Automatic identification of rhetorical relations among intra-sentence discourse segments in Arabic

Article *in* International Journal of Intelligent Systems Technologies and Applications · January 2019

DOI: 10.1504/IJISTA.2019.10020096

4 authors:


Monther Aldwairi
Zayed University


Redjimi Mohamed
University of Skikda


Samira Lagrini
Badji Moktar-Annaba University


Nabiha Azizi
Badji Moktar-Annaba University

Automatic Identification of Rhetorical Relations Among Intra-sentence Discourse Segments in Arabic

ABSTRACT

Identifying discourse relations, whether implicit or explicit, has seen renewed interest and remains an open challenge.We present the first model that automatically identify both explicit and implicit rhetorical relations among intra-sentence discourse segments in Arabic text. We build a large discourse annotated corpora following the rhetorical structure theory framework. Our list of rhetorical relations is organized into three level hierarchy of 23 fine-grained relations, grouped into seven coarse-grained relations classes. To automatically learn these relations, we evaluate and reuse features from literature, and contribute three additional features: accusative of purpose, specific connectives and the number of antonym words. We perform experiments on fine and coarse-grained relations identification. The results show that compared with all the baselines, our model achieves the best performance in most cases, with an F-score of 86.74% and an accuracy of 91.05%.

Keywords: Discourse relations, Rhetorical structure theory, automatic Arabic language processing, machine learning

1. INTRODUCTION

Automatic identifying of discourse relations (e.g., contrast, cause…) between two text segments is a crucial task for discourse analysis. Ithas been shown to be useful for many Natural Language Processing (NLP) applications such as information extraction [Cimiano et al., 2005], question answering [Sadek and Meziane, 2016; Jansen et al., 2014],sentiment analysis [Somasundaran et al., 2009], and machine translation [Tu et al., 2013].

Depending on the presence or absence of discourse connectives (such as because,but, despite…etc.), discourse, rhetorical or coherence relations can be classified into explicit and implicit relations. Discourse connectivesplay a pivotal role in identifying explicit discourse relations [Pitler et al.,2008]. For instance, “because” is a strong indicator of a causal relation. However,in the absence of clear indicators, as in “*Be careful! it’s raining*”,the relation is called implicit, and recognizing such relations remains a serious challenge.

Identifying discourse relations has seen renewed interestin the literaturewithin different frameworks: Rhetorical Structure Theory (RST) [Mann and Thompson,1988], and Penn Discourse Treebank model (PDTB)[Prasad et al.,2008]. The release of Discourse annotated corpora (RST-DT) [Carlson et al., 2003], Penn Discourse Treebank (PDTB) [Prasad et al., 2008], and Discourse Graphbank[Wolf and Gibson, 2005] has encouraged researchers to explore this area of work[Sporleder and Lascarides, 2007; Lin et al.,2009; Pitler et al., 2009; Louis et al.,2010; Pardo and Cardie, 2012; Biran and McKeown, 2013; Mihaylov and Frank,2016; Li et al., 2017].Severalapproacheshave been proposed to date. Conventional approachestreat this task as a classification problem, that directly rely on informed features extracted from the corpus and designing machine learning algorithms such as Naïve Bayes, Maximum Entropy model (ME), and Support Vector Machine (SVM) [Pitler et al., 2009; Wellner et al., 2009;Lin et al., 2009; Louis et al.,2010].A major challenge for these approaches is: which features are more representatives for texts segments’ pairs.

Various linguistic features were explored, starting from lexical, syntactic, tosemantic features. Although these approaches have proven successful, these featuresare labor-intensive task [Li and Hovy, 2014], and strongly rely on linguistic resources [Li et al., 2017].That is why, recentlydeep learning modelssuch as Convolutional Neural Network (CNN), Recursive Neural Network (RNN),are employed to alleviate the aforementioned problems and to boost the performances of implicit discourse relation recognition [Braud and Denis, 2015; Zhang et al.,2015].

Up until recently, most of the research studies focusing on discourse relations identification had been focused on English language.To the best of our knowledge, there are very few works in this field for Arabic language. The neglect of this important research area may be due to the lack of discourse-annotated corpora upon which researches in discourse analysis heavily rely.The only resource available is the Leeds Arabic Discourse Tree bank (LADTB)[Alsaif and Markert, 2010], which has been recentlypresentedas an Arabic discourse annotated corpus with only explicit relations. However, although implicit discourse relations are very common in Arabic texts, LADTB does not provide any implicit relations annotation. Therefore, we focus in this work on both explicit and implicit discourse relations.

Inthis paper, we tackle both explicit and implicit discourse relation recognitionwithin RSTframework,focusing on the intra-sentential level(i.e. discourse relations between adjacent text segments within the same sentence).To enable us to carry out this research, we develop an RST annotated corpus with high reliability. For the annotation process, we use a list of 23 fine-grained relations enriched with nuclearity annotation, belonging toseven merged relations (classes). To automatically learn intra-sentence Arabic discourse relations, we explore ten groups of features. We contribute three new and novel features: accusative of purpose,specific connectives and number of antonym words.We evaluate and reuse sevenotherfeatures from priorresearch works thathave empirically demonstrated their efficiency. The proposed approach is compared to two baselines: the majority class and discourse markers.Extensive experiments show that it significantly outperforms all the baselines. Up to our knowledge, thisis the first work that tackles both explicit and implicit Arabic discourse relations recognition within the RSTframework. Our contributions are twofold: we propose1) A new intra-sentence Arabic discourse relation hierarchy, and 2) A multi-layer perceptron network model for explicit and implicit Arabic discourse relations recognition.

The rest of the paper is organized as follows: Section 2 gives an overview of RST, whilesection 3 presents the data. Section 4 explains our list of discourse relations, andsection 5provides a detailed description of the annotation process as well as the characteristics of the gold standard. Section 6describes the proposed model and our set of features. Section 7describes theexperiments and analyzes the results. We review relevant related work in section 8 and conclude this paper in section 9

2. RHETORICAL STRUCTURE THEORY

The Rhetorical Structure Theory (RST) [Mann and Thompson,1988] is one of the mostsalient discourse theoriesin computational linguistics**.** It was originally used for text generation, and then it became a famous framework for discourse parsing [Taboada and Mann, 2006]. According to RST,labeled generative trees, called discourse trees (DT), where the leaves correspond to the non-overlapping atomic text spans, called elementary discourse units (EDUs), can represent a coherent text.Adjacent nodes, EDUs or larger discourse units, are linked via discourse relationsto form a discourse sub tree, which in turn can be linked to other neighboring nodes in the tree structure.

Conventionally, discourse analysis in RST involves three subtasks:

1) Segmenting the text into elementary discourse units (EDUs).

2) Identifying discourse relation between consecutive discourse units.

3) Linking all discourse units into a labeled tree.

