ISO-DR-core Plugs into ISO-dialogue Acts for a Cross-linguistic Taxonomy of Discourse Markers

Purificação Silvano
University of Porto and CLUP
Via Panorâmica, s/n
4150-564 Porto, Portugal
`msilvano@letras.up.pt`

Mariana Damova
Mozaika Ltd
Solunska 52
Sofia 1000, Bulgaria
`mariana.damova@mozajka.co`

Abstract

The present paper[1] proposes an interoperable taxonomy to represent the meaning of discourse markers based on ISO DR-core (ISO 24617-8) but with a plug-in to ISO-dialogue acts (ISO 24617-2). The proposed taxonomy encompasses two dimensions: the semantic, with values regarding the discourse relations signalled by discourse markers, and the pragmatic, with values concerning the communicative function realized by discourse markers. We present a proof of concept for this two-dimensional taxonomy in a multilingual parallel dataset in three languages, English, European Portuguese and Bulgarian, comprising 165 textual segments with multiword discourse makers obtained from publicly available TED Talk transcripts. We show that the two-dimensional taxonomy can successfully annotate cross-linguistically the meaning of discourse markers and discuss linguistic evidence where extension of the proposed taxonomy can be relevant.

1 Background and Motivation

Discourse markers have been largely studied in different languages (e.g. Schiffrin (1987); Fraser (1996); Knott and Dale (1993); Silvano (2010); Taboada (2006); Das (2014); Mendes et al. (2018); Stede et al. (2019), among others) due to their relevance in discourse interpretation and, simultaneously, to their complexity regarding their multifunctional nature. Some of these studies have rendered several taxonomies within different theoretical frameworks, some language independent, others - language specific, many associated to discourse relations taxonomies (eg. Mann and Thompson (1988); Sanders et al. (1992); Asher et al. (2003); Prasad et al. (2008); Zeyrek et al. (2018)), and most directed to written discourse (cf. eg. for spoken discourse González (2005); Maschler and Schiffrin (2015); Crible (2014)).

Bearing in mind, on the one hand, the diversity of frameworks described and, on the other hand, the usefulness of establishing comparisons between annotated data in the same language and across languages, there have been some efforts to reconcile different taxonomies, such Benamara and Taboada (2015) and Sanders et al. (2021). One of those unifying proposals has resulted in the *Semantic annotation framework (SemAF) — Part 8: Semantic relations in discourse, core annotation schema (DR-core) – ISO 24617-8* (Bunt and Prasad, 2016; Prasad and Bunt, 2015). ISO 24617-8 (ISO, b) stipulates an interoperable core-annotation scheme for low-level discourse relations, i.e., local dependencies. Although the aforementioned aggregating schemes are designed for annotating discourse relations, since these can be explicitly marked by discourse markers that act as cue words/ expressions to infer the proper relation of meaning, it is assumed that they can also be used to represent discourse markers semantics/pragmatics. There are, however, research that design discourse markers-oriented taxonomies experimenting in more than one language, as is the case of Crible and Zufferey (2015).

Regardless of the theoretical approach, the uni-

---

fying taxonomies lack a wide-range application to corpora across languages, genres and types of discourse to test their reliability and comprehensiveness. Regarding multilinguality, ISO (b) states that "a future part of ISO 24617 is envisaged that will complement this document by providing a complete interoperable annotation scheme for DRels (discourse relations), while also addressing the multilingual dimension of the standard", but it has not been published so far. In what concerns written and oral discourse, Crible and Degand (2019), for example, observe that "these interoperable schemes either target written corpora or the relational meanings of spoken DMs, while specific (non-relational) spoken functions still lack a similar unifying approach to date".

