

DiMLex-Bangla: A Lexicon of Bangla Discourse Connectives

Debopam Das, Manfred Stede, Soumya Sankar Ghosh, Lahari Chatterjee

Humboldt University of Berlin, University of Potsdam, Mihup Communications, Jadavpur University
dasdebop@bu-berlin.de, stede@uni-potsdam.de, soumya@mihup.com, lahari.chatterjee@gmail.com

Abstract

We present DiMLex-Bangla, a newly developed lexicon of discourse connectives in Bangla. The lexicon, upon completion of its first version, contains 123 Bangla connective entries, which are primarily compiled from the linguistic literature and translation of English discourse connectives. The lexicon compilation is later augmented by adding more connectives from a currently developed corpus, called the Bangla RST Discourse Treebank (Das and Stede, 2018). DiMLex-Bangla provides information on syntactic categories of Bangla connectives, their discourse semantics and non-connective uses (if any). It uses the format of the German connective lexicon DiMLex (Stede and Umbach, 1998), which provides a cross-linguistically applicable XML schema. The resource is the first of its kind in Bangla, and is freely available for use in studies on discourse structure and computational applications.

Keywords: discourse connectives, lexicon, discourse relation, Bangla

1. Introduction

Successful interpretation of texts (at the cognitive level) and the establishment of valid text structures (at the meta-discourse level) largely depend on the identification of the presence and type of discourse relations that hold between text segments. Discourse relations are often indicated by discourse connectives (*and, but, however, thus, etc.*), which are widely believed to be the most explicit, most prototypical and most reliable relational signals (Danlos et al., 2018). Accordingly, the role of connectives in discourse interpretation has been extensively studied in both (psycholinguistic) discourse processing (see Kleijn et al. (2019) for an overview) and computational discourse applications such as discourse parsing (Hernault et al., 2010; Lin et al., 2014), machine translation (Meyer et al., 2011), text summarization (Alemany, 2005), or argumentation mining (Kirschner et al., 2015). This has led to the production of an increasingly large number of text corpora in different languages, annotated for discourse connectives (most often, in addition to other signals), based on different discourse frameworks (for references, see Das and Taboada (2019)).

Parallel to the practice of developing discourse-annotated corpora, initiatives have also been taken to construct ancillary resources such as lexicons of discourse connectives, which can provide the list of connectives in a language, along with useful information about their syntactic and semantic-pragmatic properties. Such lexicons are developed and are becoming increasingly available in different languages, beginning with DiMLex in German (Stede and Umbach, 1998), later DPDE for Spanish (Briz et al., 2008) and LexConn for French (Roze et al., 2012), and more recently LICO for Italian (Feltracco et al., 2016), CzeDLex for Czech (Mírovský et al., 2017), LDM-PT for Portuguese (Mendes et al., 2018), DisCoDict for Dutch (Bourgonje et al., 2018) and DiMLex-Eng for English (Das et al., 2018).

We extend this line of work by developing a new lexicon of discourse connectives in Bangla, called DiMLex-Bangla. The lexicon, upon completion of its first version, contains

123 Bangla discourse connectives, which are primarily accumulated from the linguistic literature on Bangla and translation of English discourse connectives into Bangla. The lexicon compilation is further augmented by including more connectives from a currently developed corpus, called the Bangla RST Discourse Treebank (Das and Stede, 2018). Each connective entry in DiMLex-Bangla is accompanied by information on the possible orthographic/stylistic variants of the connective, its syntactic category, non-connective usage (if any), and the set of discourse relations indicated by the connective (with examples). For organizing the lexicon, we follow the German connective lexicon, DiMLex¹ (Stede and Umbach, 1998), whose most updated version (Scheffler and Stede, 2016) includes an exhaustive list of 275 German discourse connectives. For each connective entry, DiMLex provides a number of features, characterizing its syntactic, semantic and pragmatic behaviour.

