

A Discourse-Based Approach for Arabic Question Answering

JAWAD SADEK, University of Salford¹
FARID MEZIANE, University of Salford

The treatment of complex questions with explanatory answers involves searching for arguments in texts. Because of the prominent role that discourse relations play in reflecting text-producers' intentions, capturing the underlying structure of text constitutes a good instructor in this issue. From our extensive review, a system for automatic discourse analysis that creates full rhetorical structures in large scale Arabic texts is currently unavailable. This is due to the high computational complexity involved in processing a large number of hypothesized relations associated with large texts. Therefore, more practical approaches should be investigated. This paper presents a new Arabic *Text Parser* oriented for question answering systems dealing with لماذا “why” and كيف “how to” questions. The *Text Parser* presented here considers the sentence as the basic unit of text and incorporates a set of heuristics to avoid computational explosion. With this approach, the developed question answering system reached a significant improvement over the baseline with a *Recall* of 68% and *MRR* of 0.62.

• Information systems → Information retrieval • Computing methodologies → Natural language processing

Additional Key Words and Phrases: Arabic Question Answering, Discourse Analysis, Information Extraction.

ACM Reference Format:

Jawad Sadek and Farid Meziane. 2016. A Discourse-based Approach for Arabic Question Answering.

1. INTRODUCTION

There is a high demand for systems that could return a precise answer to a user's query and avoid the thousands of links returned by traditional search engines. Finding answers to factoid questions such as (what, who, where) involves detecting noun phrases in texts. In contrast, non-factoid questions are expected to provide answers in the form of a meaningful discourse segment, examples of this type are “why” and “how to” questions. Early studies in this domain reported that non-factoid questions require fine-grained text analysis and reasoning capabilities (Kupice, 1999; Breck et al., 2000; Bernardi et al., 2003). Moreover, they suggested that a wise exploitation of linguistic knowledge (i.e. the knowledge about discourse structure) would allow Question Answering (QA) systems to answer this type of question.

However, writing has always been considered as a complex and demanding mental activity undertaken by human beings. This is because of the huge variety of linguistic forms used by writers to achieve their communicative objectives in addition to the tricky nature of the text itself which frequently develops into debatable issues when it comes to grasping these intentions. Accordingly, deriving hierarchical structures of this kind of rich medium is a time-intensive effort.

In a previous work, we developed a system for answering “why” and “how to” questions (Sadek et al., 2012) in which we employed Rhetorical Structure Theory (RST) and used cue phrases to both determine the elementary units and the set of rhetorical relations. The experiment was conducted on short texts (approximately 200 words) derived mainly from Arabic news websites and achieved a *Recall* of 55%. However, we argued that handling larger specialized texts would reduce the overall performance. On the one hand, the computational cost of hypothesising all possible relations within a large text is high, and on the other hand, cue phrases alone are unable to handle all syntactical categories and lexical items embedded within sentences of a specialized text.

Consequently, it is crucial to adopt an improved methodology that would be able to reduce the search space and to cover the great diversity of syntactical structure of the Arabic language. This improvement can be achieved by decomposing the task of discourse structure derivation into two sub-tasks; detecting relations within sentences and locating relations between sentences. The problem of discovering *intrasentential* relations (those existing within a sentence) has been studied by Sadek and

Authors' address: J. Sadek, School of Computing Science and Engineering, University of Salford, M54WT, UK; email: j.sadek@hotmail.com; F. Meziane, School of Computing Science and Engineering, University of Salford, M54WT, UK; email: f.meziane@salford.ac.uk.

Meziane (2016) in which a set of linguistic patterns was constructed to discover discourse information in a single sentence.

In the present study we take one step forward and build our *Text Parser* on top of sentences already annotated with *intrasentential* relationships. The *Text Parser* incorporates a list of discourse markers as relation indicators. Furthermore, a set of heuristics scores are incorporated so that the *Text Parser* produces the most suitable text structure in the framework of RST. In the context of this work, كيف “how to” refers to the type of questions that enquire about the manner in which something is done. The assumption that underlies the process of text structure annotation stems from the fact that the text is well-constructed i.e. cohesive and coherent.