Fig.1 shows an example ofan RST-tree for the following text fragment consisting oftwo sentences and five EDUs.

[*The impact won't be that great,]1[said Graeme Lidgerwood of First BostonCorp.]2 [This is in part because of the effect]3[of having to average the number of shares outstanding,]4 [she said.]5*

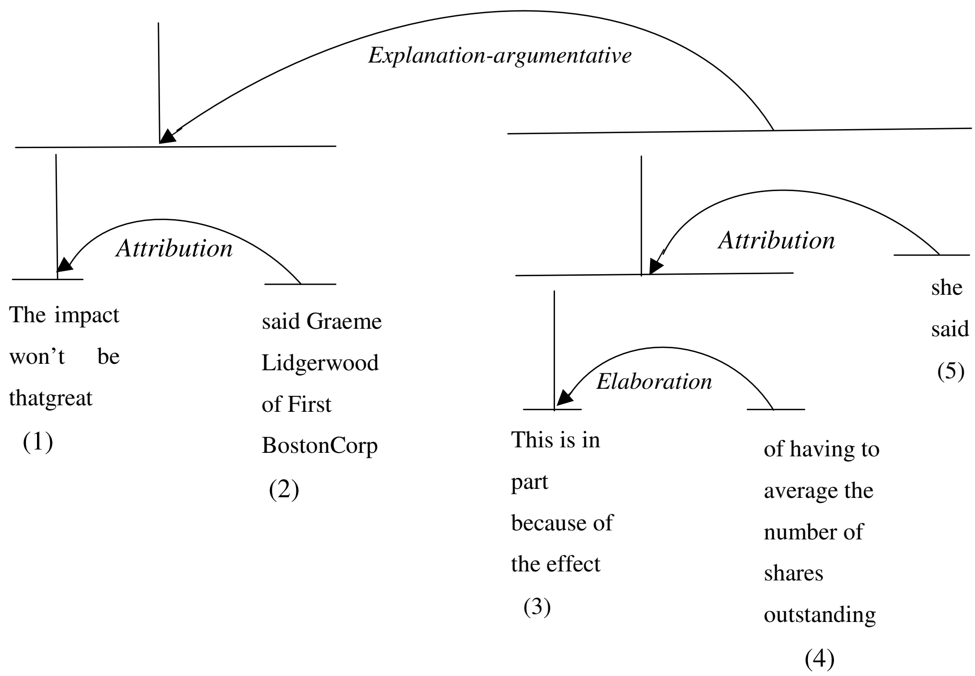

Figure 1. Example of an RST- tree for two sentences in RST–DT.

Discourse units linked bydiscourse relations are annotated as either *nucleus* or *satellite* depending on their relative importancein the text. The nucleus is more importantto the writer purpose, while the satellite provides secondary information about the nucleus. Horizontal lines in the figure indicate discourse units, while vertical lines denote the nucleus. Satellites are linked to their nucleus by curved arrows.Relations thatlinktwo discourse units having different status,a nucleus and a satellite, are calledhypotactic (mononuclear). For instance, in Figure 1, *Elaboration*is ahypotactic relation between the nucleus (EDU3), denoted by vertical line, and the satellite (EDU4). On the other hand,when the relationjointstext segments of equal importance, it is called paratactic(multinuclear).Such relation may link two or more discourse units.

Many resources were createdfollowing the RST principles. The firstRST-annotated corpus for English language is the RST-Discourse Treebank (RST-DT) [Carlson et al., 2003]. It consists of 385wall street journalarticlesselected fromthe Penn Treebank. In the RST-DT, relations annotation is done using a set of 53 hypotactic and 25 paratactic relations. These relations are clustered into16 coarse-grained relations, depending on their rhetorical similarity. For more details about this resource, refer to [Carlson et al., 2003]. It is to be noted that there are similar annotated corpora for Spanish [Da Cunha et al., 2011],Dutch [Van Der Vliet et al.,2011],Portuguese [Pardo et al., 2004] and German [Stede, 2004].

3. THE DATA

ArabicCorpus[1] is a large untagged data collectionavailable online for exploration. The corpus consists of several resources distributed into five categories: newspapers, modern literature, nonfiction, Egyptian colloquial and premodern.The newspapers categorycontains approximately 135 million words from articles published between 1996 and 2010 in different Arabic countries**.**

Werandomly selected 140 documents from the newspapers category, because it covers many topics.The corpus has been rhetorically annotated following the reference manual used to develop RST-DT[Carlson and Marcu, 2001]. Although this manual was used for English language, it can also beapplied to Arabictext withsome adaptations according to the peculiarities of Arabic language. These adaptations include segmentation rules, Arabic discourse markers and discourse relations they can signal.

The annotation of our corpus involved threemain subtasks:1)defining intra-sentence Arabic discourse relationsand organizing them into a new hierarchy, 2)defining the annotation manual,and 3) the manual annotation of all documents. To achieve these tasks, we select 30 documentsfor the first subtask, and 25 other documents for measuring annotation reliability. Since we mainly focus on discourse relation identification, discourse tree building is discarded from the annotation process.

Next section presents our list of Arabic discourse relations, followed by the details of the annotation processin section 5.

4. INTRA-SENTENCEARABIC DISCOURSE RELATIONS

To defineintra-sentence Arabic discourse relations following RST framework, we chose to use a semantically driven approach as done in [Keskes et al., 2014]. We started from the reduced set of 16relationsalready defined within the RST-DT. We refine them based on corpus analysis and Arabic rhetorical senses defined in Arabic rhetoric literature [Abubakre, 1989;Abdul-Raof,2006].

In the RST-DT, relation annotation is done using a set of 78fine-grained relations partitioned into 16 classes: Attribution, Background, Cause, Comparison, Condition, Contrast, Elaboration, Enablement, Evaluation, Explanation, Joint, Manner-Means, Topic-Comment, Summary, Temporal and Topic-Change.

Our aim is to first analyzehow each relation from the previous set is instantiated in our Arabic corpus,and what is its corresponding Arabic rhetorical sense.Secondly, looking for the rhetorical relations already defined in Arabic rhetoric that are instantiated in our corpus, but not included in the relationsset of RST-DT.

FourArabic linguistics' professors are involved in this deep analysis, resulting in a novel hierarchyof seven classes as shown in Figure 2.Two relationsthat we found instantiated in our corpus were added to our list: manner and simile.

