies, causality is quite easily inferred, as humans tend to see causal relations everywhere (Sanders, 2005). Our quantitative approach to discourse configurations confirms that causal relations, either positive (consequence) or negative (concession), can occur in many different linguistic configurations and are not constrained by context as much as other relations such as specification or contrast.

To sum up so far, after frequency-based relative configurations (see Table 6), we are now able to identify relation-specific predictive signals (Table 7), extracted from the multivariate statistical model (conditional inference tree). In the remainder of the analysis, I restrict the predictive signals to these statistically relevant features. I also include indicative words (AltLex) for all relations since, by definition, AltLexes are conceptually specific to a relation (e.g. *as well* for addition, *others* for contrast, *by that logic* for consequence, *one of…* for specification or *supposed to* for concession).

Table 7. Predictive signals by relation

Addition	Specification	Consequence	Concession	Contrast
non-full units	presentational construction; hyponymy	n/a	n/a	antonymy; parallelism

I now report on the distribution of these predictive signals across relation types, in order to test the first hypothesis regarding the signalling of cognitively complex relations. Figure 5 shows the proportion of occurrences which are signalled by predictive signals (in addition to the DM) across relation types.

INSERT FIGURE 5 HERE

Figure 5. Presence of predictive signals across relations

As we can see, at least half of all relations do not present any other strong signals besides the DM. The proportion varies, however, between 11.25% of signalling for addition to 51.32% for contrast. Additive DMs are reinforced the least: this basic relation does not require extra marking, as expected by its low complexity, which also relates to Murray's (1997) continuity hypothesis. Concession has a more surprising low proportion of strong signals (19.22%), since negative relations have been shown to be more complex than positive relations (e.g. Hoek et al., 2017). I have already suggested that this lack of predictive signals may be due to the causality component of concession, which also explains the low proportion of signals in consequence relations (26.3%). The same pattern was found for causal relations (including concession) in Das and Taboada (2019). Specification is often signalled syntactically and/or semantically (47.4%): despite its relatively low complexity (additive positive relation), specification is quite constrained in the type of context in which it can occur. Lastly, contrast is the most signalled relation, with 51.32% of occurrences presenting some syntactic or semantic signal in addition to the DM, which may suggest a higher degree of complexity. In other words, cognitive complexity only partially explains the distribution of strong predictors, since relations differ within a degree of complexity (addition vs. specification; concession vs. contrast).

Moving on to the genre hypothesis, we can refine these results by breaking down the proportion of predictive signals across relations and genres (Figure 6). As a reminder, DMs are expected to be reinforced by predictive signals at a higher rate in formal writing compared with informal speech, with informal writing at an intermediate position. Genre differences in the presence of such signals are not significant for consequence, concession and contrast. Signals are more frequent in essays than in chat conversations for the additive relation (z = 2.195, $p <$ 0.05), but the difference with spoken discussion is not significant. Finally, against our hypothesis, signals for specification are more frequent in discussion than in chat (z = -2.825, p < 0.01) and than in essays (z = -3.049, p < 0.01), which goes in the opposite direction.

INSERT FIGURE 6 HERE

Figure 6. Presence of predictive signals across relations and genres

The hypothesis is therefore not confirmed, with very little variation across genres. Formality or planning pressure thus seem to have little impact on the presence of these predictive signals, which are much more influenced by the type of relation. I will now turn to the last potential factor on signalling variation, namely the ambiguity or strength of the DM in the relation.

4.2 Configurations vs. signals across degrees of DM strength

In this section, I will refer to the log scores of DM strength measured in Section 3.2. Given the null effect of genre identified in the previous section, the data will be presented for the whole corpus. Tables 8 to 12 are summaries of the major configuration features and the presence of predictive signals by DM (with different strength scores) for each relation. The DMs are ranked by increasing marking strength in the tables. As a reminder, it is expected that typical configurations and predictive signals will co-occur more frequently with weak DMs in order to reinforce the interpretation of the relation, whereas stronger DMs will be more independent from these signals. More specifically, the rate of "no signal" (i.e. no relation-specific predictive signal in the context) should increase from top to bottom of each table (as the log score increases); positive polarity is expected to increase with DM strength in concession and contrast; full units should be more frequent with stronger additive and contrastive DMs, as dependent (non-full) units may act as signals for weak DMs in these relations; the rate of "no construction" should increase with DM strength; finally, AltLex should be less frequent with stronger DMs. The hypothesis of DM strength is confirmed for most relations. The statistical significance of differences has been measured through a *z*-score test.