The taxonomy of discourse markers put forward in this paper addresses these two types of shortage. On the one hand, by combining ISO DR-core (ISO 24617-8) with ISO-dialogue acts (ISO 24617-2), we can represent not only the semantic meaning of discourse markers (or their relational meanings, as described by Crible and Degand (2019)) with the values of discourse relations but also their pragmatic meaning (or non-relational meaning, as proposed by Crible and Degand (2019)), making use of communicative functions. On the other hand, by applying to a multilingual dataset, which will eventually be published, we demonstrate to what extent the taxonomy is truly interoperable.

2 Related Work

One can opt for narrower and broader notions regarding discourse markers. For instance, Schiffrin (1987) presents "a definition which encompasses both "connectives" (e.g. *and, but, because, actually*) and pragmatic particles more specific to speech (e.g., *well, I mean, you know*). As the author puts it, this is intentionally a vague definition, not to limit the set of discourse markers. Schiffrin (1987) assigns to discourse markers a bracketing role, which Crible and Degand (2019) consider too restricting.

Schiffrin (1987) describes the multifunctionality of discourse markers distinguishing between (1) ideational structure, with relations between propositions, e.g. a cohesion relation, a topic relation or a functional relation; (2) action structure, which describes the organisation and constraints on the use of speech acts; (3) exchange structure, which is "the outcome of decision procedures by which speakers

alternate sequential roles and define those alternations in relation to each other" (Schiffrin, 1987). The author argues that discourse markers may simultaneously have roles within these three structures. Other authors have discussed the multifunctionality of discourse markers. Hovy (1995) considers that discourse markers convey rhetorical structure, interpersonal/ intentional structure, semantic structure, stylistic variants and guidance information. Additionally, CribleDegand+2019+71+99 put forward an annotation taxonomy of discourse markers in spoken language featuring two independent layers of semantic-pragmatic information, domains and functions. The four domains are the follwing: ideational, rhetorical, sequential or interpersonal. The model includes 15 functions (eg. addition, contrast), some based on Prasad et al. (2007)). They have tried the model in different languages (French, English, Polish, Spanish) and modalities (spoken, written, signed), attesting to their reliability and suitability for cross-lingual analysis.

Petukhova and Bunt (2009) also prove with corpus analysis that discourse markers can have multiple meanings concurrently because one dialogue act can serve several goals simultaneously. These authors adopt an empirically-based and formal approach to the semantic functions of discourse markers in dialogue capable of capturing their multifunctional nature. Within the semantic framework of Dynamic Interpretation Theory (Bunt et al., 2020), they propose a multilayered and multidimensional taxonomy with a set of communicative functions, which was the precursor of the *Semantic annotation framework (SemAF) — Part 2: Dialogue acts*, ISO 24617-2 (ISO, a), an interoperable dialogue act annotation framework with dimensions, communicative functions and qualifiers to annotate dialogue acts.

Besides the part that deals with dialogue acts, ISO 24617 comprises part 8 (ISO, b), which stipulates an interoperable core-annotation scheme for low-level discourse relations, i.e., local dependencies, according to the meaning of the relation's arguments. Despite having been designed to annotate discourse relations, ISO 24617-8 has, nevertheless, been used to develop discourse markers lexicon such as PDTB (Prasad et al., 2008), Lex-Conn (Roze et al., 2010), LDM-PT (Mendes et al., 2018), but always taken as triggers of discourse relations.

To sum up, in the face of the diversity of frame-

works described, on the one hand, and, on the other hand, the usefulness of establishing comparisons between annotated data in the same language and across languages, there have been some efforts to reconcile different taxonomies, and at the same time, there have been some proposals to develop an overarching model for discourse markers annotation. Some of those taxonomies can be used to annotate the meaning of discourse markers, but only a few are specifically designed for that purpose. Moreover, none attempts to use ISO standards that can capture both the semantic and pragmatic meaning of discourse markers. Furthermore, most discourse markers-oriented taxonomies lack a wide-range application to corpora across languages, genres and types of discourse to test their reliability and comprehensiveness.

