

Enriching a Lexicon of Discourse Connectives with Corpus-based Data

Anna Feltracco^{1,2,3}, Elisabetta Jezek², Bernardo Magnini¹

¹Fondazione Bruno Kessler, Via Sommarive 18, 38100 Trento, Italy

²University of Pavia, Strada Nuova 65, 27100 Pavia, Italy

³University of Bergamo, Via Salvecchio, 19, 24129 Bergamo, Italy

feltracco@fbk.eu, jezek@unipv.it, magnini@fbk.eu

Abstract

We present the results of the effort of enriching the pre-existing resource LICO, a Lexicon of Italian COnnectives retrieved from lexicographic sources (Feltracco et al., 2016), with real corpus data for connectives marking contrast relations in text. The motivation beyond our effort is that connectives can only be interpreted when they appear in context, that is, in a relation between the two fragments of text that constitute the two arguments of the relation. In this perspective, adding corpus examples annotated with connectives and arguments for the relation allows us to both extend the resource and validate the lexicon. In order to retrieve good corpus examples, we take advantage of the existing Contrast-Ita Bank (Feltracco et al., 2017), a corpus of news annotated with explicit and implicit discourse contrast relations for Italian according to the annotation scheme proposed in the Penn Discourse Tree Bank (PDTB) guidelines (Prasad et al., 2007). We also use an extended -non contrast annotated- version of the same corpus and documents from Wikipedia. The resulting resource represents a valuable tool for both linguistic analyses of discourse relations and the training of a classifier for NLP applications.

Keywords: discourse connectives, contrast relation, corpus examples

1. Introduction

Discourse relations and the linguistic elements marking them in text, commonly referred to as *discourse connectives*, have recently been at the core of several annotation efforts for multiple languages (including English, German, French, Italian, Portuguese, see Stede and Umbach (1998), Roze et al. (2012) among others). In this paper, we present the results of the effort of enriching the pre-existing resource LICO, a lexicon of Italian connectives retrieved from lexicographic sources (Feltracco et al., 2016), with real corpus data, thus allowing us to both extend the resource and validate the lexicon. Our goal in this contribution is limited to the class of connectives marking contrast, and the additional relations such connectives might convey, some of them being polysemous. The motivation beyond our effort is that connectives can only be interpreted and disambiguated when they appear in context, that is, in a relation between the two fragments of text that constitute the two arguments of the relation. In order to retrieve good examples, we take advantage of the existing Contrast-Ita Bank (Feltracco et al., 2017), a corpus of news annotated with explicit and implicit discourse contrast relations for Italian according to the annotation scheme proposed in the Penn Discourse Tree Bank (PDTB) guidelines (Prasad et al., 2007). Contrast-Ita Bank contains corpus annotations for 19 discourse connectives of contrasts, 14 of them are included in LICO; these provide the starting point for our work. We pick additional examples from the larger news corpus from which Contrast-Ita Bank is derived. The resulting resource represents a valuable tool for both linguistic analyses and the training of a classifier for NLP applications. The paper is structured as follows. Section 2 reports the definition of discourse connective we assume in our work. Section 3 introduces LICO and related lexica for other languages, while Section 4 reports the methodology and Section 5 presents the results of the effort of enriching the resource. The pa-

per ends with concluding observations and hints for further work.

2. Discourse connectives

We define discourse connectives as lexical markers that are used to express relations between parts of the discourse. This definition is inspired by Ferrari (Ferrari and Zampese, 2000; Ferrari, 2010): she defines a *connective* as “each of the invariable forms [...], which introduce relations that structure “logically” the meanings of the sentence and of the text”¹.

Ferrari clarifies that relations marked by connectives hold between events or assertions, and includes as arguments for the relation also nominalisations (e.g. “after the pressing invitation ...”), i.e. cases that contain an event introduced through a nominal expression. On the other hand, she excludes those grammatical elements that introduce relative clauses or pronouns (as *who* in “I don’t know who you are.”) to be connectives. This is in line with the definition provided for the *arguments* of a *connective* in the Penn Discourse Tree Bank (PDTB) 2.0 project, for which connectives relate two events, states, and propositions, that can be realized mostly as clauses, nominalisations, and anaphoric expressions (Prasad et al., 2007). From this group are excluded general *cue phrases* or *discourse markers*, words/phrases that do not have the function of connectives but are used for instance to change the topic in a discourse or to initialize it, such as “but” in “But, what are you doing?”. According to Ferrari (2010), connectives belong to different syntactic classes, the same defined in the PDTB schema: i) subordinating conjunctions or subordinating expressions; ii) coordinating conjunctions or coordinating ex-

¹Original text: “Il termine *connettivo* indica in linguistica ciascuna delle forme invariabili [...], che indicano relazioni che strutturano ‘logicamente’ i significati della frase e del testo”(Ferrari, 2010).

pressions; iii) adverbs or adverbial expressions; iv) prepositions or prepositional expressions.