Table 8. Summary of configurations and signal rates (%) for additive DMs

	Log score	**no signal**	**positive**	**full units**	**no constr.**	**AltLex**
but	-1.3979	0	38.46	76.92	23.08	46.15
and	0.6785	90.37	73.75	75.75	94.68	2.66
also	0.7642	95.16	69.35	96.77	96.77	6.45
plus	0.7649	100	79.17	100	91.67	0

For the relation of addition, there is no effect of DM strength on the presence of predictive signals. Only *but* expresses addition in the highly constrained and rare *not only… but also* construction, which could be considered as a complex DM in itself. The only notable variation lies in the higher proportion of full units with stronger DMs such as *also* and *plus*.

Table 9. Summary of configurations and signal rates (%) for specification DMs

	Log score	**no signal**	**positive**	**full units**	**no constr.**	**AltLex**
and	0.3522	50.55	72.69	93.73	67.16	9.23
actually	0.8202	64.52	67.74	96.77	74.19	6.45
for example	1.3159	58.33	62.5	100	100	29.17

For specification, the rate of signals is higher for the two stronger DMs. Moreover, there are fewer constructions with *for example* than with *and* or *actually*, which shows that this strong syntactic pattern is mostly used as a reinforcing signal for weaker DMs, while the stronger DM can signal specification on its own.

Table 10. Summary of configurations and signal rates (%) for consequence DMs

	Log score	no signal	positive	full units	no constr.	AltLex
and	-0.7212	50.00	76.00	90.00	92.00	40.00
then	0.8704	36.67	61.67	100	98.33	5.00
so	0.9609	90.57	55.35	97.48	96.23	10.06
therefore	0.9689	95.83	20.83	75.00	91.67	4.17
thus	0.9694	86.67	60.00	53.33	100	6.67

In consequence relations, there is a significant difference between the proportion of signals with *and* and with stronger DMs in the expected direction (except *then*, which is very often signalled by the co-occurring DM *if* in the *if... then* construction). This means that *and* is reinforced by an AltLex verb when it expresses consequence, but this is not the case for stronger DMs like *so* or *therefore*. In fact, the difference in the proportion of AltLex is also highly significant between *and* on the one hand and stronger DMs on the other.

Table 11. Summary of configurations and signal rates (%) for concessive DMs

	Log score	no signal	positive	full units	no constr.	AltLex
and	-0.79588	66.67	39.74	79.49	98.72	16.67
actually	0.4548	86.79	62.26	94.34	98.11	1.89
while	0.4594	80.39	49.02	98.04	100	5.88
however	0.5647	81.72	56.99	100	94.62	11.83
on the other hand	0.6294	38.89	61.11	94.44	88.89	0
but	0.6493	90.67	51.33	97.33	97.33	4.00
though	0.6884	90.00	51.67	93.33	100	3.33
although	0.6963	86.79	62.26	93.33	93.33	8.89
yet	0.6963	63.16	73.68	78.95	100	10.53
even if	0.6963	86.54	53.85	94.23	96.15	3.85

The pattern is similar for concession, where *and* is significantly more often reinforced by predictive signals than stronger concessive DMs, with only two exceptions (*on the other hand* and *yet*). Furthermore, *and* differs from other markers of concession with a smaller proportion of positive units and a higher proportion of AltLex. This suggests that stronger concessive DMs are more independent from their context than *and*, since they are less associated with negation and are less lexically reinforced.

Table 12. Summary of configurations and signal rates (%) for contrastive DMs

	Log score	no signal	positive	full units	no constr.	AltLex
and	-0.2291	30.77	41.03	56.41	79.49	12.82
but	0.5587	56.98	42.46	81.01	90.5	6.15
on the other hand	0.7356	33.33	100	100	100	0
while	0.8663	11.76	76.47	94.12	82.35	29.41
however	0.9987	54.55	33.33	90.91	84.85	9.09

whereas	1.5671	37.50	72.73	93.94	78.79	9.09

Lastly, for contrast, the hypothesis is not met for all DMs. Cases without predictive signals are significantly less frequent with *and* (31%) than with *but* or *however* (around 55%), as expected. The difference is also significant between *and* and *while* but in the opposite direction (more signals with *while*), although contrastive uses of *while* are quite rare (N=17), so that it is difficult to generalize over this data. However, there is no difference between *and* and the stronger DMs *on the other hand* and *whereas* (around 35% of "no signals"). *Whereas* is typically reinforced by antonyms and/or parallelism, as in (13). The main difference between *and* and *whereas* lies in the proportion of positive polarity, which is significantly higher in *whereas* (as in strong DMs of concession).