Considering what has already been done and what could be done to contribute to a better understanding of discourse markers, we propose a comprehensive interoperable discourse markers taxonomy able to represent not only the semantic meaning of discourse markers but also their pragmatic meaning, and we determine its reliability by applying it to a sample of a multilingual dataset.

3 The ISO-based Unifying Taxonomy

In our proposal, we assume that discourse markers subsume words or expressions that link utterances and play different pragmatic functions (Schiffrin, 1987; Fraser, 2009; Crible, 2014). Thus, we include in this group - connectives (*as a consequence, on the one hand*) and pragmatic particles (*you know, I mean*). As is well established in the literature, we assume discourse markers to be multifunctional in the sense that they can have, in some contexts, different semantic and pragmatic meanings and also that they can have multiple meanings simultaneously (Petukhova and Bunt, 2009).

We propose an ISO-based unifying taxonomy of discourse markers to annotate both written and spoken discourse cross-linguistically. We adopt the set of core discourse relations provided by ISO 24617-8 (ISO, b), which was defined on the grounds of different theoretical approaches and annotation endeavours. According to this framework, the discourse relations are of two types: symmetric, in which case the two arguments assume relation-specific semantic role, and asymmetric, when the arguments take the same semantic role. The discourse relations are used to ascertain the semantic

meaning of discourse markers such as "as a result of" (Cause) (cf. ex.(1)), "for example" (Exemplification) (cf. ex. (2)).

(1) It turns out that rarely do we practice under the types of conditions we're actually going to perform under, and **as a result**, when all eyes are on us, we sometimes flub our performance.

(2) Ah, earth's oceans. They are beautiful, inspiring, life-sustaining. They are also, as you're probably quite aware, more or less screwed. In the Seychelles, **for example**, human activities and climate change have left corals bleached. Overfishing has caused fish stocks to plummet.

Notwithstanding, not all discourse markers convey a relational meaning, and instead play an interactional function, not accounted for by ISO 24617-8. It should be noted that this part of the SemAF admits pragmatic variants of discourse relations (Bunt and Prasad, 2016), that is, for each discourse relation, there is the possibility of one or both arguments expressing an implicit belief or a dialogue act. In those instances, the relevant arguments, and not the discourse relations, are annotated with that information because, according to the authors, the inference of a belief or a dialogue act depends on the arguments, and not on the discourse relation. This distinction is not, however, relevant for our taxonomy, since we aim at a typology which encodes the meaning of the discourse marker and not the nature of the discourse relation. To properly represent the interactional (or pragmatic) meaning of some discourse markers, we deemed it best to add an annotation plug-in to Semantic annotation framework (SemAF) — Part 2: Dialogue acts (ISO, a), (Bunt, 2019), (Bunt et al., 2020). This mechanism is introduced by Bunt (2019) and Bunt et al. (2020) with the inverse direction, from ISO 24617-2 to ISO 24617-8, to solve the problem of annotating semantic content of dialogue acts. In our taxonomy, we utilize the plug-in to overcome the limitations of the discourse relations set in ISO DR-core, enabling the encoding of the pragmatic meaning of discourse markers such as *you know*, which can convey the communicative function *Opening* (cf. ex.(3)), and *of course*, which expresses certainty, hence the qualifier *Certain* (cf. ex.(4)). Although the meta-model designed for ISO 24617-2 involves dimensions, communicative functions and qualifiers, for our taxonomy the last two suffice.

(3) (Applause) Lakshmi Pratury: Just stay for a second. Just stay here for a second. (Applause) **You know**, when I heard Simon's – please sit down; I just want to talk to him for a second –

(4) You've dissolved the barrier between you and other human beings. And this, **of course**, is the basis of much of Eastern philosophy

Table 1 summarizes the different values for each dimension.