Bangla (also known by its English exonym *Bengali*) is an Indo-Aryan language which is natively spoken in the Indian subcontinent (India and Bangladesh). Bangla has long been a relatively well-studied language (although more from the viewpoint of traditional grammars than from the modern linguistic perspectives). More recently, Bangla has also become subject to NLP-based research, mainly pertaining to areas like morpho-syntax, speech or typography. Unfortunately, there exist very little resources for research on discourse structure. We aim to reduce this shortcoming by developing DiMLex-Bangla, the first (to our knowledge) lexicon of Bangla discourse connectives, which is freely available for use. In this paper, we present our lexicon, and describe the design features and development steps.

2. Defining Bangla Connectives

Discourse connectives, strictly speaking, refer to lexical items (conjunctions, adverbials and prepositional phrases)

¹DiMLex Marker LEXicon: <https://github.com/discourse-lab/dimlex>

which encode a two-place meaning relationship between propositions functioning as their arguments. This excludes lexical expressions like discourse markers (Schiffrin, 1987), which we regard as pragmatic markers with interactional/modal import (Cuenca and Marín, 2009; Cuenca and Crible, 2019), Alternative Lexicalizations (AltLex) in the Penn Discourse Treebank (Prasad et al., 2007), secondary connectives in the Prague Discourse Treebank (Rysová and Rysová, 2015), or lexical signals, such as indicative words/phrases, as proposed in (Das et al., 2015).

Our selection of connective entries follows from what we consider as Bangla discourse connectives. The connective definition in Bangla is partially based on the definition used in DiMLex (Stede and Umbach, 1998), following Pasch et al. (2003). The definition is, however, modified by incorporating certain language-specific features of Bangla. We consider a word or phrase x as a connective in Bangla if it fulfills the following criteria:

1. x can be a single-word or multiword expression, but it cannot be inflected.
2. The meaning of x is a two-place relation.
3. The arguments of this relation are abstract objects that represent propositions, events, states, or processes.
4. Syntactically, the arguments are typically expressed as clausal/sentential structures, and sometimes by other structures such as participial/infinitive constructions and noun phrases (denoting abstract objects).
5. x can either be a fixed expression (completely grammaticalized) or modifiable expression (not yet fully grammaticalized). We refer to the former type as *primary* connectives and the latter as *modifiable* connectives.
 - (a) x , being a *primary* connective, cannot be semantically compositional with respect to its component parts. This criterion excludes phrases that consist of a connective and an intensifier/adverb such as /shudhu yadi/² ‘only if’ or /kebal jakhan/ ‘only when’ (here, /yadi/ ‘if’ and /jakhan/ ‘when’ are considered to be (primary) connectives).
 - (b) x , being a *modifiable* connective, are subject to modification either through compounding (mostly with adjectives) or by the attachment of an anaphoric element, as shown in Table 1.

3. Sources of Bangla Connectives

The first step of creating a lexicon is to select and classify the entries to be included therein. In our case, since there is no prior list of Bangla discourse connectives available, we start from scratch, using multiple sources to compile a list of Bangla connectives, as described below.

3.1. Bangla Grammars

One can manually build a repository of connectives from existing resources: Stede and Umbach (1998) develop the initial version of DiMLex, compiling connective entries from the standard German dictionaries and grammars;

²The characters in the Bangla examples are written in the Roman script (ITRANS style).

Alonso et al. (2002) gather a large set of Spanish connectives from previous work in addition to a corpus study; Roze et al. (2012) construct LexConn, a lexicon of French connectives manually extracted from a corpus called FranText. In our case, we note that connectives, according to the traditional Bangla grammars, fall under the class of ‘indeclinables’, a category used to refer to words that are not subject to inflection. We make a list of such 45 connectives from the available literature on the standard Bangla grammars (David, 2015; Thompson, 2010; Chaki, 1996; Chatterji, 1988)³.