(13) those southern states will just have one general view **whereas** the northern states will have another view

It appears from this systematic overview that there are quantitative and qualitative differences in the frequency and types of signals that co-occur with DMs with different scores of strength. To summarize on the factors impacting the presence or absence of predictive signals in the context of DMs, a logistic regression model (acceptable predictive power, C-index = 0.73) was run on the data and returns the following main effects:

i. the likelihood of predictive signals decreases with higher log-transformed strength scores ($\beta = -0.81$, $SE = 0.11$, $p < 0.001$)
ii. the likelihood of predictive signals increases with all relations compared to addition ($p < 0.05$ to $p < 0.001$)

Overall, the study confirms that stronger DMs are more independent from contextual configurations and from predictive signals compared to relations expressed by *and*, which shows that the low informativeness of *and* is compensated in context. Apart from the presence of predictive signals, other contextual features such as negative polarity (for contrast and concession), alternative lexicalizations or type of unit were also shown to be more associated with weak DMs.

5. Discussion: limitations of the present approach to co-occurring signals

This study tested the impact of three potential factors on the co-occurrence of DMs with other types of discourse signals, namely the complexity of the relation, the genre and the strength of the DM. The first two of these factors were found to only have a null or moderate effect in our data, despite hypotheses drawn from previous studies. In this section, I discuss tentative explanations for these null results and relate them to some limitations of our approach.

The type of coherence relation, and in particular the cognitive complexity of the relation, was first expected to be associated with the presence of discourse signals in such a way that more complex relations would co-occur more frequently with signals, in order to ensure their interpretation. The hypothesis was mainly based on previous research showing that negative relations are more complex than positive ones, as attested by their later acquisition (Evers-Vermeul & Sanders, 2009) and higher rate of explicit translation (Hoek et al., 2017; see also Xu et al., 2015 for evidence from comprehension experiments). In our data, this hypothesis was partially confirmed: there were indeed strong differences between relation types, but internal differences between relations with a similar degree of complexity (viz. contrast vs. concession, both negative relations) were observed.

One possible explanation for this relates to the fact that all relations in the present study already contained an explicit DM: because relations are already signalled by the DM, additional signals are less crucial and only act as reinforcing cues, rather than main triggers for the interpretation. By contrast, studies such as Hoek et al.'s (2017) compare relations with and without an explicit DM, in which case the gap in terms of informativeness is perhaps larger, and hence more affected by cognitive complexity. The restriction to relations already expressed by a DM might therefore explain why the complexity hypothesis was not supported by the present study. Another possibility is that signals pertaining to syntax and semantics are not affected by cognitive factors because they form the core of sentences, as opposed to the optionality of DMs: using a DM when the sentence would be correct without one may thus be more impacted by complexity than other signals such as polarity or coreference chains, which cannot be "removed" from the sentence.

The low frequency of signals in consequence and concession relations (both causal) in our data, compared to specification and contrast (both additive relations), further suggests that the relationship between cognitive complexity and the "basic operation" of the relation (additive vs. causal; Sanders et al., 1992) is a complex one. Causal relations are logically more complex, but their processing is largely automatic and was found to be faster than additive relations (Sanders & Noordman, 2000). Furthermore, Hoek et al. (2017) found no difference in the rate of implicitation (presence vs. absence of a DM in translation) between additive and causal relations. These findings suggest that basic operation is only loosely related to cognitive complexity, while negative polarity has repeatedly been related to processing difficulty (e.g. Xu et al., 2015). The present study, however, showed that the marking of coherence relations is more impacted by basic operation than by polarity, with causal relations less often reinforced by non-DM signals. This may be because causal relations such as consequence and concession rely on world knowledge and general inferencing mechanisms, whereas "additive" relations such as contrast or specification strongly rely on the identification of semantic relationships (antonymy and hyponymy, respectively) between the connected segments. Such local semantic links are quite likely to be lexicalized. Therefore, it appears that the basic operation of relations (additive vs. causal) could explain their signalling tendency more than their polarity (negative vs. positive). This finding confirms Das and Taboada's (2019) observations, although methodological differences prevent further comparison.