In line with this definition, Stede (2012) distinguishes connectives as never inflected, closed-class lexical items, which belong to the above mentioned syntactic categories. He also specifies that these lexical elements can only be interpreted successfully when they appear in a relation between two discourse segments.

Ferrari (2010) also proposes a non hierarchical classification for connectives depending on the “type of logical relation they convey”, e.g. *temporal* and *causal*. The PDTB 3.0 project (Webber et al., 2016) proposes a hierarchical classification composed by three levels (Table 1).

I level	II level	III level	
CLASS	TYPES	SUBTYPES	
TEMPORAL	Synchronous	–	
	Asynchronous	Precedence Succession	
CONTINGENCY	Cause	Reason Result	
		Condition	Arg1-as-cond Arg2-as-cond
	Negative Condition	Arg1-as-negcond Arg2-as-negcond	
		Purpose	Arg1-as-goal Arg2-as-goal
	COMPARISON	Contrast	–
		Similarity	–
		Concession	Arg1-as-denier Arg2-as-denier
EXPANSION	Conjunction	–	
	Disjunction	–	
	Equivalence	–	
	Instanciatiion	–	
	Level-of-detail	Arg1-as-detail Arg2-as-detail	
		Substitution	Arg1-as-subst Arg2-as-subst
	Exception	Arg1-as-exception Arg2-as-exception	
		Manner	Arg1-as-manner Arg2-as-manner

Table 1: The PDTB 3.0 hierarchy of relations (Webber et al., 2016).

In the first level of the hierarchy, the *class level*, sense tags are grouped in four major classes (first column of Table 1). The second level of the hierarchy (second column of Table 1) specifies further the semantics of the class level: the *type level*. For example, the tag *TEMPORAL.Synchronous* indicates the type *Synchronous* of the class *TEMPORAL* and is used for connectives that indicate that the arguments of the relation are simultaneous (e.g. “When she arrived, he was leaving”); differently, the *TEMPORAL.Asynchronous* tag is used when the connective indicates a before-after relation (e.g. “She arrived before he left”). The third level, the *subtype level* (third column of Table 1), reflects the *direction of the relations*. For example, the type *CONTIN-*

GENCY.Cause represents an asymmetric relation between two arguments: being one the cause, and the other the result. The subtype *CONTINGENCY.Cause.Reason* is used if the argument introduced by the connective (Arg2) is the reason for the situation in the other argument (Arg1) (e.g. “I stayed at home because it was raining”), while *CONTINGENCY.Cause.Result* is used if it represents the result/ef-
fect (e.g. “It was raining, therefore I stayed at home”). Notice that not every *type* has a further *subtype*: for example, the arguments involved in a temporal relation of type *Synchronous* do not play different roles and no subtype has been proposed.

3. LICO: Lexicon for Italian Connectives

According to our knowledge, LICO (Feltracco et al., 2016), Lexicon for Italian Connectives, is the highest coverage resource of discourse connectives available for Italian.

Connectives in LICO. In LICO connectives are listed together with orthographic, syntactic, semantic information and also possible alignments with lexica of connectives in other languages. LICO is organized in 173 entries, each one corresponding to a connective and its orthographic or lexical variants. In fact, the invariability criterion proposed by Ferrari (2010) which does not include variable forms (i.e. those forms which are subject to morphological modifications) is partially dropped in LICO. Specifically, the resource does not include forms which exhibit morphological inflection or conjugation, but includes connectives which show a certain degree of lexical variability, that is, multi-word expressions which are not totally rigid from a lexical point of view (e.g. *ad esempio/per esempio* ‘for example’ are both registered in the resource, as two variants of unique entry).

Connectives in LICO are retrieved from three sources: i) the list of connectives mentioned by Ferrari for the entry *connettivi*², ii) the list of connectives tagged as *congiunzione testuale* in Sabatini Coletti 2006 (Sabatini and Coletti, 2007), except for the ones of literary use, and iii) the list of the equivalent Italian terms of the German connectives in the DimLex resource (Stede, 2002) (see Related Lexica).