3.2. Translation of English Connectives

In order to collect more connectives, we adopt a translational strategy, generating a list of translational equivalents in Bangla for English connectives. For this purpose, we use as our source a lexicon of English connectives, called DiMLex-Eng (Das et al., 2018), which comprises 149 English connectives compiled from resources such as the Penn Discourse Treebank (Prasad et al., 2008) and RST Signalling Corpus (Das et al., 2015). To translate these connectives into Bangla, we consult the available standard bilingual dictionaries (Biswas, 1957; Ghosh, 1999), mainly for items that comprise only a single word (*furthermore*, *since*, etc.). For translating the multiword connectives of English (*as a result of*, *quite the contrary*, etc.), we use a well-recognized online translation tool Google Translate. Using the translational method, we include 58 additional connectives in DiMLex-Bangla.

3.3. The Bangla RST Discourse Treebank

We also use a corpus of Bangla texts, called the Bangla RST Discourse Treebank or Bangla RST-DT, to check the presence of any new connectives not identified using the first two methods above. The corpus provides annotations of discourse relations following Rhetorical Structure Theory or RST (Mann and Thompson, 1988), and is currently under production. It contains a collection of 266 Bangla texts (comprising over 70,000 words) from a popular Bangla daily called *Anandabazar Patrika*⁴ published in India. For more information about the corpus, annotation guidelines and annotation procedure, see (Das and Stede, 2018) and Das (2018). A careful perusal in the corpus helps us identify 20 more new connectives.

In this way, we include in DiMLex-Bangla a total of 123 (45 + 58 + 20) connectives, which are further divided into 100 primary connectives and 23 modifiable connectives. As it is evident from the compilation procedures we follow, the list of discourse connectives is probably not exhaustive. However, the collection, we believe, is representative and reliable, with possible options for further incorporating more connectives into the lexicon, thus making it more comprehensive and expanded in the future.

³The Bangla linguistic literature typically provides only a few illustrative examples, rather than an exhaustive list of Bangla connectives.

⁴<https://www.anandabazar.com/>

A comparison across connectives lexicons (mentioned in Section 1.) shows that the size of a connective lexicon inventory may vary considerably from language to language: 275 for German, 210 for Spanish, 328 for French, 173 for Italian, 208 for Czech, 252 for Portuguese, 207 for Dutch and 149 for English. The inventory size appears to largely depend on the definition of discourse connectives adopted for a lexicon⁵, as a strict set of criteria may lead to the inclusion of a lower number of connectives than what would normally be resulted from a more open set of features. Another possible reason for the variation in number concerns the way a language assigns resources for relation marking. It is possible that alternative ways of relation marking may reduce or exclude the use of a discourse connective altogether, and this may determine the number of the connectives a language makes available⁶.

4. Populating the Lexicon Entries

For each entry in DiMLex-Bangla, we provide significant lexicographic information on both formal and functional properties of the corresponding Bangla connective, as listed below:

1. Type of connective (primary or modifiable)
2. Part of speech (POS) of the connective (for a modifiable connective, the POS of the root word)
3. English translation of the connective
4. Possible variants of the connective
 - Orthographic: other spelling(s)
 - Stylistic: other form(s) based on stylistic or social contexts⁷
5. Ambiguity information (whether the lexical item also has a non-connective reading)
 - If ambiguous, examples of non-connective usage
6. Discourse relations expressed by the connective

⁵For a cross-linguistic comparison of defining parameters for discourse connectives, see (Stede et al., 2019).

⁶With respect to Bangla, we observe, for example, that certain monosyllabic discourse particles (functioning as enclitics) systematically and frequently signal additive, concessive and conditional relations. In the case of English, purpose relations are mostly indicated by infinitival clauses ('in order to...') rather than by a discourse connective (Das, 2014). It may well be the case that these relations are mostly (if not exclusively) signalled by lexical signals like connectives in a different language, and this may also be a reason for the language to have a larger number and variety of discourse connectives.

⁷In formal writing, Bangla used to be (and is still infrequently) represented by an earlier variety, called *Sadhubhasha*, which is marked by the use of a higher degree of /tatsama/ words ('same words from Sanskrit') and relatively more rigid syntactic structures. The other variety, called *Chalitbhasha* ('colloquial language'), was previously used only for speech, but has later been adopted for writing Bangla texts, largely substituting the use of *Sadhubhasha* in a wide range of contexts in the present era. In DiMLex-Bangla, we primarily include Bangla connectives from *Chalitbhasha* (as they are used in the standard Bangla language). In addition, we also include the earlier forms of those connectives from *Sadhubhasha* (if available), and record them as their stylistic variants. For example, we include /yAte/ from *Chalitbhasha* and /yAhAte/ from *Sadhubhasha*, both meaning 'so that', and record the latter as a stylistic variant of the former.