2. ANSWER EXTRACTION

RST is a framework developed by Mann and Thompson (1988) which represents the structure of text in the form of a hierarchical tree. The text is broken down into parts called textual spans that are subdivided recursively until the smallest text spans are reached and these are called elementary discourse units (EDUs). RST labels relations between adjacent spans on different levels (clauses, sentences, paragraphs) using specific relationships. The span that is more important for the writer is called *Nucleus* and the other, which is considered less essential, is called *Satellite*. If both text spans have the same importance to the reader, the relation is called paratactic; however, if one span is more important to the reader than another, the relation is called hypotactic.

Consider text (1) which explains how we employ discourse structure to find and extract answers to some kind of questions. The text is segmented into two discourse units (DUs) each with the length of a full sentence.

- (1) [قالت دراسة نشرت في صحيفة بريتيش ميديكال إن الشاي الأسود الذي تم إعداده عند درجة حرارة تزيد عن 70 درجة مئوية يزيد من خطر الإصابة بالسرطان.]¹ [وعليه يمكن تفسير ارتفاع الإصابة بسرطان المري بين بعض الشعوب الغير غربية.]²

[The research published in the British Medical Journal found that black tea made at temperature greater than 70 °C, can raise the risk of cancer.]¹[and that may be the cause of high rates of esophageal cancer among non western people.]²

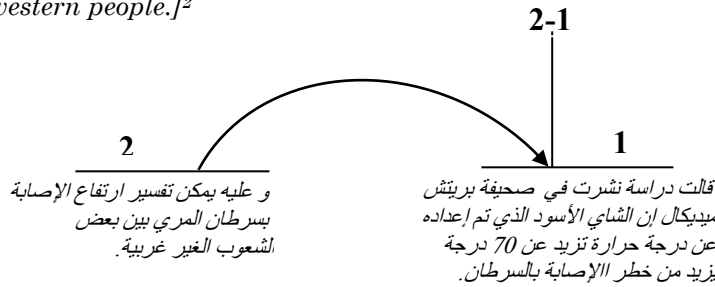


Figure 1: A schema representation of text (1).

In the case of the following question:

{ لماذا تعد الإصابة بسرطان المري مرتفعة بين الشعوب الغير غربية؟ }

{Why does esophageal cancer has high rates among non western people?}

We notice that the question corresponds to unit₂, so the other part of the relation will be the answer for the question, which is the unit₁.

3. DISCOURSE MARKERS

Discourse Markers (DMs) form a heterogeneous class of words and expressions which draw mainly from the categories of conjunctions, prepositional and adverbial phrases. DMs have an important linking function that link adjacent segments of discourse together to achieve coherence and cohesion. More importantly, DMs are frequently used by writers to avoid possible unintended interpretations of the text. Most studies tackling the task of automatic discourse analysis share the assumption that DMs are the most important type of signals in texts whose function is primarily to link linguistic units at any level, i.e. the main function of DMs are to structure the discourse (Schneuwly, 1997; Sanders and Noordman, 2000, Marcu, 2000b).

However, adopting a specific list of DMs is a challenging task, as a given word or expression may be classified as a DM by one researcher but not by another (Farsler, 1996; Schiffrin et.al. 2001; Blakemore, 2003). This is due to the disagreement among researchers on the features and functions that exactly constitute a DM. Thus, there is no generally agreed list recognized by all researchers.

A number of studies in linguistic literature refer to Arabic DMs broadly in the course of their research while discussing other language phenomena (Wright, 1896; Fareh and Hamdan, 1999). Very few studies focused on the analysis of the role DMs can play to tie units together at the discourse level. A more recent account was proposed by Al Kohlani (2010) in which she identified a list of Arabic DMs used in opinion articles. She utilized RST in discovering the functional relations that occur between sentences. As such, the outcome of her analysis should be consistent with our methodology since we employed the same framework. Moreover, the style of texts she investigated i.e. opinion articles is characterized as being of an argumentative and evaluative nature that aim to influence reader's perceptions of facts and events. This implies that, whenever writers seek to argue facts or express point of view, they tend to use the same DMs. Accordingly, employing these DMs is particularly useful for the objective of the present study.