LICO Structure. For each entry LICO specifies:

- possible lexical variants (e.g. *dopo di ché* and *dopo di ciò*) and orthographic variants (e.g. *ciò nonostante* and *ciononostante*);
- whether the connective (or its variants) is composed by a single token or by more than one token;
- whether the connective is composed by correlating parts (e.g. *da una parte [...] dall'altra*) or not (e.g. *ciononostante*);
- the syntactic category: adverbs, prepositions, subordinating or coordinating conjunctions;
- the semantic relation(s) which the connective indicates, according to the PDTB 3.0 schema of relations (Webber et al., 2016);

²[http://www.treccani.it/enciclopedia/connettivi_\(Enciclopedia-dell'Italiano\)/](http://www.treccani.it/enciclopedia/connettivi_(Enciclopedia-dell'Italiano)/)

- possible alignments with lexica of connectives in German;
- examples of usage of the connective for each semantic relations it indicates.

The examples in the first version of the resource are translation of the German examples already present in the DimLex resource (Scheffler and Stede, 2016; Stede, 2002; Stede and Umbach, 1998). In adopting a corpus-based approach we aim at enriching LICO with data-driven examples and validating the information in the resource.

Related lexica. LICO has been inspired by the DimLex project for German (Scheffler and Stede, 2016; Stede, 2002; Stede and Umbach, 1998)³, an XML-encoded resource that provides information on orthographic variants, syntactic category, semantic relations in terms of PDTB3.0 (Webber et al., 2016) sense tags, and usage examples for 274 connectives. DimLex is used for automatic discourse parsing, and also for semiautomatic text annotation using the ConAno tool (Stede and Heintze, 2004). A similar repository for French is LEXCONN (Roze et al., 2012)⁴, which contains more than 300 connectives with their syntactic categories and discourse relations from Segmented Discourse Representation Theory (Asher and Lascarides, 2003). The lexicon has been constructed manually, using a corpus as empirical support.

LICO is freely distributed under a CC-BY licence⁵ and can be browsed with DIMLEX and LEXCONN at <http://connective-lex.info/>.

4. Enriching Connectives of Contrast in LICO

We aim at enriching the connectives of contrast in LICO with examples from corpora. Collecting these examples, we can observe in context how each connective is used: for instance, if the connective is used at the beginning of a sentence, if it requires the following verb to be in a conjunctive form, etc.

We focus on the connectives signalled in LICO as conveying a contrast relation. As we said, for each entries in LICO, the semantic relation the connectives convey is signalled adopting the PDTB schema of sense tags: we include in our research those that are tagged in LICO with the *COMPARISON.Contrast* tag (e.g. *contrariamente a*) and the *COMPARISON.Concession* tag (e.g. *nonostante*). For convenience, we will refer to these senses as *CONTRAST* and *CONCESSION*.

The entries tagged with one of these two senses (or both) are 38. For each of them, we retrieve 5 examples from a corpus; and, for those that are polysemous, we retrieve 5 examples for each of the senses they convey according to LICO. For example, the connective *mentre* has been tagged with *CONTRAST*, *CONCESSION*, *TEMPORAL* thus we retrieve 5 examples for the connective in these three senses.

This lead us to complete the information about each connective with useful examples of its usage in all its senses.

In order to collect corpus-driven examples, we use two strategies: i) we retrieve examples from the resource Contrast-Ita Bank; ii) we pick examples from the bigger corpus from which Contrast Ita-Bank is derived and from Wikipedia documents.

Connectives and Examples from Contrast-Ita Bank. Contrast-Ita Bank (Feltracco et al., 2017) is a corpus of 169 news (65,053 tokens) annotated with explicit and implicit discourse contrast relations in Italian.

More specifically, the documents correspond to articles published in a local newspaper “L’Adige” in two different days and include reports, news about politics, news about economics, sport results. They contain narrations and quotes of oral interviews.

Contrast-Ita Bank (henceforth CIB) follows the schema proposed in the PDTB guidelines (Prasad et al., 2007) both in terms of sense tags, i.e., *CONTRAST* and *CONCESSION* are tagged in the corpus, and in terms of information annotated, i.e., for each explicit relations, the connective that conveys the relation is marked together with its arguments (named *Arg1* and *Arg2*)⁶. For instance, in Example (1) the connective is underlined, *Arg1* is in italics, and *Arg2* is in bold.