- According to the Penn Discourse Treebank (PDTB) 3.0 sense hierarchy (Webber et al., 2016)
- According to the RST (Rhetorical Structure Theory) taxonomy used in the Potsdam Commentary Corpus (PCC) (Stede, 2016)
- Examples of coherence relations, mostly from the Bangla RST-DT (Das and Stede, 2018)

We provide relational information for each connective based on two discourse frameworks: PDTB 3.0 and RST (as in PCC). In the PDTB 3.0, discourse relations (known as *senses*) are organized in a hierarchical taxonomy, organized into three levels: sense classes, sense types and sense subtypes (as shown in Table 2). On the other hand, the PCC relational taxonomy (also used in annotating the Bangla RST-DT) is based on the original RST taxonomy (Mann and Thompson, 1988), and it includes a set of 31 relations which are divided with respect to two parameters: (1) whether a relation is mononuclear or multinuclear; and (2) whether a relation is semantic, pragmatic or textual in nature (as shown in Table 3).

For encoding the lexicographic features in DiMLex-Bangla, we use the format of the German lexicon DiMLex, which provides a cross-linguistically applicable XML schema. As an illustration, we provide the lexicographic information for the connective entry /bale/ 'because' in Figure 1. The entry shows that /bale/ is a subordinating conjunction functioning as a primary connective. The connective is used to indicate two possible discourse relations: (1) Contingency > Purpose (according to PDTB) or Purpose (according to RST) and (2) Contingency > Cause-Belief > Reason (according to PDTB) or Reason (according to RST). The connective /bale/ has a non-connective reading. It also has a stylistic variant, /baliyA/.

5. Summary and Outlook

We have presented DiMLex-Bangla, a lexicon of discourse connectives first ever compiled for Bangla. The lexicon, upon the release of its first version, includes a total of 123 connectives entries, accumulated from the linguistic literature, translational outputs and evidence from corpus. Connectives in DiMLex-Bangla, based on their morphological attributes, are classified into two types: primary and modifiable connectives. The lexicon is modelled after the German connective lexicon, DiMLex, which provides a cross-linguistically applicable XML schema. Each entry in DiMLex-Bangla is accompanied with significant information on the formal characteristics (grammatical category, orthographical variant, possible modifiers, etc.) and semantic-discourse properties (discourse relations, non-connective usage, stylistic variant, etc.) of the corresponding connective.

As mentioned in Section 3., the majority of the DiMLex-Bangla entries come from multiple sources other than corpora. Only a subset of DiMLex-Bangla entries is compiled from the Bangla RST-DT, which, unlike the PDTB-style corpora, is not annotated for connectives yet. Our future work includes (in addition to the expansion of the lexicon)

Modifiable connective	Modifier	Modifier type	Modified connective
/dike/ ‘side.locative’	/anya/ ‘other’	adj	/anyadike/ ‘on the other side’
	/ulTo/ ‘opposite’	adj	/ulTodike/ ‘in contrast’
/janya/ ‘for’	/se/ ‘that’	anaphoric	/sejanya/ ‘for that (reason)’
	/e/ ‘this’	anaphoric	/ejanya/ ‘for this (reason)’
/bhAbe/ ‘in the manner’	/anurUp/ ‘similar’	adj	/anurUpbhAbe/ ‘similarly (in the similar manner)’
	/AlAdA/ ‘different’	adj	/anurUpbhAbe/ ‘differently (in a different manner)’
	/sei/ ‘that’	anaphoric	/seibhAbe/ ‘in that manner’