4. RECOGNIZING DISCOURSE RELATIONS

The *Text Parser* takes as input a set of discourse units each of which extends to a full sentence length associated with intrasentential relations, and outputs all possible rhetorical relations that may hold between these sentences. In most cases, a sentence is directly rhetorically related to a sentence that occurred before or after it.

4.1 Recognition of Adjacent Relations

The recognizer first discovers rhetorical relations between adjacent sentences using linguistic devices which were specifically gathered from (Al Kohlani, 2010). The recognizer scores each of the identified rhetorical relations according to its likelihood that this relation actually holds, and to its significance in building the text structure. Heuristic scores are discussed in Section 5.1.

In most cases the absence of discourse markers correlates with a preference to consider the statement in the unmarked sentence as a continuation of the topic of the sentence that precedes it (Segal et al., 1991). Two possible relations can be hypothesized to hold between two unmarked sentences. The first one is *Elaboration* relation when a pair of sentences tackles the same point. The second one is *Joint* which can be assumed to exist in case a topic shift occurs at the boundary between two sentences.

Arabic writers use demonstrative pronouns frequently to refer to the idea (question, proposition or event) that has been posed in the preceding context (Zaki, 2011). In this regard, demonstrative pronouns - which normally precede a noun made definite by prefixing the definite article- play a particularly useful role as referring expressions. Thus, the sentence in which they are located elaborates on the preceding one. The demonstrative pronoun “هذه” (this, feminine) that appears at the head of sentence (2) of text (2) illustrates this fact.

- (2) [تفحص الطائرات بشكل دوري للتأكد فيما إذا كان هناك خلل في أي جزء من جسم الطائرة.¹] [هذه الاختبارات ضرورية لتجنب أي مشاكل محتملة.²]

[The Aircraft is inspected regularly for any damage to any part of the fuselage.¹] [These checks are crucial in order to avoid any potential problem.²]

However, a demonstrative pronoun can also be used to refer to some other entity appearing in the same sentence. Consider for example text (3) in which the pronoun “هذا” (this, masculine) refers to the idea stated at the beginning of the sentence. This effect can be attributed to the position of the pronoun as it is located approximately in the middle of the sentence. During our experiments, we observed that whenever a demonstrative pronoun occurs within a window comprising the first third of a sentence, it most likely refers to an entity located in the previous sentence. After all rhetorical relations are hypothesized; a *Joint* relation is applied to connect all adjacent sentences that are not connected by other relations.

- (3) من المرجح أن تزيد كمية المعطيات المتاحة للتحليل والتقييم بشكل كبير مع مرور الوقت وهذا الأمر يعني فسح المجال لفرض عمل تتطلب التفكير بطريقة حاسوبية من أجل ترتيب المعلومات وجعلها قابلة للاستخدام.

The amount of data available for evaluation and analysis is likely to increase drastically with the passage of time and **this** means an opening of job opportunities that require computational thinking in order to sort out the information and make it usable.

4.2 Recognition of Distance Relations

Given our commitment to the assumption we made i.e. the text to be annotated is well-constructed, it is possible that one sentence in the middle of the text might be related to another in the beginning. In his work, Marcu (2000a) associated each discourse marker with the feature “*Maximal distance*” which specifies the number of sentences that separate the textual units that are related by the discourse marker. However, the outcome of this approach comes at the cost of computational complexity, as the number of hypothesized relations increases, the number of sub trees increases exponentially. Corston-Oliver (1998) used a different method by checking all pairs of clauses in the text in an effort to hypothesize all possible discourse relations. These hypothesized relations are then grouped into bags of mutually exclusive relations. Nevertheless, for large texts, the time complexity for examining the constraints corresponding to all possible relations could be also high.

Transitivity in natural languages contributes to annotating this sort of relation. Discourse transitivity reveals that there is an implicit relation over hypotactic relations. The sentences in text (4) demonstrate this fact. We notice that sentence (2) elaborates the idea mentioned in sentence (1); also, the discourse marker “لذلك” “Therefore” signals a rhetorical relation of *Result* between sentences (3) and (2). However, the information stated in sentence (3) is still considered as a result of the idea mentioned in sentence (1). Hence, according to the transitivity property, we can say that a hypotactic relation of *Result* also holds between sentences (3) and (1). The graph in Figure 2 shows the discourse analysis of text (4).