- (1) *Il ministro del Lavoro e delle Pensioni britannico, Andrew Smith, ha rassegnato ieri le dimissioni nonostante i tentativi del premier Tony Blair di convincerlo a rimanere.*⁷
tag: *CONCESSION.Arg1.as.denier*

In our work we take advantage of the information associated to the connectives of contrast in CIB. In fact, the annotated connective (marked as *CONTRAST* or *CONCESSION*) together with its arguments constitute the examples we retrieved for the enrichment of LICO. For instance, Example (1) from CIB has been retrieved as an example of *nonostante* to enrich LICO. Moreover, we can get information of how the connective is used with reference to the relation it conveys; for example, we can inspect if it is found between the arguments it links, before them, if it requires a conjunctive form of the verb just in its *arg2*, etc... We keep this data in LICO by taking care of reporting the span of text of the two arguments as part of the example and by encoding the *token_id* of the connective as registered in CIB: this works as a pointer for users who can reconstruct the entire annotation of the contrast relation in CIB.

Corpus examples picking. Not all the connectives tagged as *CONTRAST* and *CONCESSION* in LICO are present in CIB, and the resource does not provide 5 examples for all the connectives of contrast it contains (some of them appear just once). Moreover, we want to retrieve examples also for the non contrastive use of the connectives, and these are not tagged in CIB. To reach our goal, we extend the search to other documents. On one hand, we con-

⁶Specifically, *Arg2* is the argument that is syntactically bound to the connective, and *Arg1*, the other one (Prasad et al., 2007).

⁷Eng: *Secretary of State for Work and Pensions Andrew Smith resigned yesterday, despite Prime Minister Tony Blair’s attempts to persuade him to stay.*

³<https://github.com/discourse-lab/dimlex/>

⁴<http://www.linguist.univ-paris-diderot.fr/~croze/D/Lexconn.xml/>

⁵<https://hlt-nlp.fbk.eu/technologies/lico>

entry-id	4
orth	cont
part	phrasal
	a dire la verità
orth	cont
part	phrasal
	in verità
orth	cont
part	phrasal
	per la verità
POS	adverbial
sem relation	COMPARISON:Concession:Arg2-as-denier
examples	id="1" source = "CIB_405635_185,186" text: "La nostra idea - dice Rita - è quella di aprire un bar normale. In cui il pubblico possa consumare anche qualche cibo precotto, patatine fritte, poi i soliti panini. Magari ascoltandosi il juke-box". In verità, proprio normale, normale non sarà il bar. id="2" source = "Adige_414186" text: Sulla carta l'opposizione è di dodici consiglieri. In verità siamo rimasti in cinque id="3" source = "wiki_34540_Piario" text: Soltanto nel 1792 si verificò la definitiva divisione di quello che era il comune di "Oltressenda bassa" in Piario e Villa d' Oghna . A dire la verità i due centri vennero nuovamente uniti in un' unica realtà amministrativa durante il periodo fascista (anno 1929) , per poi separarsi in modo definitivo nel 1958 . id="4" source = "wiki_39733_Papa.Pio.XII" text: .. id="5" source = "wiki_1041014.Divisione.Nazionale" text: ..
sem relation	EXPANSION:Level-of-detail:Arg2-as-detail
	example id="1" source = "Adige.413952" text: ..

Figure 1: The connective *a dire la verità* and its variants in LICO.

sidered other 357 documents of the newspaper “L’Adige” (same source of CIB); we will refer to this source as *Adige*. On the other hand, we search for additional examples in documents from Wikipedia⁸. While the retrieving has been done automatically, the selection of the examples has been conducted manually. This is because we need to distinguish cases in which the connective plays such a role from cases in which it does not introduce a discourse relation. These latter cases are also known as *discourse markers* already mentioned in Section 2, and are used, for instance, to take the turn in a conversation (interactive function) or as indicators of reformulation (metatextual function) -see (Bazzanella, 1995) and (Ferrari, 2010). Once the connective is identified, we also need to disambiguate it, in order to associate it with the sense of the relation it is actually conveying. A manual annotation is thus needed for the creation of a resource of reference for both linguistic and computational uses.

5. Results

The results of our work will be presented in three sections considering that: the format used for registering the information in LICO has been updated since we introduce new elements; a new list of connectives has been created since we validated and modified the first version of LICO; a revision of the polysemy of the connectives has been carried out with the new data.