Accordingly, there are discourse markers with a semantic dimension that receive one of the values from the first column. The discourse markers with a pragmatic dimension can be assigned a general communicative function (first column from the pragmatic dimension) or a more specific communicative function (second column from the pragmatic dimension), as discussed in example (3) above. Their interpretation may require an additional value related to notions of certainty, conditionality, and sentiment, like in examples (5), where the discourse marker plays a communication function *confirm*, in addition to carrying a value represented by the qualifier *Certain*. The multifunctional nature of discourse markers is evidenced by example (6), where the discourse marker *of course* has, concurrently, a semantic and pragmatic value, signalling the discourse relation *Expansion* and having the communication function *Confirm* and the qualifier *Certain*.

(5) And that is, there is a sudden emergence and rapid spread of a number of skills that are unique to human beings like tool use, the use of fire, the use of shelters, and, **of course**, language, and the ability to read somebody else's mind and interpret that person's behavior.

(6) Instead, so far, the measurements coming from the LHC show no signs of new particles or unexpected phenomena. **Of course**, the verdict is not definitive. In 2015, the LHC will almost double the energy of the colliding protons,

We acknowledge that both the semantic and pragmatic dimensions of the annotation scheme we propose can be in themselves multi-dimensional[2]. However, although a text span can convey more

---

[2]This observation was made by one reviewer, to whom we thank.

than one communicative function and/ or be linked to another by more than one discourse relation, the same is not as frequent with discourse markers. In other words, the same discourse marker can be assigned different communicative functions and discourse relations in different contexts, but, as we will demonstrate in the next section, the concurrence of two semantic meanings or two communicative functions in the same discourse marker in the same context is rarely observed in our annotation framework and data.

4 The Proof of Concept

With the goal of determining the reliability and coverage of the proposed taxonomy, we devised a short experiment with a dataset of 165 multiword discourse makers occurrences in three languages, English, European Portuguese and Bulgarian. We selected multiword expressions because we have also been working on cross-lingual and language-agnostic methods for discourse markers prediction, and multiword discourse markers pose relevant problems when dealing with automatic detection. The data for this experiment were extracted from publicly available TED Talk transcripts. They represent a subset from a larger parallel multilingual corpus covering English, European Portuguese, Lithuanian, Bulgarian, German, Macedonian, Hebrew, Romanian, Italian and Polish, where English has been established as a pivot language for all language pairs of the dataset. A baseline annotation was performed by a linguist for the English data. Whenever necessary, annotation decisions were discussed in the working group. After establishing the gold standard, an annotation manual was created. While all languages have been annotated, we present evidence from three of them in this paper. Table 2 illustrates the result of applying the taxonomy to the three datasets.

Table 2 reveals that ISO 24617-8 adequately represents the meaning of most of the discourse markers found in the three datasets. However, the plug-in to ISO 24617-2 enables a more suitable classification of a group of discourse markers, even if they are few. A very small number of discourse markers can be classified using both dimensions (*of course*, *de facto*, разбира се)

In the set of 165 multiword discourse makers occurrences in three languages, English, European Portuguese and Bulgarian, we observed that the majority of the discourse markers convey a seman-

Table 1: Taxonomy of discourse markers.

Semantic dimension	Pragmatic dimension		
Cause	CheckQuestion	AutoNegative	conditional
Expansion	Inform	AlloPositive	unconditional
Asynchrony	Agreement	AlloNegative	certain
Concession	Disagreement	FeedbackElicitation	uncertain
Elaboration	Correction	Stalling	positive
Exemplification	Answer	Pausing	negative
Manner	Confirm	InteractionStructuring	
Condition	Disconfirm	Opening	
Negative Condition	Offer	TopicShift	
Purpose	Promise	SelfError	
Exception	AddressRequest	Retraction	
Substitution	AcceptRequest	SelfCorrection	
Conjunction	DeclineRequest	InitGreeting	
Contrast	AddressSuggest	InitSelfIntroduction	
Synchrony	DeclineSuggest	Apology	
Similarity	Request	Thanking	
Disjunction	Instruct	InitGoodbye	
Restatement	Suggest	Compliment	
	AddressOffer	Congratulation	
	AcceptOffer	SympathyExpression	
	DeclineOffer	ContactCheck	