- (4) [يعتقد خبراء التجميل أن البشرة الصافية والخالية من البقع وحب الشباب والتجاعيد والمسامات العريضة هي من مقومات الجمال.¹] [حتى ذهب البعض إلى القول "لا جمال بدون بشرة جميلة".²] [لذلك نرى اهتمام الجميع بالمحافظة على بشرة جميلة.³]

[Beauty experts believe that one of the fundamentals of beauty is to have a skin that is free of spots, acne and wrinkles.]¹ [Some even went as far as saying: “there is no beauty without a beautiful skin”.]² [Therefore, everybody is keen about having a beautiful skin.]³

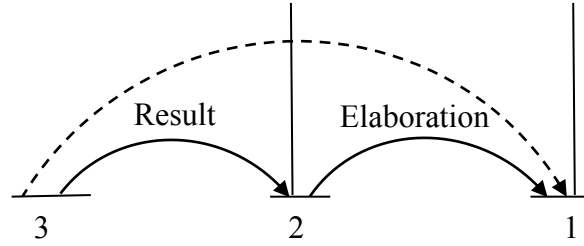


Figure 2: A rhetorical analysis of text (4).

Keywords repetition (Marcu, 2000a; Timmerman, 2007) can also be used as an indicator of the presence of a distance relation. The idea behind this technique relies on a facet of text cohesion that is adequate for determining sentences that have a single theme i.e. if two sentences deal with the same concept, it is likely that they involve the same elements of nouns. In this sense, we can say that a hypotactic relation relates those two sentences. However, it is tricky to accurately recognize which type of relation exists without world knowledge; hence the added relation would always be of an *Elaboration* relation where the sentence that comes later, *satellite*, elaborates on the topic of the sentence that went before, *nucleus*.

The matching process is carried out as follows: the words that are associated with noun tags are initially extracted from the texts that correspond to the sentences and then all affixes - grammatical units that is attached to the beginning or end of a word stem- of these nouns are removed using a word stemmer. Thereafter, each sentence is compared to the following sentences in turn. If similar nouns are found, a new relation is hypothesized to hold between the two sentences under consideration provided that neither sentence is rhetorically related to another one; this condition is neglected in the case where the sentence has the *nucleus* status.

Consider the four sentences in text (5), we notice that a rhetorical relation of *Result* is signalled between sentences (1) and (2) based on the occurrence of the marker “مما أدى” at the head of sentence (2). Since sentence (1) is the *nucleus* of this relation, it is matched with sentences (3) and (4) for possible mutual nouns. Finally two relations of *Elaboration* are added to the relations set because sentences share the nouns “ارض - نيزك” “earth - meteorite”. The graph in Figure 3 shows the discourse analysis of text (5).

[أكد بعض العلماء أن نيزكا كبيرا اصطدم بالأرض في حقبة الديناصورات منذ ملايين السنين.]¹ [مما أدى الى هلاك هذه الديناصورات (5) والأحياء الأخرى التي عاشت في تلك الفترة.]² [وتم التعرف على آثار النيزك من خلال طبقة الرواسب المتخلفة عن السحابة الغبارية التي غطت كوكب الأرض بعد الاصدام.]³ [إن دراسة اصدام النيزك بالكرة الأرضية يمكن أيضا أن تساعد العلماء على فهم الظروف التي نشأت فيها الحياة على هذا الكوكب بشكل أفضل.]⁴

[A team of researchers have confirmed that a large meteorite had collided with Earth at the age of dinosaurs millions of years ago.]¹ [This was responsible for the mass extinction of dinosaurs and all other species living on Earth.]² [The meteorite was identified from the layer of sediment deposited from the dust cloud that enveloped the Earth after the impact.]³ [Studying the meteorite’s impact with the Earth could also help scientists to better understand the conditions under which early life on the planet evolved.]⁴

The *Text Parser* incorporates the two methods mentioned above i.e. relations transitivity and keywords repetition in order to discover long distance relations. It operates as follows, every pair of adjacent sentences in the list of DUs is checked for possible relations on the basis of discourse marker occurrences; thereafter the list is examined again for long-distance relations between sentences that were not already hypothesized to be related to another DU as a satellite unit. The parser applies the heuristics introduced in Section 5.1 to add a scoring value for each hypothesized discourse relation.