Table 1: Examples of modifiable connectives in Bangla

```

<entry id="c37" word="বলে" type="primary">
<modifiers/>
<english_equivalent>because</english_equivalent>
<orths>
<orth type="cont" canonical="1" onr="38o1">
<part type="single">বলে</part>
</orth>
<orth type="cont" canonical="0" onr="38o2">
<part type="single"></part>
</orth>
</orths>
<stylistic_variations>
<stylistic_variatiant>বলিয়া</stylistic_variatiant>
</stylistic_variations>
<ambiguity>
<non_conn>1</non_conn>
<sem_ambiguity>1</sem_ambiguity>
</ambiguity>
<non_conn_reading>
<example>আজকের রায়ে এই সব অভিযোগই মিথ্যা বলে প্রমাণিত হয়েছে।</example>
<english_example_equivalent>All allegations have been proven to be lies by today's verdict.</english_example_equivalent>
<part_of_speech>postposition</part_of_speech>
</non_conn_reading>
<syn>
<cat>subordinating-conjunction</cat>
<sem>
<pttb3_relation sense="Contingency:Purpose"/>
<rst_relation sense="Purpose"/>
<example>আমি তার সাথে দেখা করবো বলে তার অফিসে গিয়েছিলাম।</example>
<english_equivalent>I went to his office in order to meet him.</english_equivalent>
</sem>
<sem>
<pttb3_relation sense="Contingency:Cause-belief:Reason"/>
<rst_relation sense="Reason"/>
<example>অভীক পাপ করেছে বলে তার সাজা হয়েছে।</example>
<english_equivalent>Abhik has been punished because he committed sins.</english_equivalent>
</sem>
</syn>
</entry>

```

Figure 1: DiMLex-Bangla entry for the connective /bale/ ‘because’

validation of the non-corpus lexicon entries, both in terms of their presence in natural language texts and the accuracy of their lexicographic information as provided in the lexicon. In particular, we would like to examine:

1. Whether and how frequently the (non-corpus) connectives included in the lexicon actually appear in corpora. Furthermore, for ambiguous connectives, we examine how many of them occur in corpora with respect to their connective vs. non-connective meanings.
2. Whether the lexicographic features assigned to the (non-corpus) connectives in the lexicon actually correspond with their occurrence in corpora (more importantly, whether the connectives in corpora are used to convey the same relations as assigned to them in the lexicon).

DiMLex-Bangla is freely available⁸, and has been added as part of the Connective-Lex database⁹ at the University

⁸<https://github.com/discourse-lab/>

DiMLex-Bangla

⁹<http://connective-lex.info/>

of Potsdam. The lexicon can be used in areas such as language learning and contrastive linguistics. Also, the connectives in DiMLex-Bangla, together with other relational signals, can be used as a valuable resource for discourse parsing and related applications. Furthermore, the lexicon can also serve as a model for constructing similar connective lexicons in other Indic languages like Hindi, Punjabi or Marathi.

6. Bibliographical References

- Aleman, L. A. (2005). *Representing discourse for automatic text summarization via shallow NLP techniques*. PhD dissertation, Universitat de Barcelona, Spain.
- Alonso, L., Castellón, I., Gibert, K., and Padró, L., (2002). *An Empirical Approach to Discourse Markers by Clustering*, volume 2504 of *Lecture Notes in Computer Science*, pages 173–183. Springer, Berlin.
- Biswas, S. (1957). *Songsod Bangla Obhidhan*. Sahitya Sangsad, Kolkata, India.
- Chaki, J. (1996). *Bangla Bhashar Byakaran*. Ananda Publishers, Kolkata, India.

Level-1	Level-2	Level-3
Temporal	Synchronous	-
	Asynchronous	Precedence Succession Reason Result NegResult Reason+Belief Result+Belief
Contingency	Cause	Reason+SpeechAct Result+SpeechAct
	Cause+Belief	Arg1-as-Cond Arg2-as-Cond
	Cause+SpeechAct	-
	Condition	Arg1-as-NegCond Arg2-as-NegCond
	Condition+SpeechAct	-
	Negative-Condition	Arg1-as-Goal Arg2-as-Goal
	Negative-Condition+SpeechAct	Arg1-as-Denier Arg2-as-Denier
	Purpose	Arg2-as-Denier+SpeechAct
		-
Comparison	Concession	-
	Concession+SpeechAct	-
	Contrast	-
	Similarity	-
Expansion	Conjunction	-
	Disjunction	-
	Equivalence	-
	Exception	Arg1-as-Excpt Arg2-as-Excpt
	Instantiation	Arg1-as-Instance Arg2-as-Instance
	Level-of-Detail	Arg1-as-Detail Arg2-as-Detail
	Manner	Arg1-as-Manner Arg2-as-Manner
	Substitution	Arg1-as-Subst Arg2-as-Subst
		-
		-
		-