tic meaning represented by nine different discourse relations, which are *Exemplification, Elaboration, Synchrony, Contrast, Concession, Conjunction, Restatement, Cause and Expansion*. The values of *Restatement* - inferred when the discourse marker links two arguments that represent the same situation but from different perspectives (ISO, b) -, and *Expansion* - assigned when the second argument is a situation involving some entity/entities present in the first argument, expanding the narrative or expanding on the setting relevant for interpreting the first argument (ISO, b) -, are, in our dataset, expressed by more multiword discourse markers, at least for English and European Portuguese. Although, in the case of *Restatement*, the discourse markers are variants or have very similar meanings (eg. in Portuguese, *por outras palavras, noutras palavras*), looking at the discourse markers that carry the value of *Expansion*, we can observe, for English and European Portuguese, more lexical variety (eg. *in fact, that is, of course*). In fact, regarding the set of discourse relations, it is not surprising that more specific ones would permit a more fine-grained distinction of the discourse markers semantic value. ISO 24617-8 already assumes that this applies to *Expansion*. It also postulates

that Elaboration subsumes the discourse relation *Summary* proposed by Mann and Thompson (1988). However, discourse marker *sum up* encodes a different meaning when compared to *in particular*, for instance. Other discourse markers such as *in fact*, *de facto*, всъщност would be better represented with a more informative discourse relation, like, for instance, *Affirmation*.

In what concerns the pragmatic dimension, despite the extensive list of communicative functions (cf. Table1), the sample of discourse markers subject to this experiment displays little variety, only four, to be precise. The communicative functions that the discourse markers fulfill are the following: *CkeckQuestion*, used to determine, from the addressee, whether a proposition, which forms the semantic content, is true (ISO, a); *Confirm*, utilized to inform the addressee that the proposition which constitutes the semantic content is true (ISO, a); *Opening*, to show to the addressee that the sender is ready to start the dialogue (ISO, a); and *AlloPositive*, employed to inform the addressee that the sender believes that the addressee is processing what is being said (ISO, a). The fact that the same discourse marker can signal different communicative functions, as is the case of *you know* and its

equivalents in the other three languages, or discourse relations, like *on the other hand* with a *Contrast* and *Concession*), or even simultaneously a discourse relation and a communicative function, like *in fact, de facto*, attests the polyfunctionality of discourse markers. Furthermore, the same discourse marker can carry a communicative function and an additional value, represented in our proposal by qualifiers, which are predicates that can "narrow down the meaning of a communicative function, called restrictive qualifiers, and those that add something to the meaning of a communicative function, called additive qualifiers" (Bunt et al., 2012). In our dataset, we only came across one discourse, *of course*, *claro*, разбира се to which a certainty qualifier (restrictive) was assigned.

Table 2 includes all the cases where the discourse markers translated from English to European Portuguese and Bulgarian have the same semantic and/or pragmatic in the three languages. However, on close inspection, the cross-lingual analysis of the dataset reveals that one and the same English expression gets translated with different expressions conveying distinct meanings. In Bulgarian in different contexts, for example, we encounter правилно, Bulgarian for the English words (*right, correct*), conveying a value of *CheckQuestion* (cf. ex.(7), (8)), and not всъщност in a context where in English *in fact* with the meaning of *Expansion* is used.

(7) и рожденият ден на Лейди Гага. Не ви ли звучат невероятно? Но повечето хора не са съгласни. Правилно, защото техните умове не се вписват, в това което обществото смята за нормално, често биват избягвани и неразбрани.

(8) and Lady Gaga's birthday. Don't they sound incredible? But most people don't agree. And **in fact**, because their minds don't fit into society's version of normal, they're often bypassed and misunderstood.