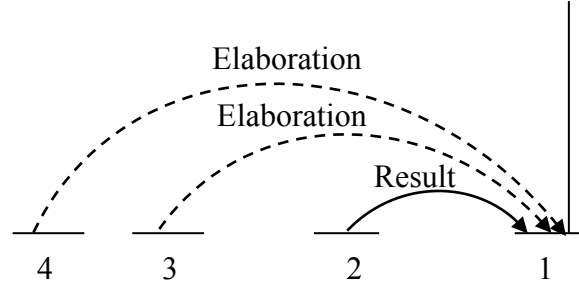


Figure 3: A rhetorical analysis of text (5).

All generated relations are stored in a sorted set according to their heuristic score, in case that more than one relation is found to connect the same two sentences, the relation with the highest heuristic score is retained and all the others relations are discarded. At this point all sentences are supposed to be connected as the text is presumed to be coherent. Algorithm 1 finds possible relations for a given text where $[l, r]$ are the left and right boundaries of a rhetorical relation, rr is a temporary set of all rhetorical relations that could be hypothesised by a discourse marker.

ALGORITHM 1: Hypothesizing rhetorical relations.

Input: A sequence $S[n]$ of sentences annotated with intrasentential relations.

Output: A set RR of relations that hold among sentences in $S[n]$.

1. $RR := \text{null}$;
2. Determine the list DM_s of all Discourse Markers occur at the head of each sentence in $S[n]$;
3. For each marker $M \in DM_s$
4. $rr := \text{null}$;
5. While there is a relation that *marker* can relate
6. $rr := rr \oplus rhet_rel(\text{name}_{(M)}, \text{score}_{(M)}, l_{(M)}, r_{(M)})$;
7. $RR := RR \cup \{rr\}$;
8. End For
9. For each pair (i, j) of adjacent sentences in $S[n]$
10. If more than one relation found in RR to hold between (i, j)
11. $rr := rr \oplus rhet_rel(\text{name}, \text{score}_{(max)}, i, j)$;
12. End For
13. $RR := RR \cap \{rr\}$;
14. For each pair (x, z) of sentences in $S[n]$
15. Use cohesion and transitivity to find distance relation rrd
16. If $\text{Score}_{(rrd)} > \text{threshold}$
17. $RR := RR \cup \{rrd\}$
18. End For
19. Sort RR from the highest scored hypothesis to the lowest scored

5. CONSTRUCTING THE RHETORICAL STRUCTURE TREE

5.1 Heuristic Scores

As the aim of the current study is to provide answers to “*why*” and “*how to*” questions, rhetorical relations which are more relevant for such type of questions should be highlighted. Thus, we composed a small subset subsuming the following relations: *Result*, *Reason*, and *Interpretation*. Our target then is to prioritize this relevant subset in order to ensure that its members are always on top

Annotated corpora ought to be available to automatically learn the optimal values for heuristic scores. Unfortunately, no corpus of Arabic RST-analyzed texts exists. Hand-tuning is therefore still necessary. The heuristic scores presented in this study were obtained by trial and modification. Initial values were assigned with the aim of ensuring that preferred relations occurred at the top of the sub-Trees list i.e. *Result*, *Reason* and *Interpretation* relations are extremely good indicators of “*why*” and “*how to*” questions, we can therefore assign a high initial value. The other relations’ and DMs’ values scores were initially based on intuition as the authors are native speaker of the Arabic Language. We carried out a regression test on a set of Arabic texts; these are full articles extracted from the contemporary Arabic corpus². The outcome of *Text Parser* is always checked to determine whether it produces a tree that spans over the whole text. Whenever the heuristic scores are modified for a new text, to achieve the tree-like structure, the new scores are tested to ensure that texts that were previously analyzed correctly in the regressions set are not affected.