Table 2: PDTB 3.0 sense hierarchy

Relation type	Semantic	Pragmatic	Textual
Mononuclear	Circumstance	Background	Preparation
	Condition	Antithesis	Restatement
	Otherwise	Concession	Summary
	Unless	Evidence	
	Elaboration	Reason	
	E-Elaboration	Reason-N	
	Interpretation	Justify	
	Means	Evaluation-N	
	Cause	Evaluation-S	
	Result	Motivation	
	Purpose	Enablement	
	Solutionhood		
	Multinuclear	Sequence	
		Contrast	
		Conjunction	
		List	
		Joint	

Table 3: RST relation taxonomy in PCC

Chatterji, S. K. (1988). *Bhasha-Prakash Bangla Vyakaran*. Rupa and Company, New Delhi, India.

Cuenca, M. J. and Crible, L. (2019). Co-occurrence of discourse markers in English: From juxtaposition to composition. *Journal of Pragmatics*, 140:171–184.

Cuenca, M. J. and Marín, M. J. (2009). Co-occurrence of discourse markers in Catalan and Spanish oral narrative. *Journal of Pragmatics*, 41(5):899–914.

Danlos, L., Rysová, K., Rysová, M., and Stede, M. (2018). Primary and secondary discourse connectives: definitions and lexicons. *Dialogue and Discourse*, 9(1):50–78.

Das, D. and Stede, M. (2018). Developing the Bangla RST Discourse Treebank. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 1832–1838, Miyazaki, Japan. European Language Resources Association (ELRA).

Das, D. and Taboada, M. (2019). Multiple Signals of Coherence Relations. *Discours*, 24 [online].

Das, D. (2014). *Signalling of Coherence Relations in Discourse*. PhD dissertation, Simon Fraser University, Canada.

Das, D. (2018). Discourse Segmentation in Bangla. In Girish Nath Jha, et al., editors, *Proceedings of the 4th Workshop on Indian Language Data Resource Evaluation (WILDRE-4)*, Miyazaki, Japan. European Language Resources Association (ELRA).

David, A. B. (2015). *Descriptive Grammar of Bangla*. Mouton-CASL Grammar Series 2. DE GRUYTER MOUTON, Germany.

Ghosh, G. P. (1999). *Everyman’s Dictionary [English-Bengali]*. Ramkrishna Pustakalaya, Kolkata, India.

Hernault, H., Prendinger, H., duVerle, D., and Ishizuka, M. (2010). HILDA: A Discourse Parser Using Support Vector Machine Classification. *Dialogue and Discourse*, 1(3):1–33.

Kirschner, C., Eckle-Kohler, J., and Gurevych, I. (2015). Linking the thoughts: Analysis of argumentation structures in scientific publications. In *Proceedings of the 2015 NAACL-HLT Conference*, pages 1–11. Association for Computational Linguistics.

Kleijn, S., Maat, H. L. P., and Sanders, T. J. (2019). Comprehension Effects of Connectives Across Texts, Readers, and Coherence Relations. *Discourse Processes*, 56(5-6):447–464.

Lin, Z., Ng, H. T., and Kan, M.-Y. (2014). A PDTB-Styled End-to-End Discourse Parser. *Natural Language Engineering*, 20:151–184.

Mann, W. and Thompson, S. (1988). Rhetorical structure theory: Towards a functional theory of text organization. *TEXT*, 8:243–281.