Secondly, the hypothesis on genre variation was not confirmed either, since the presence of signals was relatively stable across genres or did not vary in the expected direction for most relations. While it is undeniable that speech and writing are different when it comes to syntax and discourse (e.g. Biber, 2006; Chafe, 1982), the particular effect of genre (here, different modalities and formality degrees) on discourse signals seems less prevalent. All texts in the corpus were produced by the same group of participants, with a very similar sociolinguistic profile, and talked about the same six topics in a very argumentative manner. It may well be that different patterns of discourse signalling will emerge in other types of discourse (e.g. narrative, explanatory) and across different speaker profiles. In addition, the focus on relations already expressed by a DM might further explain the absence of a genre effect, as I already suggested for the complexity effect: genre might be responsible for the presence or absence of DMs, but other (non-optional) signals are more central to the sentences and not subject to genre variation.

Overall, restrictions in the scope of the study (explicit relations only, highly homogeneous data) provide some explanations for the negative results of the analysis. They call for further investigation, for instance into the different marking of additive vs. causal relations (see Das & Taboada, 2019), and once more illustrate how complex and multivariate discourse phenomena can be.

6. Conclusion

This study provided a comprehensive portrait of the typical configurations of a set of coherence relations (i.e. addition, specification, consequence, concession and contrast) in English. It distinguished between, on the one hand, features that are relatively more frequent in some relations but present in all of them (e.g. negative verb polarity), and on the other, features that are specific to a given relation (e.g. antonyms for contrast). Similarly, DMs were not treated as a uniform category but ranked according to their "strength", that is, their use in single vs. multiple relations. The traditional divide between explicit and implicit relations has thus been refined through considerations of "signal strength", applied to both DMs and other devices.

This study further revealed that genre had no effect on the distribution of signals, while there was a main effect of DM strength. In particular, the underspecified DM *and* tends to be used in contexts where the coherence relation is already expressed through other signals. By contrast, stronger DMs are more self-sufficient and do not require a very explicit context. This finding corroborates cognitive accounts of language production, such as Uniform Information Density (Levy & Jaeger, 2007) or Rational Speech-Act theory (Frank & Goodman, 2012), which suggest that speakers tend to distribute information evenly in order to avoid (or compensate for) ambiguity. The ambiguity of the DM thus prevails over genre in explaining the presence of co-occurring signals. As for the effect of relation type, the study sheds new light on the respective roles of negativity (polarity) and causality (basic operation) in the complexity of coherence relations and suggests that the latter is more predictive of signalling patterns, with fewer signals in causal relations.

The study is innovative in at least two ways. Firstly, the methodology combined systematic coding of a large number of features (regardless of their relevance for the particular coherence relation at stake) and multivariate statistical models, which teased apart relative configurations from predictive signals that are significantly specific to the given relation identified by a conditional inference tree. By covering both description and prediction of coherence relations, this study refines previous accounts of discourse signals. Another innovative element of the study is the inclusion of more than one text genre, and in particular the inclusion of spoken data. The bulk of research on coherence relations and discourse signals focuses on writing (and in particular on press articles), for the very rational reason that most discourse-annotated corpora are only available in this modality. The corpus at hand is representative of a wider range of uses and contextual settings. As a result, more types of potential signals were included in the taxonomy (e.g. pauses, disfluencies) compared to previous models.

Future perspectives are numerous in the field of discourse signalling. First, the present method has yet to be replicated to other coherence relations such as cause or condition, in order to provide their syntactic and semantic portrait. It could also be extended to relations without a DM, as in Das and Tabaoda (2018, 2019). The computational applications of such an endeavor are far-reaching, in terms of automatic identification and sense disambiguation. Another avenue is psycholinguistic in nature and concerns the online processing cost of coherence relations with and without (different types of) discourse signals, as a function of the type of DM (Crible & Pickering, 2020). Such an endeavor is not trivial since, as Das and Taboada (2018: 767) already observed, it involves manipulating the intricate syntax and semantics of utterances, without modifying their pragmatic interpretation. I hope that the present study will encourage more research in this direction.