It is important to emphasize that we did not embrace an exhaustive list of all the relations identified by Al Kohlani (2010). Rather, the relations employed in this study, comprised a set of ten relations that occur more often among sentences as indicated by the DMs frequency in (Al Kohlani, 2010). These relations are sufficient for reflecting the writer’s attitudes and viewpoints in discourse from the cohesion-based perspective. The other relations are hardly signalled in text. Table 1 shows the set of adopted rhetorical relations along with the corresponding maximum score. The maximum score represents the highest value that relations may be assigned. For example, when positing an *Elaboration* relation between two sentences, we add the value of 15 whenever a pair of nouns matches unless the relation reaches its max score, i.e. 80.

Table 1
Rhetorical relations scores

Relation	Max Score	Relation	Max Score
Result	100	Background	60
Reason	100	Evaluation	50
Interpretation	100	Certainty	50
Elaboration	80	Sequence	50
Contrast	70	View	50

Table 2
A list of DMs and corresponding heuristic score.

Marker	Rhetorical relation	Score
من هنا “thus, therefore”	Evaluation – Result	50 / 40
من اجل ذلك “because of that”	Result	100
إلا أن “however, but”	Contrast	70
ثم “then”	Sequence	50
Shared noun	Elaboration	+15
Intrasentential relations	Relevant subset	+45
Demonstratives	Elaboration	+60

² <http://www.comp.leeds.ac.uk/eric/latifa/research.htm>

We have taken each of the DMs identified by Al Kohlani (2010) and considered its potential contribution in hypothesizing the rhetorical relations. In the case where a DM correlated with only one particular relation, the relation was thus indicated with relatively high level of confidence and accordingly the DM was associated with a score equal to the maximum value of this relation. In case where a DM signalled different discourse relations, as was the case with “من هنا” “therefore/thus”, it is perceived as a weaker evidence and is associated with a lower score. Table 2 shows a set of scores that correspond to some of the discourse markers.

5.2 Building the Tree

Given a text segmented into DUs at the sentence level and a set of rhetorical relations that have been hypothesized to hold between those sentences, we now build the possible RST Trees.

In this study, we assumed full conformity to the principle of compositionality proposed by Marcu (2000a) in order to join two adjacent sub-trees: “*whenever two large text spans are connected through a rhetorical relation, that rhetorical relation holds also between the most important parts of the constituent spans*”. Accordingly, each rhetorical relation is associated with a promotion set that reflects the compositionality criterion. Promotions sets are the set of units that constitute the most important parts of the text that is spanned by the node. For a terminal node, the promotion set consists only of the terminal node itself. For an asymmetric sub-tree, the promotion set consists of a single element, the nucleus. For a symmetric sub-tree, the promotion set consists of the union of the promotion sets of the co-nuclei.

The *Text Parser* applies the posited discourse relations with high heuristic scores before those with lower scores in a bottom-up manner, grouping contiguous clauses into a hierarchical representation. Afterwards the parser establishes a list of sub-trees by gathering text spans produced in the previous step into contiguous new textual units. Sub-trees are being built up by iterating over all pairs in the relations set. The *Text Parser* starts by selecting the relations ranked highest according to their scores since they constitute the most promising path and then moves to the second pair in the relations set. Heuristic scores are being accumulated throughout by adding up all scores in the sub-trees produced so far. This step is repeated until the list of sub-trees contains only one tree including all sentences in the text. If no relations are found between two adjacent sub-trees, the sub-trees could be assembled with the *Joint* relation because the text is considered to be a connected structure that spans across all its units. Algorithm 2 produces a discourse tree spanning over the whole text. Each sub-tree takes the following form:

SubTree(L,R,Status,Type,Promotion,Score,left_SubTree,right_SubTree) where L,R are the left and right boundaries of a sub-tree.

ALGORITHM 2: Building up the valid tree structure.

Input: A text T of N sentences $S[N]$.

A sorted list RR of relations that hold among the sentences in $S[N]$.