Meyer, T., Popescu-Belis, A., Zufferey, S., and Cartoni, B. (2011). Multilingual Annotation and Disambiguation of Discourse Connectives for Machine Translation. In *Proceedings of the SIGDIAL 2011 Conference*, SIGDIAL ’11, pages 194–203, Stroudsburg, PA, USA. Association for Computational Linguistics.

Pasch, R., Brauße, U., Breindl, E., and Waßner, U. H. (2003). *Handbuch der deutschen Konnektoren*. Walter de Gruyter, Berlin/New York.

Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L., and Webber, B. (2007). The Penn Discourse Treebank 2.0 Annotation Manual. Technical report, Institute for Research in Cognitive Science, Philadelphia.

Roze, C., Danlos, L., and Muller, P. (2012). LexConn: A French Lexicon of Discourse Connectives. *Discours*, 10.

Rysová, M. and Rysová, K. (2015). Secondary Connectives in the Prague Dependency Treebank. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, pages 291–299. Uppsala University, Uppsala, Sweden.

Schiffirin, D. (1987). *Discourse Markers*. Studies in Interactional Sociolinguistics. Cambridge University Press, Cambridge.

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 and 36th Annual Meeting of the Association for Computational Linguistics (COLING/ACL'98)*, pages 1238–1242, Montréal, Canada.
- Thompson, H.-R. (2010). *Bengali: A comprehensive grammar*. Routledge Comprehensive Grammars. Routledge, London, New York.
- ### 7. Language Resource References
- Bourgonje, P., Hoek, J., Evers-Vermeul, J., Redeker, G., Sanders, T., and Stede, M. (2018). Constructing a Lexicon of Dutch Discourse Connectives. *Computational Linguistics in the Netherlands Journal*, 8:163–175.
- Briz, A., Bordería, S. P., and Portolés, J. (2008). Diccionario de partículas discursivas del español.
- Das, D. and Stede, M. (2018). Developing the Bangla RST Discourse Treebank. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 1832–1838, Miyazaki, Japan. European Language Resources Association (ELRA).
- Das, D., Taboada, M., and McFetridge, P. (2015). RST Signalling Corpus, LDC2015T10.
- Das, D., Scheffler, T., Bourgonje, P., and Stede, M. (2018). Constructing a Lexicon of English Discourse Connectives. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 360–365, Melbourne, Australia. Association for Computational Linguistics.
- Feltracco, A., Jezek, E., Magnini, B., and Stede, M. (2016). Lico: A Lexicon of Italian Connectives. In *Proceedings of the 3rd Italian Conference on Computational Linguistics (CLiC-IT 2016)*, Napoli, Italy.
- Mendes, A., del Rio, I., Stede, M., and Dombek, F. (2018). A Lexicon of Discourse Markers for Portuguese: LDM-PT. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC)*, pages 4379–4384, Miyazaki, Japan.
- Mírovský, J., Synková, P., Rysová, M., and Políková, L. (2017). CzeDLex - A Lexicon of Czech Discourse Connectives. 109(1):61–91.
- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., and Webber, B. (2008). The Penn Discourse Treebank 2.0. In *Proc. of the 6th International Conference on Language Resources and Evaluation (LREC)*, pages 2961–2968, Marrakech, Morocco.
- Roze, C., Danlos, L., and Muller, P. (2012). LexConn: A French Lexicon of Discourse Connectives. *Discours*, 10.
- 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)*, pages 1008–1013, Portorož, Slovenia.
- 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 and 36th Annual Meeting of the Association for Computational Linguistics (COLING/ACL'98)*, pages 1238–1242, Montréal, Canada.
- Stede, M., Scheffler, T., and Mendes, A. (2019). Connective-Lex: A Web-Based Multilingual Lexical Resource for Connectives. *Discours*, 24 [online].
- Stede, M. (2016). *Rhetorische Struktur*. Universitätsverlag, Potsdam.
- Webber, B., Prasad, R., Lee, A., and Joshi, A. (2016). A discourse-annotated corpus of conjoined VPs. In *Proceedings of the 10th Linguistic Annotation Workshop, ACL*, pages 22–31, Berlin, Germany.