[1]http://arabicorpus.byu.edu/index.php

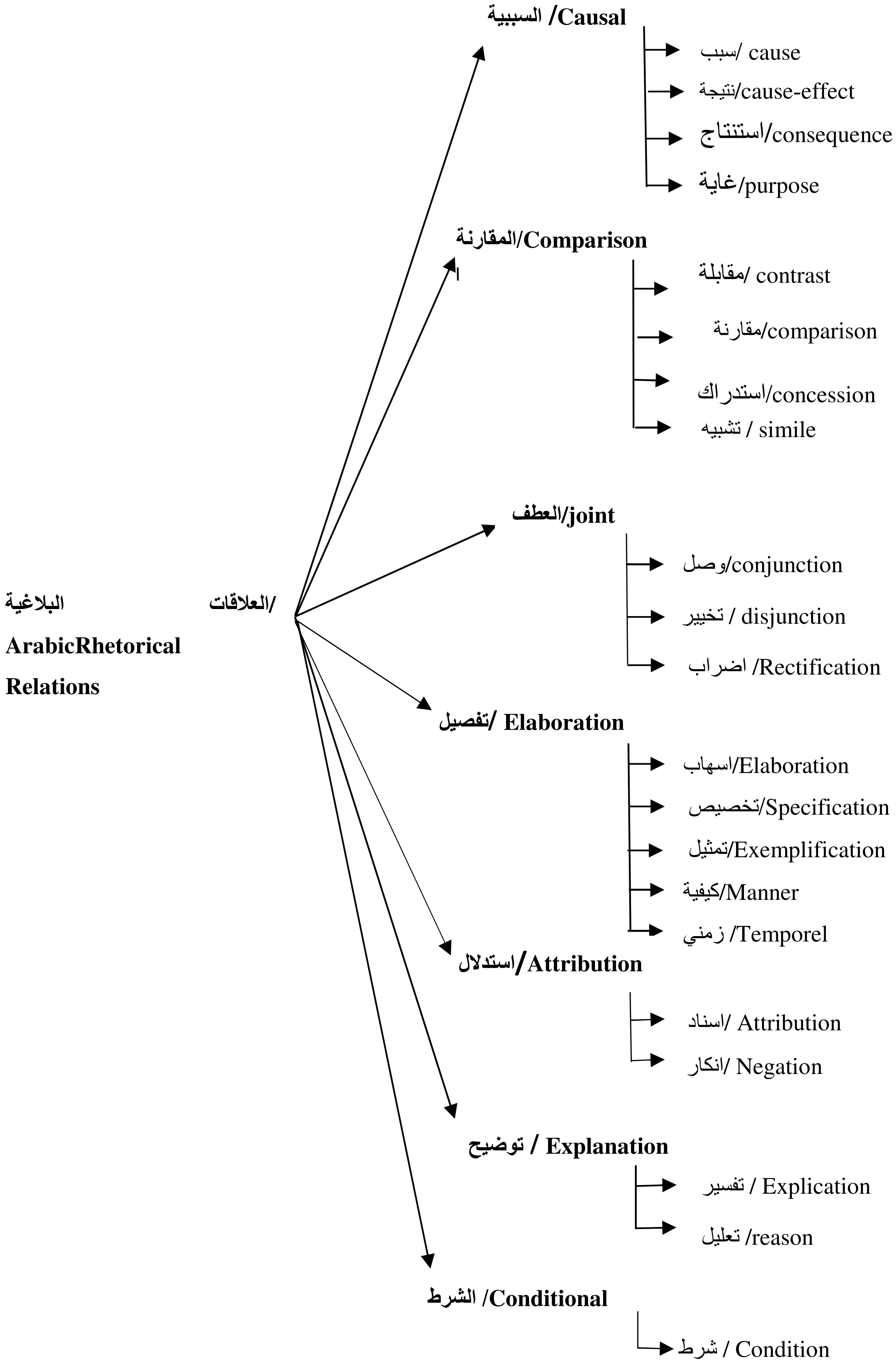

Figure 2. Intra-sentence Arabic discourse relations

Simile in Arabic rhetoric, referred to as (فن التشبيه/ the art of analogy), is an aesthetic mode of discourse who sepragmatic aim sis to bring two significations close to each other and to compare a given entity with another sharing the same feature. For instance, when we say: (وجهه كالقمر/ his face like the moon), we are comparing (وجه/ his face) to (القمر/ the moon) in term of beauty and brightness. Generally, simile in Arabic is used to achieve the function of hyperbole. Therefore, we consider it as a sub-type of comparison relation. Simile in Arabic rhetoric and its different types are clearly detailed in [Abdul-Raof, 2006].

5. ANNOTATION PROCESS

Four linguistics' professors were asked to annotate our corpus. They were provided by an exhaustive annotation manual, in which the meaning of each discourse relations is clearly defined, accompanied with a list of possible markers that signal each relation.Ambiguous markers that can signal multiple discourse relation are also cited with multiple examples. Furthermore, then nu clearity principle in the RST framework and how determining the relative importance of the two text segmentslinked by a specific discourse relation are also discussed.

Annotators were asked to insert discourse relation between two adjacent EDUs, when appropriate, and then assign a rhetorical status (*nuclei/satellite*) to each unit that forms part of a relation, that's what we call nuclearity assignment. Distinguishing between nucleus and satellites is not a simple task. Sometimes, identical contents can result in dissimilar nuclearity assignment, depending on the context.

All annotators went through a training phase first, in which they were encouraged to discuss any disagreements and to give their feedbacks on the manual when necessary.

Then, a blind annotation on 25 documents is performed in order to compute inter-annotator agreement. Four annotators who do not look at each other's annotation annotate each document. Agreement was measured for relation identification and nuclearity assignment. For relation identification, we got a Cohen's kappa of 0.80, and 0.89for nuclearity assignment, which can be interpreted as very good agreements for discourse annotation task.

Finally, and after discussing the key points of disagreements, annotators were asked to reconcile their differencesand produce a unified gold corpus. The mains characteristics of this corpus are presented in Table1.

The total number of annotated relations is 4963. Their distribution in the gold standardis presented in Table2. Each document is annotated using the full set of discourse relation extracted from our corpus enriched with nuclearity annotation resulting in 23 labels(fine-grained relations) such as (سبب/Cause[N][S]), (مقابلة/Contrast[N][N]),..etc. The label (سبب/ Cause[N][S])means that (سبب/cause) relation holds between the left and the right EDUs, where the left EDU is the nucleus and the right EDU is the satellite. These relations are partitioned into 7 classes, coarse-grained relations, that share some type of rhetorical meaning: السببية/causal, المقارنة/comparison, العطف/ joint, اسناد/attribution, تفصيل/elaboration, توضيح/explanation, andالشرط/conditional.