Acknowledgments

This research was supported by the H2020 Marie Skłodowska-Curie Individual Fellowship, grant number 794575.

Notes

1. In this paper, ‘discourse marker’ is used as an umbrella term encompassing connectives (relational devices) and other pragmatic markers (not strictly relational) such as *well* or *I mean*. The present study focuses on markers of coherence relations, yet it is part of a larger project on discourse markers in general, which motivates this terminological decision.

2. Asr and Demberg’s (2012) formula is as follows: $p(r|cue) = \frac{p(cue|r)}{p(cue)} * p(r)$, where *r* is the discourse relation and *cue* is the DM.

3. AltLex expressions differ in this taxonomy from the original AltLex in the PDTB 2.0 (Prasad et al., 2008), which mainly correspond to phrases such as *The reason for this is*. It is here extended to any word or phrase that explicitly relates to the meaning of the relation. It corresponds to what Das and Taboada (2018) term ‘indicative word’.

Crible, L. 2022. The syntax and semantics of coherence relations. From relative configurations to predictive signals. *International Journal of Corpus Linguistics* (online first).

References

Asr, F., & Demberg, V. (2012). Measuring the strength of linguistic cues for discourse relations. In E. Hajičová, L. Poláková, & J. Mírovský (Eds.), *Proceedings of the COLING Workshop on Advances in Discourse Analysis and its Computational Aspects (ADACA)* (pp.33–42). The COLING 2012 Organizing Committee.

Biber, D. (2006). *University Language: A Corpus-based Study of Spoken and Written Registers*. John Benjamins.

Blakemore, D. (1987). *Semantic Constraints on Relevance*. Blackwell.

Cain, K., & Nash, H. (2011). The influence of connective on young readers' processing and comprehension of text. *Journal of Educational Psychology*, *103*(2), 429–441.

Chafe, W. (1982). Integration and involvement in speaking, writing and oral literature. In D. Tannen & R. Freedle (Eds.), *Spoken and Written Language* (pp. 83–113). Academic Press.

Clark, H. (1996). *Using Language*. Cambridge University Press.

Crible, L. (2017). Discourse markers and (dis)fluency in English and French. Variation and combination in the DisFrEn corpus. *International Journal of Corpus Linguistics*, *22*(2), 242–269.

Crible, L., & Degand, L. (2019). Domains and functions: A two-dimensional account of discourse markers. *Discours*, *24*. https://doi.org/10.4000/discours.9997

Crible, L., & Demberg, V. (2020). When do we leave discourse relations underspecified? The effect of formality and relation type. *Discours*, *26*. https://doi.org/10.4000/discours.10848

Crible, L., & Pickering, M. J. (2020). Compensating for processing difficulty in discourse: Effect of parallelism in contrastive relations. *Discourse Processes*, *57*(10), 862–879.

Cuenca, M. J. (2013). The fuzzy boundaries between discourse marking and modal marking. In L. Degand, B. Cornillie, & P. Pietrandrea (Eds.), *Discourse Markers and Modal Particles. Categorization and Description* (pp. 191–216). John Benjamins.

Das, D., & Taboada, M. (2018). Signalling of coherence relations in discourse, beyond discourse markers. *Discourse Processes*, *55*(8), 743–770.

Das, D., & Taboada, M. (2019). Multiple signals of coherence relations. *Discours*, *24*.

Das, D., Taboada, M., & McFetridge, P. (2015). *RST Signalling Corpus, LDC2015T10*. https://catalog.ldc.upenn.edu/LDC2015T10

Evers-Vermeul, J., & Sanders, T. (2009). The emergence of Dutch connectives; how cumulative cognitive complexity explains the order of acquisition. *Journal of Child Language*, *36*, 829–854.

Frank, A., & Goodman, N. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998.

Crible, L. 2022. The syntax and semantics of coherence relations. From relative configurations to predictive signals. *International Journal of Corpus Linguistics* (online first).

Goldstein-Stewart, J., Goodwin, K. A., Sabin, R. E., & Winder, R. K. (2008). Creating and using a correlated corpora to glean communicative commonalities. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008*. European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2008/

Hansen, M.-B. M. (2006). A dynamic polysemy approach to the lexical semantics of discourse markers (with an exemplary analysis of French *toujours*). In K. Fischer (Ed.), *Approaches to Discourse Particles* (pp. 21–41). Elsevier.