This leads to considerations that the different translations of the same expression can signal different meanings or communication functions and to the assumption that the thorough cross-lingual analysis can provide insight into the application and the further enrichment of the proposed taxonomy. Further, observation points to the interdependence between some conjunctions with discourse markers. It is not rare to see *in fact* preceded by *and*,

for example preceded by *so*, and much more. Although out of the scope of the present work, these phenomena present interesting evidence related to the classification and identification of the roles of discourse markers in discourse and their representation.

5 Final Remarks

In conclusion, when compared to other proposals, our taxonomy has the following strengths: a) it was specifically designed to codify the meaning of discourse markers; b) the two dimensions, semantic and pragmatic, are featured by values that are specific to those dimensions (and not generic); c) the dimensions-oriented values properly account for the role or roles each discourse marker can play in discourse; d) being the values extracted from parts of ISO 24617, tried out in different genres and text modalities and languages, grants our proposal more reliability and allows for interoperability.

Nonetheless, we still have some work to do. First, we will stabilize the taxonomy by adding more discourse relations to account for pertinent distinctions of meaning, by applying the taxonomy to a larger dataset both composed of monologues and dialogues and by defining a smaller set of relevant communicative functions taking into consideration their occurrence on the corpora. Then we will proceed to large–scale annotation, which means the annotation of the complete corpus using inter-annotator agreement. Finally, we will develop an empirical-based multilingual lexicon of discourse markers to be used as LLOD.

Acknowledgements

References

a. ISO 24617-2. 2020. language resource management-semantic annotation framework (SemAF) - part 2 - dialogue acts. Standard, Geneva, CH.

b. ISO 24617-8. 2016. language resource management, part 8: Semantic relations in discourse (DR-Core). Standard, Geneva, CH.

Nicholas Asher, Nicholas Michael Asher, and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.

Farah Benamara and Maite Taboada. 2015. Mapping different rhetorical relation annotations: A proposal. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 147–152, Denver, Colorado. Association for Computational Linguistics.

Harry Bunt. 2019. Plug-ins for content annotation of dialogue acts.

Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis, and David Traum. 2012. ISO 24617-2: A semantically-based standard for dialogue annotation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 430–437, Istanbul, Turkey. European Language Resources Association (ELRA).

Harry Bunt, Volha Petukhova, Emer Gilmartin, Catherine Pelachaud, Alex Fang, Simon Keizer, and Laurent Prevot. 2020. The ISO standard for dialogue act annotation. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 549–558.

Harry Bunt and Rashmi Prasad. 2016. ISO DR-Core (ISO 24617-8): Core concepts for the annotation of discourse relations. In *Proceedings 12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-12)*, pages 45–54.

Ludivine Crible. 2014. Identifying and describing discourse markers in spoken corpora. annotation protocol v.8. Technical report.

Ludivine Crible and Liesbeth Degand. 2019. Reliability vs. granularity in discourse annotation: What is the trade-off? *Corpus Linguistics and Linguistic Theory*, 15(1):71–99.

Ludivine Crible and Sandrine Zufferey. 2015. Using a unified taxonomy to annotate discourse markers in speech and writing.

Debopam Das. 2014. *Signalling of coherence relations in discourse*. Ph.D. thesis, Arts & Social Sciences: Department of Linguistics.

Bruce Fraser. 1996. Pragmatic markers. *Pragmatics*, 6:167–190.

Bruce Fraser. 2009. An account of discourse markers. *International Review of Pragmatics*, 1(2):293–320.

Montserrat González. 2005. Pragmatic markers and discourse coherence relations in english and catalan oral narrative. *Discourse Studies*, 7(1):53–86.

Eduard Hovy. 1995. The multifunctionality of discourse markers. In *Proceedings of the Workshop on Discourse Markers*. Egmond-aan-Zee, The Netherlands.

Alistair Knott and Robert Dale. 1993. Using linguistic phenomena to motivate a set of rhetorical relations. *Human Communication Research Centre, University of Edinburgh*.