Table 1. Characteristics of the gold corpus

Texts	140
Sentences	2298
EDU	5382

Table2. Distribution of discourse relations in the gold corpus

Coarse-grained relations	**Rhetorical relations**	**Fine-grained relations**	**Frequency**
causal	cause	Cause[N][S]	312
	cause-effect	Cause-effect[N][N]	325
	consequence	Consequence[S][N]	39
	purpose	Purpose[N][S]	311
	Total		**987**
comparison	contrast	Contrast[N][N]	94
	comparison	Comparison[N][N]	433
	concession	Concession[N][S]	309
	simile	Simile[N][N]	98
	Total		**934**
joint	conjunction	Conjunction[N][N]	429
	disjunction	Disjunction[N][N]	25
	rectification	Rectification[N][S]	36
	Total		**490**
Elaboration	elaboration	Elaboration[N][S]	590
	specification	Specification[N][S]	238
	exemplification	Exemplification[N][S]	109
	manner	Manner[N][S]	208
	temporal	Temporal[N][S]	157
		Temporal[S][N]	33
	Total		**1343**
explanation	explication	Explication[N][S]	225
	raison	Raison[N][S]	244
	Total		**469**
attribution	attribution	Attribution[S][N]	524
		Attribution[N][S]	69
	negation	Negation[S][N]	45
	Total		**638**
conditional	condition	Condition[N][N]	**102**

6. PROPOSED MODEL

We propose a supervised learning on amulti-layer perceptron classifier for both explicit and implicit relation recognition.Our instances are composed of adjacentEDUs pairs that are linked by a discourse relation.To perform a supervised learning on the gold standard, we compute a feature vector for each instance.We implemented ten groups of features.Three features groups namely accusative of purpose, specific connectives and number of antonyms words are novel. Al-masdar isinspired from [Alsaif and Markert, 2011].The six remaining features areinspired from prior works that demonstrate theireffectivityfor recognizing explicit and implicit discourse relations [Marcu, 2000; Subba and Di Eugenio, 2009; Huang and Chen, 2011; Keskes et al.,2014].
In this section, a description of each features group is presented. S and N indicate the nucleus and satellite in the illustrative examples in square brackets at the beginning of each (EDU).

F1) Discourse markers

We use three numeric features to encode the discourse markers using a rich lexicon, manually constructed during the annotation process. For each marker we encode its:

-Class: which defines the rhetorical relation that it can signal.
- Type:(unambiguous/ambiguous). Unambiguous markers are those that signal only one specified discourse relation such as بالمقابل/in contrast, من أجل/in order to, لأن/ because. Ambiguous markers on the other hand, are associated with multiple discourse relations. For example the markers حتى/until can signal a temporal relation as in (cf.example.1), and a conditional relation as in (cf.example.2).
- Position: (first EDU/second EDU/ null).

(1)[N] وكلوا واشربوا [S]حتى يتبينلكم الخيط الأبيض من الخيط الأسود من الفجر [1]
(1) [N]And eat and drink [N] until the white thread of dawn becomes distinct from the black thread.
(2) [N]لن تنالوا البر[N]حتى تنفقوا مما تحبون[2]
(2) [N]Never will you attain the good[N]until you spend from that which you love.

F2) Modality

We use two binary features to check the presence of modal verbs in each EDU using a manually constructed lexicon composed of 41 Arabic modal verbs such as كشف /expose, أعلن/announce, قال/say, أكد/*confirm*, أوضح/*explain*. The presence of modal verbs is a strong indicator that signals the relationاسناد/Attribution (cf.example.3).

(3) [S]قال الوزير:[N] ان أسعار النفط في انخفاض مستمر
(3) [S]The minister said:[N]that the price of oil continues to drop.

F3) Specific connectives

We are interested only in three Arabic connectives useful for our task: the connective أن/>n/that, and the justification particles (لـ/l/for, بـ/b/with). A binary feature is used to check the presence of the connective أن />n/that in the beginning of the second EDU.This connective is generally used with modal verbs to express attribution relation. For the justification particles, two numeric features are used to encode the justification particle string and position. These particles are generally prefixed to words to indicate causation or explanation in sentences.

F4) Al-masdar/المصدر

A binary feature is used to check whether the first word of each EDU contains al-masdar noun. Al-masdar is a noun category indicating events without tense such as تجنب/avoidance,اعتذار/apology, استعمال/use. In Arabic texts, justification particle followed by al-masdar nouns signal generally some discourse relation such asغرض/purpose (cf. example (4)), and manner relation (cf.example (5))

[N](4) تواصل الدولة المفاوضات مع النقابات [S]للحد من الإضرابات المتكررة
(4) [N] State continues negotiations with trade unions [S]to curb repeated strikes
(5) [N] تسعى الدولة لتطوير اقتصادها [S]بالاعتماد على مصادر متنوعة
(5) [N] State seeks to develop its economy [S]by relying on a variety of sources

We rely onthe Arabic morphological analyzer[Boudchiche et al., 2017] to construct al-masdar noun using well-known morphological patterns such as فعالة ,تفعيل , فعال .

[1]A verse 187 from Surat al-Baqarah - The Holy Quran.
[2]A verse 92 from Surataal-i-Imraan – The Holy Quran.

F5)Punctuation marks

We have 4 binary features to test for the presence of particularpunctuations ((:) and (,)) as well asof typographical markers («» and ()). These featurescould help us in identifying some discourserelations, such asاسناد/attribution.

F6) Accusatives ofpurpose/ المفعول لأجله

Accusatives of purpose / المفعول لأجله is an indefinite noun in the accusative case *manṣūb* (منصوب) used to specify the purpose, motive or reason behind an action. For example the word تجنبا/ to avoid, in example (6) explains why an urgent meeting is organized yesterday. The presence of accusative of purpose within a sentence strongly signals purpose relation. Thus, a binaryfeatureis used toindicate whether each EDUcontainsanaccusatives of purpose/المفعول لأجله, andanother numeric feature is used to located it(first or second EDU).

(6) [N] عقد اجتماع عاجل مع اللاعبين مساء امس [S]تجنبا لاحتجاجاتهم

(6) [N]An urgent meeting with the players was held yesterday [S]to avoid their protests

F7) Named entities

A binary features is used to indicate whether each text segment containsnamed entities. Such information could be useful for recognizing some discourse relation such as attribution, comparison and elaboration. To check the presence of named entities in the text segment, we use the Arabic gazetteer namely ANER gazet [Benajibaet al., 2007].

ANER gazet is composed of three types of gazetteers: people containing 2309 entries, locations containing 2182 entries and organizations containing 403 entries.

F8) Length feature

This is a numeric feature that denotes which EDU is longer than the other (length in term of words) is. In some relation like تفسير/explanation or اسهاب/elaboration, the second EDU is generally longer than the first. This is why we think that encoding this information could be useful for our task.

F9) Numerical data

We use two binary features to indicate whether each text segment contains numerical datasuch as dates, times, and numbers. It helps us to recognize some discourse relation such asمقارنة/comparison and زمني/temporal relations.

F10) Number of antonym words

Semantic information has a great importance in discourse relation recognition, especially for Arabic language, where the identification of some implicit discourse relation is based mainly on semantic relations between words. For instance, the relation مقابلة/contrast in Arabic (cf.example.7) is signaled by the presence of more than one words in the first text segment and their antonyms in the second text segment.That is why, we define a numeric feature that compute the number of words from the first and second EDUsthat are antonyms.