The resulting format. Figure 1 shows the connective *a dire la verità* and its variants in LICO. In the central part of the figure, we can see how examples from the three different sources (i.e. CIB, Adige, Wikipedia) have been reported

in LICO. Specifically, the examples have their own id and are identified with the tag “source” which attribute is a two or three position code standing for: *i*) the source corpus (i.e. CIB or Adige or Wiki), *ii*) the source document id as identified in the source corpus, *iii*) just for example from CIB, the id of the tokens of the connective in the source document. For instance, in Figure 1 the first example is from document 405635 of CIB, and in that document the connective corresponds to tokens 185 and 186. Notice also that the source documents are distributed with LICO, as external material at users disposal.

A new list of connectives of contrast. One important benefit of considering CIB, a corpus that has been exhaustively annotated with contrast relations, concerns the enrichment of the list of 38 connectives of contrast in LICO. As can be seen in Table 2, five of the connectives tagged with CONTRAST or CONCESSION, or both relations, in CIB are not present in LICO. More precisely *al contrario di* and *seppure* were not present in LICO as connectives, while *e*, *in realtà* and *se* were in LICO but associated to a non contrastive sense.

On the other side, the corpus investigation bring us to eliminated 3 connectives from this list: *con tutto questo*, *a onor del vero*, *persino*. The three of them were found in the corpus as connectives but not conveying a contrast relation; they have been kept in LICO as connectives of the PDTB sense EXPANSION:Conjunction.⁹

⁹A deep examination highlights that the first two were retrieved from the resource Sabatini Coletti, which does not specify the relation they convey. The assignment to the contrast sense is probably derived by the fact that their synonyms (i.e. *nonostante tutto* for *con tutto questo* and *a dire la verità* for *a onor del*

⁸Documents are from Italian Wikipedia, February 2010.

connective	CIB	L	connective	CIB	L
a dire la verità (in verità)	x		eppure	x	x
a dire il vero		x	in realtà	x	
a dispetto di	x	x	invece	x	x
a onor del vero		x	ma	x	x
ad ogni modo		x	malgrado		x
al contrario	x	x	malgrado ciò		x
al contrario di	x		mentre	x	x
anche		x	nondimeno		x
anche se	x	x	nonostante	x	x
benché		x	per contro		x
bensì		x	per quanto		x
ciononostante		x	però	x	x
cionondimeno		x	persino		x
comunque	x	x	pur / pure	x	x
con tutto questo		x	quantunque		x
contrariamente a		x	se	x	
da un canto.. dall'altro		x	sebbene		x
da un lato.. dall'altro lato		x	sennonché		x
da una parte.. dall'altra parte	x	x	seppure	x	
e	x		solamente		x
			solo		x
			tuttavia	x	x
			viceversa		x
			Total	19	38

Table 2: Connectives of contrast in Contrast-Ita-Bank (CIB) and in LICO (L). Three were removed as they do not appear to convey the contrast relation.

We thus update the list of connectives (of contrast) in LICO including the connectives from Contrast-Ita Bank and discarding those found not conveying this relation in the corpus; the final list is 40 connectives (38 in LICO - 3 removed + 5 from CIB). This result, even limited, highlights the importance of a corpus investigation in order to enrich a lexical resource.

Checking polysemy The use of corpora not only permits to discover new connectives of contrast or contrast uses of connectives that was already in LICO, but it leads us to review the polysemy of the already listed connectives of contrast. For example, we add the sense EXPANSION:Exception:Arg2-as-except to the connectives *solamente* and *solo* (both Eng. 'only') since we find example as the following, in which it introduces an exception.

- (2) L'attaccante quindi genera un certificato server falso, totalmente uguale al certificato vero, solamente che non è firmato dalla stessa CA.¹⁰

Table 3 shows the data of the enrichment. Notice that glob-

vero) are in fact conveying the contrast relation; in context however, they seem to convey more the EXPANSION:Conjunction relation. For what concerns *persino*, it has been retrieved as a translation of the German *auch* which however, is not associated with the contrast relation in the DimLex lexicon; we do not find example for its use as contrastive.

¹⁰Eng. "The attacker then generates a fake server certificate, totally equal to the true certificate, only that it is not signed by the same CA."

ally the average number of examples for each entry almost doubled, even though we added only two entries to the resource. As already specified, the corpus analysis also led us to discover new senses for the connectives under examination (for a total of 214 senses over 175 entries with respect to the previous 205 over 173 connectives).