Output: The RS-tree of T .

1. $SubTreesList := \text{Null}$;
2. For $i = 1$ to N
3. Convert sentence into the form: $SubTree(i, i, NONE, LEAF, [S_i], 0, NULL, NULL)$;
4. $SubTreesList := SubTreesList \cup SubTree$;
5. End For
6. While RR contains at least one element and $SubTreesList$ has more than one element
7. For each $rr \in RR$

8. Search in the *SubTreesList* for elements with the promotions specified by *rr*;
9. If match not found or combining the two *subTrees* would result in crossing lines
10. Remove *rr*;
11. Else create new *subTree* by joining the two *subTrees* as specified by *rr* and add the heuristic score accordingly;
12. Update *SubTreesList* and *RR* accordingly;
13. End While
14. If *SubTreesList* has more than one element
15. Join all elements in *SubTreesList* into one tree that spans the whole text;

6. WORKED EXAMPLE

The operation of the text derivation proposed in this paper is illustrated by the example below. The text is segmented into five discourse units (DUs) each with the length of a full sentence. The *Pattern Recognizer* discovers relations within sentences. This process yields two intrasentential relations: Cause-Effect and Method-Effect from sentences (B) and (C) respectively; each relation involves two slots to be filled by the first and second part of the relation (Sadek, 2013).

(8)

[تتعرض عظام المرأة والرجل بعد سن الخمسين إلى التنخر الداخلي وتناقص مستوى الهرمون المساعد على بنائها.]^A [بالتالي <تكون العظام ضعيفة أمام الضربات والحوادث>^{cause} مما يؤدي إلى <حدوث الكسور التي تتطلب أحيانا الاستبدال العظمي.>^B [والأدوية الموجودة حاليا تعمل على تقليل نسبة حدوث الكسور في العظام<effect> من خلال <المحافظة على نسبة الكالسيوم الطبيعية في الدم.>^C [لكن الدواء الجديد ويدعى فورتيو ، فهو على العكس يعمل على زيادة سرعة عملية البناء والتي تقوم بها الخلايا البانية للعظم.]^D [وفي التجارب التي أجريت على هذا الدواء ظهر أن احتمالات إصابة العمود الفقري بالكسور قلت بنسبة 65-90% بينما قلت احتمالات إصابة الكاحل والمعصم و الورك والأضلاع والقدم بالكسور بنسبة 54%.]^E

[After the age of fifty, bones of men and women are exposed to internal necrosis and reduction in the level of the hormone that helps building bone structure.]^A [As a result, <they become weak in cases involving blows and accidents>^{cause} which causes <fractures that sometimes require bone replacement.>^{effect}] ^B [*Nowadays, the existing drugs reduce the possibilities of bones fracture injuries*>^{effect} by <maintaining the ratio of natural calcium in blood.>^{method}] ^C [On the contrary, the new drug, Forteo, accelerates the speed of the construction process carried out by bone-constructing cells.]^D [The experiments conducted on this drug revealed that the possibilities of spine fracture injuries decreased by 65% to 90%, whilst possibilities of ankle, wrist, hip, ribs and foot injuries are reduced by 54%.]^E

Given sentences tagged with interasential relations, the *Text Parser* then starts to identify rhetorical relations between these sentences. The *Relation Recognizer* first examines all pairs of the adjacent sentences and produces the hypothesized discourse relations given in Figure 5.

$$\left\{ \begin{array}{l} rhet_rel (Result, 85, A, B) \\ rhet_rel (Evaluation, 50, A, B) \\ rhet_rel (Contrast, 70, C, D) \\ rhet_rel (Elaboration, 60, D, E) \end{array} \right.$$

Figure 5: Adjacent relations for text (8).

We notice that two relations, *Result* and *Evaluation*, are posited between sentences (A) and (B) based on the occurrence of the DM "بالتالي" at the head of sentence (B). The score of the *Result* relation is calculated by adding 45 points to the base value 40 because sentence (B) is tagged with a *Causal* relation. The relation with the higher likelihood between sentences (A) and (B) is kept and the other

one is discarded i.e. the *Evaluation* relation. Also, an *Elaboration* relation is hypothesized between sentences (D) and (E) based on the occurrence of the demonstrative pronoun “هذا” in the first third of sentence (E).