(7) [N] فَلْيَضْحَكُوا قَلِيلًا [N] وَلْيَبْكُوا كَثِيرًا[1].

(7) [N] So let them laugh a little [N] and weep more.

Antonym words are extracted from Arabic WordNet(AWN), the well-known lexical database for Arabic language.

[1]A verse 82 from Surat Al – Tawbah – The Holy Quran.

7. EXPERIMENTS AND RESULTS

We experimented with a multilayer perceptron classifier, implemented in weka[1]environment using various combinations of features.For all experiments, we usedmultilayer perceptron's default settingswith one hidden layer, as it yielded very good results.All experiments were evaluated using 10-fold cross-validation. We report our experiments on identifyingcoarse-grained relations (7 merged relations) and fine-grained relations (23 relation).As there are few instances for Consequence[S][N],Disjunction[N][N],Rectification[N][N], Temporal[S][N] and Negation[N][S] (as shown by Table 2), we removed these relations from further consideration, and we used only the remaining 18 relations.

Our models are compared to two baselines. The first one is the majority class, all instance are classified as اسهاب /elaboration [N][S] forfine-grained relations classification, and asتفصيل/Elaboration for coarse-grained relation classification. The second baselineis a model based on discourse markers features (F1).

7.1. Overall results

Firstly, we evaluate the performance of our classifiers on fine-grained relationsclassification (18 classes), and on coarse-grained relations classification (7 classes).The majority baseline yielded an accuracy of 20.93 and 51.03, respectively for fine and coarse-grained relationsclassification. The results are presented in Table 3 in terms of macro average F-score and accuracy(number of correctly classified instances over the total number of instances).Precision (P), recall(R)and F-score for each classare calculated as follow:

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + FN}$$

$$\mathrm{F} - \mathrm{score} = \frac{2 * \mathrm{P} * \mathrm{R}}{\mathrm{P} + \mathrm{R}}$$

The macro averaged F-score is computed as:

$F\ (macro - averaged)\ = 1/k \sum_{i=1}^{k} F - score$, where K is the number of class decisions.

Table3. Overall results for fine and coarse-grained relations classifications

	Fine grained classification		Coarse grained classification	
	F score	Accuracy	F score	Accuracy
Baseline1(majority class)	-	20.93	-	51.03
Baseline2(discourse markers)	71.93	74.12	81.69	85.54
Our model	78.49	83.39	86.74	91.03

It is clear that, our models perform notably better than the two baselines for both levels. At fine-grained relations classification, it outperforms baseline2 in term of accuracy and F-score by 9.27 and 6.56,respectively. It also outperforms the baseline1, in term of accuracy by 62.46. Similarly, at coarse-grained relations classification level, the performance of our classifier is higher than the baseline2 in term of F-score by 5.05. For both levels, we can also observe that baseline2 gets very good results compared to the majority baselines.

[1] https://sourceforge.net/projects/weka/

To check the effectiveness of each group of features on fine-grained classification, we run several experiments in which we gradually added each features group to the discourse markers features, which serve as a baseline. Table 4 reports the results of these experiments.

Table 5. Impact of each group of features on fine-grained relations classification

Features	F-score
Baseline2 (F1)	71.93
+Modality (F2)	73.65
+Specific connectives (F3)	73.89
+ Al-masdar (F4)	74.76
+Punctuation marks (F5)	74.76
+Accusative of purpose (F6)	76.03
+Named entities (F7)	76.44
+Length EDU (F8)	76.83
+Numerical data (F9)	77.64
+Antonym words (F10)	78.49

When analyzing the impact of each features group individually, we can seethat overall performance is most impacted by discourse markers and modality features. Indeed,when using discourse markers' features the overall accuracy was improved by 53.19over the majority baseline.Moreover,whenadding modal verbs tobaseline2 the macro average F-score was improved by 1.72and 73.01 over baseline2 and baseline1, respectively.Adding both specific connectives and Al-masdar features improved the F-score by 1.11. A significant improvement was also detected when adding accusative of purpose features. To our surprise, punctuation marks features had no impact, and adding this feature did notlead to any improvement in terms of F-score.It is to be noted that this result is inconsistent with the conclusion reported in [Keskes et al., 2014]. Adding Named entities and length EDU lead to marginal improvement in terms of F-score, while antonym words and numerical data features have almost a similar and significant impact on overall performance.

Once we have evaluated the efficacy of each group of features, we have then measured the performanceof our classifier to predict implicit discourse relations. Our results show that predicting implicit discourse relation is 18.05 (in term of accuracy) lower than its capacity to predict explicit ones. One possible reason is the partial coverage of implicit discourse relations in our gold corpus (21%). Therefore, we believe thatmore instances of implicit discourse relations are required to boost the classification performance.

7.2. Fine-grained classification

Table 6 shows the detailed results in term of F-score for fine-grained relations classification.Columns 3 to 9show the results of different features groupsgradually added to baseline2 (model based on discourse markers F1). The last column shows results using all features. From synthetizing these results, we can say that our model achieved the best performance using all feature groups.

Table 6. Detailed accuracy for fine-grained relations classification

Relation class	(F1)	+F2	+F3+F4	F5	+F6	+F7	+F8	+F9	+F10 (All)
Cause[N][S]	66.32	62.32	65.18	65.18	**69.95**	69.83	70.03	71.10	**71.13**
Cause-effect[N][N]	78.06	78.33	78.21	78.21	82.29	82.17	82.19	82.28	**82.22**
Purpose[N][S]	72.63	72.54	**77.12**	77.11	**84.04**	84.04	84.25	84.22	**84.19**
Contrast[N][N]	45.28	46.17	46.22	46.22	46.18	46.22	48.12	47.71	**61.48**
Comparison[N][N]	78.25	79.02	79.13	79.13	79.41	**79.66**	80.12	**86.23**	**88.27**
Concession[N][S]	73.32	73.27	74.01	74.01	74.12	75.08	75.19	75.22	**75.31**
Simile[N][N]	56.45	56.45	55.11	55.11	56.22	58.03	58.11	58.02	**58.00**
Conjunction[N][N]	76.23	76.28	76.19	76.17	76.31	76.31	76.22	76.22	**76.27**
Elaboration[N][S]	78.15	78.65	78.66	78.66	79.66	**81.02**	82.17	82.22	**82.19**
Specification[N][S]	85.79	85.82	85.84	85.84	85.87	85.88	86.25	86.23	**86.09**
Exemplification[S][N]	84.36	84.36	84.28	84.28	84.28	85.12	85.32	85.32	**85.31**
Manners[N][S]	69.25	69.22	**80.72**	80.72	80.49	81.38	81.22	81.51	**81.33**
Temporal[N][S]	75.97	75.97	75.95	75.96	75.96	75.96	75.98	**83.49**	**83.46**
Explication[N][S]	79.35	79.39	76.49	76.48	**79.88**	**80.12**	82.82	82.82	**82.76**
Reason[N][S]	73.23	73.14	76.88	76.88	**78.21**	78.18	78.21	78.21	**78.03**
Attribution[S][N]	79.95	**98.21**	98.21	98.23	98.23	98.23	98.25	98.25	**98.28**
Attribution[N][S]	67.23	**81.62**	81.29	81.29	81.29	82.33	82.18	82.18	**82.26**
Condition [N][N]	54.94	54.94	56.24	56.24	56.24	56.47	56.41	56.40	**56.41**
Macro average F-score	71.93	73.65	74.76	74.76	76.03	76.44	76.83	77.64	**78.49**