Data	Pre	Post
# Connectives of Contrast in LICO	38	40
Average polysemy of the connectives in LICO	1.18	1.22
# examples per connective sense (average)	1.73	3.37

Table 3: Data on connectives of contrast in Lico pre and post corpus enrichment.

6. Conclusion and Further works

We have presented our project aiming at enriching the pre-existing resource LICO (Feltracco et al., 2016), a lexicon of discourse connectives for Italian, with real corpus data for connectives marking contrast.

The adopted methodology partially takes advantage of a pre-existing resource in which discourse connectives of contrast are annotated, along with the arguments of the discourse relation they make explicit. A complementary investigation has been conducted picking examples of the connectives in corpora and manually disambiguating their senses in the retrieved textual contexts. In particular, this latter strategy is replicable to enrich LICO with information about discourse connectives that convey relations other than contrast (e.g. temporal, causal) and it can also be adopted to enriched lexica of connectives in other language.

Corpus investigation can also be carried out in a more automatic way: for example, Bourgonje et al. (2017) use parallel corpus to discover correspondences between connectives in different language and highlight point to gaps in the examined resources. In particular, the authors report on experiments to validate the list of connectives in DimLex (Stede, 2002) and LICO in an effort to constructing a bilingual lexicon on connectives that are connected via their discourse senses.

In our case, since we want to extract clear examples and disambiguate their senses in the context, we believe the manual disambiguation of connectives was necessary. The resulting resource represents a valuable tool for both linguistic analyses and the training of a classifier for NLP applications.

7. Bibliographical References

- Asher, N. and Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.
- Bazzanella, C. (1995). I segnali discorsivi. *Grande grammatica italiana di consultazione*, 3:225–257.
- Bourgonje, P., Grishina, Y., and Stede, M. (2017). Toward a bilingual lexical database on connectives: Exploiting a german/italian parallel corpus. In *Proceedings of the Fourth Italian Conference on Computational Linguistics (CLiC-it 2017)*.
- Feltracco, A., Jezek, E., Magnini, B., and Stede, M. (2016). LICO: A Lexicon of Italian Connectives. In *Proceedings*

- of the Third Italian Conference on Computational Linguistic (CLiC-it 2016), page 141.
- Feltracco, A., Magnini, B., and Jezek, E. (2017). Contrast-ita bank: A corpus for italian annotated with discourse contrast relations. In *Proceedings of the Fourth Italian Conference on Computational Linguistic (CLiC-it 2017)*.
- Ferrari, A. and Zampese, L. (2000). *Dalla frase al testo : una grammatica per l'italiano*. Zanichelli.
- Ferrari, A. (2010). Connettivi. In *Enciclopedia dell Italiano*. diretta da Raffaele Simone, con la collaborazione di Gaetano Berruto e Paolo D Achille, Roma, Istituto della Enciclopedia Italiana.
- Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L., and Webber, B. L. (2007). The Penn Discourse Treebank 2.0 Annotation Manual.
- Roze, C., Danlos, L., and Muller, P. (2012). LEXCONN: a French lexicon of discourse connectives. *Discours. Revue de linguistique, psycholinguistique et informatique.*, (10).
- Sabatini, F. and Coletti, V. (2007). Dizionario della lingua italiana 2008 (2007). *Milano: Rizzoli Larousse*.
- Scheffler, T. and Stede, M. (2016). Adding Semantic Relations to a Large-Coverage Connective Lexicon of German. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC '16)*, Portoroz, Slovenia, May.
- Stede, M. and Heintze, S. (2004). Machineassisted rhetorical structure annotation. In *Proceedings of the 20th International Conference on Computational Linguistics*, pages 425–431, Geneva.
- Stede, M. and Umbach, C. (1998). Dimlex: A lexicon of discourse markers for text generation and understanding. In *Proceedings of the 17th international conference on Computational linguistics Volume 2*, pages 1238–1242. Association for Computational Linguistics.
- Stede, M. (2002). Dimlex: A lexical approach to discourse markers. In *Exploring the Lexicon - Theory and Computation*. Edizioni dell'Orso, Alessandria.
- Stede, M. (2012). *Discourse processing*. Morgan & Claypool Publishers.
- Webber, B., Prasad, R., Lee, A., and Joshi, A. (2016). A Discourse Annotated Corpus of Conjoined VPs. *LAW X*, page 22.