William Mann and Sandra Thompson. 1988. Rethorical structure theory: Toward a functional theory of text organization. *Text*, 8:243–281.

Yael Maschler and Deborah Schiffrin. 2015. Discourse markers: Language, meaning, and context. *The handbook of discourse analysis*, 1:189–221.

Amália Mendes, Iria del Río Gayo, Manfred Stede, and Felix Dombek. 2018. A lexicon of discourse markers for portuguese - ldm-pt. In *LREC*.

Volha Petukhova and Harry Bunt. 2009. Towards a multidimensional semantics of discourse markers in spoken dialogue. In *Proceedings of the Eight International Conference on Computational Semantics*, pages 157–168, Tilburg, The Netherlands. Association for Computational Linguistics.

R. Prasad, Eleni Miltsakaki, Nikhil Dinesh, Alan Lee, Aravind K. Joshi, Livio Robaldo, and Bonnie Lynn Webber. 2007. The penn discourse treebank 2.0 annotation manual.

Rashmi Prasad and Harry Bunt. 2015. Semantic relations in discourse: The current state of ISO 24617-8. In *Proceedings of the 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-11)*, London, UK. Association for Computational Linguistics.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Charlotte Roze, Danlos Laurence, and Philippe Muller. 2010. LEXCONN: a French Lexicon of Discourse Connectives. In *MAD 2010 - 8th Workshop Multidisciplinary Approaches to Discourse*, Proceedings of the 8th Workshop Multidisciplinary Approaches to Discourse (MAD 2010), pages 114–125, Moissac, France.

Ted Sanders, Wilbert Spooren, and Leo Noordman. 1992. Toward a taxonomy of coherence relations. *Discourse Processes - DISCOURSE PROCESS*, 15:1–35.

Ted J.M. Sanders, Vera Demberg, Jet Hoek, Merel C.J. Scholman, Fatemeh Torabi Asr, Sandrine Zufferey, and Jacqueline Evers-Vermeul. 2021. Unifying dimensions in coherence relations: How various annotation frameworks are related. *Corpus Linguistics and Linguistic Theory*, 17(1):1–71.

Deborah Schiffrin. 1987. *Discourse Markers*. Cambridge University Press, Cambridge.

Purificação Silvano. 2010. *Temporal and rhetorical relations: the semantics of sentences with adverbial subordination in European Portuguese*. Ph.D. thesis.

Manfred Stede, Tatjana Scheffler, and Amália Mendes. 2019. Connective-lex: A web-based multilingual lexical resource for connectives. *Discours*.

Maite Taboada. 2006. Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics*, 38(4):567–592. Focus-on Issue: The Pragmatics of Discourse Management.

Deniz Zeyrek, Amália Mendes, and Murathan Kurfalı. 2018. Multilingual extension of PDTB-style annotation: The case of TED multilingual discourse bank. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Table 2: The annotation of discourse markers - illustration.

Discourse markers meaning	English DM	Portuguese DM	Bulgarian DM
Discourse relations ISO 24617-8			
Exemplification	for example, for instance	por exemplo	например
Elaboration	in particular, to sum up	em suma	особено, в частност
Synchrony	so far	até agora	до сега
Contrast	on the one hand	por um lado	от една страна
Concession	on the other hand	por outro lado	от друга страна
Conjunction	on the other hand	por outro lado	от друга страна
Restatement	in other words, I mean	por outras palavras, noutras palavras, isto é	с други думи
Cause	as a result	como resultado, como consequência	в резултат
Expansion	in fact, this is, that is, of course	de facto, na verdade, ou seja, claro	всъщност
Communicative functions and qualifiers ISO 24617-2			
CheckQuestion	you know		знаеш ли, знаете ли
Confirm	of course, in fact	claro, de facto, na verdade	разбира се
Opening	You know	sabem	знаеш ли, знаете ли
AlloPositive	you see		виждаш ли, видите ли
Certain	of course	claro	разбира се