More specifically, we found these features had severalinfluencesover different discourse relations. For instance,adding modal verbs feature (F2)highly influenced the performance of اسناد/Attribution relation, because thecorresponding F-scores of the relations Attribution[S][N] and Attribution[N][S] wereincreased by 0.18 and 0.15, respectively over baseline2. While adding specific connective (F3)and Al-masdar (F4) features boosted the performance of Manner[N][S] relation by 0.09 andغرض/purpose[N][S]by 0.04 over baseline2+F2. They also increased the F-score of تعليل/reason[N][S] by 0.03.On the other hand, accusative of purpose features (F6) highly improved the performance of bothغاية/purpose[N][S] and تعليل/reason[N][S] relations, while the same feature had a negative impact on relationتفسير/explication[N][S].Unlike the previous features, punctuation marks' features (F5) had marginalimpact on almost all relations except attribution[S][N], which was slightly improved by 0.01. We could explain this by the fact that punctuation marks are not widely used, and partially taken into account in Arabic texts.Hence, relying on this feature provided a very slight or even negative impact on relationsprediction.

On another hand,both named entities (F7) and length features (F8) boosted the performance of مقارنة/comparison[N][N],اسهاب/elaboration[N][S] and تفسير/explication[N][S] relations, while numerical data feature(F9) had a good impact only on theزمني/temporal[N][S] relation.

Concerning the others relations, it is noticeable that the relation مقابلة/contrast[N][N] reached its best performance when adding antonym words features (F-score 61.48). This is consistent with the definition of this relation in Arabic that hold when there is more than one words in the first text segment and their corresponding antonyms in the next segment. Adding this feature clearly distinguished between comparison and contrast relations and boosted their performance by 0.13 and 0.03, respectively.

In general, we can say that each features group had a different effect on the prediction of different discourse relations. Some features are crucial for predicting certain relations and at the same time, they had a slight or even negative impact on others. Discourse markers feature (F1) is very useful for explicit discourse relations, while modality (F2), accusative of purpose (F6), named entities (F7) and antonym words (F10) could be utilized for implicit ones.

Error analysis at this level showed that our classifier failed todiscriminate wellbetween the relations تشبيه/simile[N][N]and تمثيل/exemplification[S][N] (cf. example 10), and also between سبب/Cause[N][S]and اسهاب/Elaboration[N][S]when they are implicitly signaled.

(10)[1] مثل الذين ينفقون أموالهم في سبيل الله [2]كمثل حبة أنبتت سبع سنابل[1]

(10)[1] *The example of those who spend their wealth in the way of Allah [2] is like a seed which grows seven spikes*

The rhetorical relation holding between the discourse unite [1] and [2] is تشبيه/simile[N][N].However the predicted relation by our classifier wasتمثيل/exemplification[S][N].

7.3. Coarse-grained classification

Table 6 presents the detailed results for coarse-grained relationsclassification in term of precision, recall and F-score. The last row presents the average precision, recall, and F-score.

The proposed model achievedvery good results with an F-score of 86.74% and an overall accuracy of 91.05%. Furthermore, the performance of our model did not widely varied across relations' classes. The best performance was achieved by the relation attribution with an F-score of 98.32%, while the lowest results was achieved by conditional relation with an F-score of 67.71%.

Table 6. Detailed accuracy for coarse-grained relations classification

Relation class	**Precision**	**Recall**	**F-score**
Causal	90.03	94.10	92.02
Comparison	87.48	93.05	90.17
Joint	82.24	94.86	88.10
Elaboration	85.71	96.28	90.68
Attribution	99.27	97.39	98.32
Explanation	81.05	79.41	80.22
Conditional	66.32	69.18	67.71
Macro average	**84.58**	**89.18**	**86.74**

Errors analysis showed that the majority of the errors are between تفصيل/elaboration and شرط/conditional relations. We believe that this is due to the distribution of these relations in the gold corpus. Since the relation تفصيل /elaboration was the most predominant relation, the classifier tended to classify ambiguous relations as تفصيل /elaboration, which explains the high recall and low precision for this relation.

It is worth pointing out that our work differs from that of [Alsaif and Markert, 2011] in many respects.Firstly, our system identified both explicit and implicit relations between adjacent EDUs within the RST framework instead of only explicitly marked relations within the PDTB model (a different theoretical framework). In addition,their system applied JRIP classifier, whereas our system used a multilayer perceptron. Furthermore, our discourse-annotated corpus was composed of news articles selected from Arabic corpus instead of Arabic Penn Treebank. Finally, we performed classification at the more fine-grained level (relations enriched with nuclearity annotations) and at coarse-grained level using a different set of features.

8. RELATED WORK

[1] A verse 261 from Surat al-Baqarah - The Holy Quran

We present in this section, an overview of the main computational approaches on discourse relation recognition that follow RST as well as the PDTB model as theories of discourse.
Marcu and Echihabi[2002] presented the first unsupervised learning approach to identify four classes of RST relations: Contrast, Explanation-Evidence, Condition and Elaboration. They were the first to use word pair feature calculated from the two segments of relation.Extending Marcu and Echihabi'swork, Saito et al. [2006]further combined word pairs and phrasal patterns to identify implicit discourse relations in Japanese. Several authors have also proposed Semi-supervised approaches that exploited both labeled and unlabeled data. Hernault et al. [2010a]proposeda method based on the co-occurrence of features observed in unlabeled data. The authors used state of the art features including word pairs, production rules from the parse trees, as well as lexical heads. They showed that this method improved significantly classification accuracyfor infrequent relations.
With the release of discourse-annotated corpora, there was an opportunity to address this area of work using supervised learning approaches. On the RST-DT, Soricut and Marcu[2003] presented a sentence-level discourse parser (SPADE) based on lexical and syntactic features extracted from the lexicalized syntactic tree of sentence. The system had revealed empirically the correlation between syntax and discourse structure. Hernault et al. [2010b] presented a hight-level Discourse Analyzer(HILDA), a fully implemented discourse parser based on Support Vector Machines. For relation labeling, a multi-class SVM classifier was used. Several shallow lexical and syntactic features were considered including structural, lexical, and organizational features. Feng and Hirst [2012] extended this work by incorporating more linguistic features for text-level discourse parsing such as Discourse Production Rules to reflect the relatedness between different discourse relations, semantic similarities for verbs and nouns separately and more contextual features. An important decision was the discrimination between intra-sentential and inter-sentential relations, which implied the specification of features for each level;Joty et al. [2015] also addressed this in the implantation of their CODRA discourse parser using Conditional Random Fields.