Hoek, J., Zufferey, S., Evers-Vermeul, J., & Sanders, T. J. M. (2017). Cognitive complexity and the linguistic marking of coherence relations: A parallel corpus study. *Journal of Pragmatics*, *121*, 113–131.

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 & Cognition*, *25*(2), 275–309.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Shiffrin (Ed.), *Meaning, Form and Use in Context: Linguistic Implications* (pp. 11–42). Georgetown University Press.

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.

Koornneef, A., & Sanders, T. J. M. (2013). Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes*, *28*(8), 1169–1206.

Kunz, K., & Lapshinova-Koltunski, E. (2015). Cross-linguistic analysis of discourse variation across registers. *Nordic Journal of English Studies*, *14*(1), 258–288.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in Neural Information Processing Systems 19: Proceedings of the 2016 Conference* (pp. 849–856). MIT Press.

Liu, Y. (2019). Beyond the Wall Street Journal: Anchoring and comparing discourse signals across genres. In A. Zeldes, D. Das, E. Galani Maziero, J. Desiderato Antonio, & M. Iruskieta (Eds.), *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019* (pp. 72–81). Association for Computational Linguistics. https://aclanthology.org/W19-27.pdf

Mak, P., 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.

Crible, L. 2022. The syntax and semantics of coherence relations. From relative configurations to predictive signals. *International Journal of Corpus Linguistics* (online first).

Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, *33*(1), 128–147.

Murray, J. (1997). Connectives and narrative text: The role of continuity. *Memory & Cognition*, *25*(2), 227–236.

Pander Maat, H. (1999). The differential linguistic realization of comparative and additive coherence relations. *Cognitive Linguistics*, *10*(2), 147–184.

Petukhova, V., & Bunt, H. (2009). Towards a multidimensional semantics of discourse markers in spoken dialogue. In H. Bunt, V. Petuhova, & S. Wubben (Eds.), *Proceedings of the 8th International Conference on Computational Semantics* (pp. 157–168). Tilburg University. https://aclanthology.org/W09-3700.pdf

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)* (pp. 2961–2968). European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2008/

Prasad, R., Webber, B., & Lee, A. (2018). Discourse annotation in the PDTB: The next generation. In H. Bunt (Ed.), *Proceedings of the 14th Joint ACL-ISO Workshop on Interoperable Semantic Annotation* (pp. 87–97). Association for Computational Linguistics. https://aclanthology.org/W18-4700.pdf

Rohde, H., Tyler, J., & Carlson, K. (2017). Form and function: Optional complementizers reduce causal inferences. *Glossa*, *2*(1), Article 53. doi: https://doi.org/10.5334/gjgl.134

Sanders, T. J. M. (2005). Coherence, causality and cognitive complexity in discourse. In M. Aurnague, M. Bras, A. le Droualec, & L. Vieu, *Proceedings of SEM-05, First International Symposium on the Exploration and Modelling of Meaning* (pp. 105–114). https://www.researchgate.net/publication/46669022_Coherence_Causality_and_Cognitive_complexity_in_discourse

Sanders, T. J. M., & Noordman, L. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, *29*(1), 37–60.

Sanders, T. J. M, Spooren, W., & Noordman, L. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, *15*(1), 1–35.

Schiffrin, D. (1987). *Discourse Markers*. Cambridge University Press.

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

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

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*

Crible, L. 2022. The syntax and semantics of coherence relations. From relative configurations to predictive signals. *International Journal of Corpus Linguistics* (online first).

Language Resources and Evaluation (LREC 10) (pp. 2084–2090). European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2010/

Webber, B. (2013). What excludes an alternative in coherence relations? In A. Koller, & K. Erk (Eds.), *Proceedings of the 10th International Workshop on Computational Semantics (IWCS2013)* (pp. 276–287). Association for Computational Linguistics. https://aclanthology.org/W13-01.pdf

Xu, X., Jiang, X., & Zhou, X. (2015). When a causal assumption is not satisfied by reality: Differential brain responses to concessive and causal relations during sentence comprehension. *Language, Cognition and Neuroscience, 30*(6), 704–715.

Address for correspondence

Ludivine Crible
Linguistics Department
Ghent University
Blandijnberg
Ghent, 9000
Belgium

ludivine.crible@ugent.be