Recently, the PDTB [Prasad et al., 2008] has presented as the largest discourse annotated corpus following thePDTB model. It provide a clearer and more exhaustive implicit relation annotation and a valuable platform for researchers to develop discourse-centric systems. The first work that tackle implicit discourse relation recognition on PDTB was[Pilter et al., 2009]. The author used various surface and linguistically informed features such as modality, verb classes and polarity. Extending their work, Lin et al. [2009] provideda classification of implicit relations on the second-level type in the PDTB, using four class of features: word pair, arguments' context, arguments' internal constituent and dependency parses. Zhou et al. [2010] proposed a method that automatically predicted implicit connectives using language models, and then these connectives were used as additional features to recognize implicit relations.Xu et al. [2012] extended this work by employing further linguistically informed features.

With the remarkable results achieved by deep learning models in natural language processing [Socher et al., 2013; Kim, 2014; Cao et al., 2015], researches moved towards the use of deep neural networks models, and related features representation methods for implicit discourse relations recognition. Zhang et al. [2015] proposed a shallow convolutional neural network (SCNN) with one convolution layer to classify implicit discourse relation. Ji and Eisenstein [2015] employed two recursive neural networks to learn distributed representation of arguments and entity spans. To overcome the problem of data sparsity and argument representation,Liu et al. [2016] proposed a Convolutional Neural Network embedded multi-task learning system to synthesize different discourse analysis tasks, using three corpora: PDTB, RST-DT, and New York Times (NYT). Recently, Li et al. [2017] proposed a max-margin based neural network model that took into account the relation transformation property and the interactions between arguments. The authors also proposed to learn distributed features representations from words, arguments, and syntactic structures to sentences.

Compared to the large body of studies on English discourse relation recognition, there is very few works for Arabic.Alseif and Market [2011] proposed supervised algorithms to identify discourse connectives and explicit relations that hold between adjacent EDUs within the PDTB model. The author used some features already used for English implicit relation recognition. Experimental results showed that production rule features did not have a good effect on the classification performance, since the overall accuracy for fine-grained relations classification was degraded from 77% to 76% when using these features. Kesekes et al. [2014] extended their works by addressing both explicit and implicit relations holding between adjacent as well as non-adjacent units within the SDRT framework. The authors used the same features used by Alseif and Market[2011]except production rules features as well as punctuation, Contextual, Lexico-semantic and lexical features that have been successfully employed for English relation recognition. All features were automatically extracted from the Discourse Arabic Treebank corpus (D-ATB).The proposed model achieved good results with an accuracy of 77.8% on fine-grained discourse relations and an accuracy of 82.8% on class-level discourse relations. Finally, Sadek and Meziane [2016] proposed a *Pattern Recognizer* model to signal the presence of causal relations within sentences. The model includes approximately 700 linguistic patterns constructed based on various syntactic features. Experimental results have revealed the extreme importance of justification particles in detecting Arabic *Causal* relations.Which motivated us to further incorporate these particles in our study to predict explicit causal relations.

9. CONCLUSIONS AND FUTURE WORKS

In this article, the first model to automaticallyidentifying both explicit and implicit rhetorical relations among intra-sentence discourse segments in Arabic was presented.

The model was built using a supervised learning on a multi-layer perceptron classifier. The challengewas the lack of RST annotated corpora in Arabic. To overcome this problem, we annotated a large corpus following the rhetorical structure theory framework.Annotation was done using a set of 23fine-grained relations enriched with nuclearity annotation.

To automatically learn these relations, we explored several types of features, and contributed novel features including accusatives of purpose features, specific connectives and the number of antonyms words.

The experimental results showed that our model achievedexcellent results for both configurations (fine and coarse-grained relations), and substantially outperformedall the baselines.

It is worth noting that the model presented here constitutesthe main componenttoward developing a sentence-level discourse parser for Arabic. Furthermore, we believe that this model is useful to carry out several Arabic Naturel Language processing (ANLP) applications such as automatic text summarization and sentence compression.

As future work,we plan to explore deep neural network models and related features representation methods. Additionally, we will devote efforts to extend this work by developing the first sentence-level discourse parser for Arabic text.

10. REFERENCES

Abubakre, R.D., 1989, Bayan in Arabic Rhetoric: An analysis of the core of Balāgha. *Ibadan: Intec Printer Limited.*

Abdul-Raof, H., 2006, Arabic Rhetoric. A Pragmatic Analysis. Routledge, ISBN10: 0–415–38609–8.

Alsaif, A.,Markert, K., 2010, The Leeds Arabic Discourse Treebank: Annotating Discourse Connectives for Arabic. In *LREC*.

Alsaif, A.,Markert, K., 2011, Modelling discourse relations for Arabic. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 736-747). Association for Computational Linguistics

Biran, O., McKeown, K., 2013, Aggregated Word Pair Features for Implicit Discourse Relation Disambiguation. In *ACL (2)*, 69-73.

Braud, C., Denis, P., 2015, Comparing word representations for implicit discourse relation classification. In EMNLP *2015.*

Boudchiche, M., Mazroui, A., Bebah, M.O.A.O., Lakhouaja, A.,Boudlal, A., 2017, AlKhalilMorpho Sys 2: A robust Arabic morpho-syntactic analyzer. *J. King Saud University-Computer and Information Sciences*, ***29***, 141-146.

Benajiba, Y., Rosso, P., &Benedíruiz, J., 2007, Anersys: An arabic named entity recognition system based on maximum entropy. *Computational Linguistics and Intelligent Text Processing*, 143-153.

Cao, Z., Wei, F., Dong, L., Li, S., Zhou, M., 2015, January Ranking with Recursive Neural Networks and Its Application to Multi-Document Summarization. In *AAAI*, 2153-2159.

Cimiano, P., Reyle, U., Šarić, J., 2005, Ontology-driven discourse analysis for information extraction. *Data Knowl. Eng.**55***, 59-83.

Carlson, L., Marcu, D.,2001, Discourse tagging reference manual. *ISI Technical Report ISI-TR-545*, *54*, 56.

Carlson, L., Marcu, D., Okurowski, M. E. , 2003, Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Current and new directions in discourse and dialogue*, 85-112, Springer Netherlands.

Da Cunha, I., Torres-Moreno, J. M., Sierra, G., 2011, On the development of the RST Spanish Treebank. In *Proceedings of the 5th Linguistic Annotation Workshop*, 1-10, Association for Computational Linguistics.

Feng, V. W., Hirst, G., 2012, Text-level discourse parsing with rich linguistic features. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers- **1***, 60-68, Association for Computational Linguistics.

Hernault, H., Bollegala, D., Ishizuka, M., 2010a, October, A semi-supervised approach to improve classification of infrequent discourse relations using feature vector extension. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing* 399-409, Association for Computational Linguistics.

Hernault, H., Prendinger, H., Ishizuka, M., 2010b, HILDA: A discourse parser using support vector machine classification. *Dialogue & Discourse*, ***1***.

Huang, H. H., Chen, H. H., 2011, Chinese Discourse Relation Recognition. In *IJCNLP*, 1442-1446.

Joty, S., Carenini, G., Ng, R. T., 2015, CODRA: A novel discriminative framework for rhetorical analysis. *Computational Linguistics*.

Jansen, P., Surdeanu, M., Clark, P., 2014, Discourse Complements Lexical Semantics for Non-factoid Answer Reranking. In *ACL (1)*, 977-986.

Ji, Y., Eisenstein, J., 2015, One vector is not enough: Entity-augmented distributed semantics for discourse relations. *Transactions of the Association for Computational Linguistics* 3, 329–344.

Keskes, I., Zitoune, F. B., Belguith, L. H., 2014, Learning explicit and implicit Arabic discourse relations. *J. King Saud University-Computer and Information Sciences*, ***26***, 398-416.

Kim, Y., 2014, Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Lin, Z., Kan, M. Y., Ng, H. T., 2009, Recognizing implicit discourse relations in the Penn Discourse Treebank. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, 343-351, Association for Computational Linguistics.

Louis, A., Joshi, A., Prasad, R., Nenkova, A., 2010, Using entity features to classify implicit discourse relations. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 59-62, Association for Computational Linguistics.

Li, H., Zhang, J., Zong, C., 2017, Implicit Discourse Relation Recognition for English and Chinese with Multiview Modeling and Effective Representation Learning. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, ***16***, 19.

Li, J., Li, R.,Hovy, E. H., 2014, Recursive Deep Models for Discourse Parsing. In *EMNLP*, 2061-2069.

Liu, Y., Li, S., Zhang, X., Sui, Z., 2016, Implicit Discourse Relation Classification via Multi-Task Neural Networks. In *AAAI*, 2750-2756.

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***, 243-281.

Marcu, D., 2000, The rhetorical parsing of unrestricted texts: A surface-based approach. *Computational linguistics*, ***26***, 395-448.

Marcu, D., Echihabi, A., 2002, July, An unsupervised approach to recognizing discourse relations. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 368-375. Association for Computational Linguistics.

Mihaylov, T., Frank, A., 2016, Discourse Relation Sense Classification Using Cross-argument Semantic Similarity Based on Word Embeddings. In *CoNLL Shared Task*, 100-107.

Pardo, T., das Graças Volpe Nunes, M.,Rino, L., 2004, Dizer: An automatic discourse analyzer for brazilianportuguese. *Advances in Artificial Intelligence–SBIA 2004*, 224-234.

Park, J., Cardie, C., 2012, Improving implicit discourse relation recognition through feature set optimization. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue* ,108-112, Association for Computational Linguistics.

Prasad, A., Miltsakaki, R., Dinesh, E., Lee, N., Joshi, A., Webber, B., 2008, The Penn discourse Treebank 2.0. In LREC.

Pitler, E., Raghupathy, M., Mehta, H., Nenkova, A., Lee, A., Joshi, A. K., 2008, Easily identifiable discourse relations. In Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008).

Pitler, E., Louis, A., &Nenkova, A., 2009, Automatic sense prediction for implicit discourse relations in text. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, 683-691, Association for Computational Linguistics.

Sadek, J., &Meziane, F. (2016). Extracting arabic causal relations using linguistic patterns. *ACM Transactions on Asian and Low-Resource Language Information Processing*, ***15***, 14.

Saito, M., Yamamoto, K., Sekine, S., 2006, Using phrasal patterns to identify discourse relations. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, 133-136, Association for Computational Linguistics.

Somasundaran, S., Namata, G., Wiebe, J., Getoor, L., 2009, Supervised and unsupervised methods in employing discourse relations for improving opinion polarity classification. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, 170-179. Association for Computational Linguistics.

Sporleder, C., Lascarides, A., 2007, Exploiting linguistic cues to classify rhetorical relations. *AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES 4*, *292*, 157.

Stede, M., 2004, The Potsdam commentary corpus. In *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, 96-102, Association for Computational Linguistics.

Subba, R., Di Eugenio, B. 2009, An effective discourse parser that uses rich linguistic information. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 566-574). Association for Computational Linguistics.

Soricut, R., Marcu, D. 2003, Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1* (pp. 149-156). Association for Computational Linguistics.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., Potts, C., 2013, Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, 1631-1642.

Taboada, M., Mann, W. C. (2006). Rhetorical structure theory: Looking back and moving ahead. *Discourse studies*, ***8***, 423-459.

Tu, M., Zhou, Y.,Zong, C., 2013, A Novel Translation Framework Based on Rhetorical Structure Theory. In *ACL (2)*, 370-374.

Van Der Vliet, N., Berzlánovich, I., Bouma, G., Egg, M., Redeker, G., 2011, Building a discourse-annotated Dutch text corpus. *BochumerLinguistischeArbeitsberichte*, *3*, 157-171.

Wolf, F., Gibson, E., 2005, Representing discourse coherence: A corpus-based study. *Computational Linguistics*, ***31***, 249-287.

Wellner, B., Pustejovsky, J., Havasi, C., Rumshisky, A., Sauri, R., 2009, Classification of discourse coherence relations: An exploratory study using multiple knowledge sources. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, 117-125. Association for Computational Linguistics.

Xu, Y., LAN, M., Lu, Y., Niu, Z. Y., Tan, C. L. 2012, Connective prediction using machine learning for implicit discourse relation classification. In *Neural Networks (IJCNN), The 2012 International Joint Conference on*, 1-8, IEEE.

Zhang, B., Su, J., Xiong, D., Lu, Y., Duan, H., & Yao, J., 2015, Shallow Convolutional Neural Network for Implicit Discourse Relation Recognition. In *EMNLP*, 2230-2235.

Zhou, Z. M., Xu, Y., Niu, Z. Y., Lan, M., Su, J., Tan, C. L., 2010, Predicting discourse connectives for implicit discourse relation recognition. In *Proceedings of the 23rd International Conference on Computational Linguistics,* 1507-1514, Association for Computational Linguistics.