The *Relation Recognizer* proceeds with discovering long distance relations. It compares nouns in each possible pair of sentences and assigns a likelihood based on the number of similar nouns. The *Relation Recognizer* only adds an *Elaboration* relation if it receives a score above the minimum score. For example, only the noun “عظام” is shared between sentences (A) and (C), thus such relation is not added to the relations list. Also, sentences (D) and (E) contain the noun “الدواء” which indicates the presences of an *Elaboration* relation with a likelihood of 15. However, since an *Elaboration* relation has been hypothesized between the same sentences in the previous step this value is added up to the total score. At this stage all sentences are connected and the final relation set is shown in Figure 6.

$$\left\{ \begin{array}{l} rhet_rel (Result, 85, A, B) \\ rhet_rel (Contrast, 70, C, D) \\ rhet_rel (Elaboration, 75, D, E) \end{array} \right.$$

Figure 6: Relations set for text (8).

Next, the *Tree Builder* parses the relations list generated by the *Relation Recognizer*. It initially converts all sentences into terminal nodes represented as sub-trees each having a single member in its promotion set - the sentence itself. The *Tree Builder* then attempts to apply all the rhetorical relations starting with the one which has the highest score. Figure 7 illustrates the sub-trees list content resulting from the application of the first and third hypothesis in the relations set, sentences written in curly braces specify the promotion set of each sub-tree. The *Tree Builder* moves on to consider the *Contrast* relation, it searches the sub-trees list for a sub-tree whose promotion set includes sentence (C) and a sub-tree whose promotion set includes sentence (D). It finds the terminal node (C) and the sub-tree [D-E], it thus combines them to form a new sub-tree covering sentences (C) through (E) as shown in Figure 8. The *Tree Builder* is unable to find a relation that connects sub-tree [A-B] and [C-E], and therefore a *Joint* relation is applied to combine the two sub-trees. Figure 9 depicts the Tree that covers the entire input text.

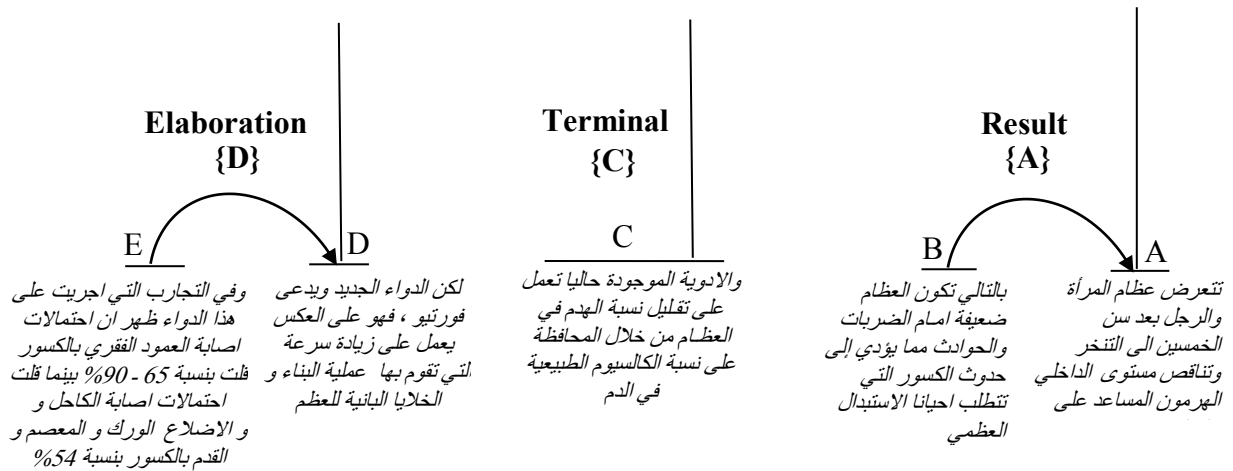


Figure 7: Sub-trees list after applying the *Result* and *Elaboration* relations.

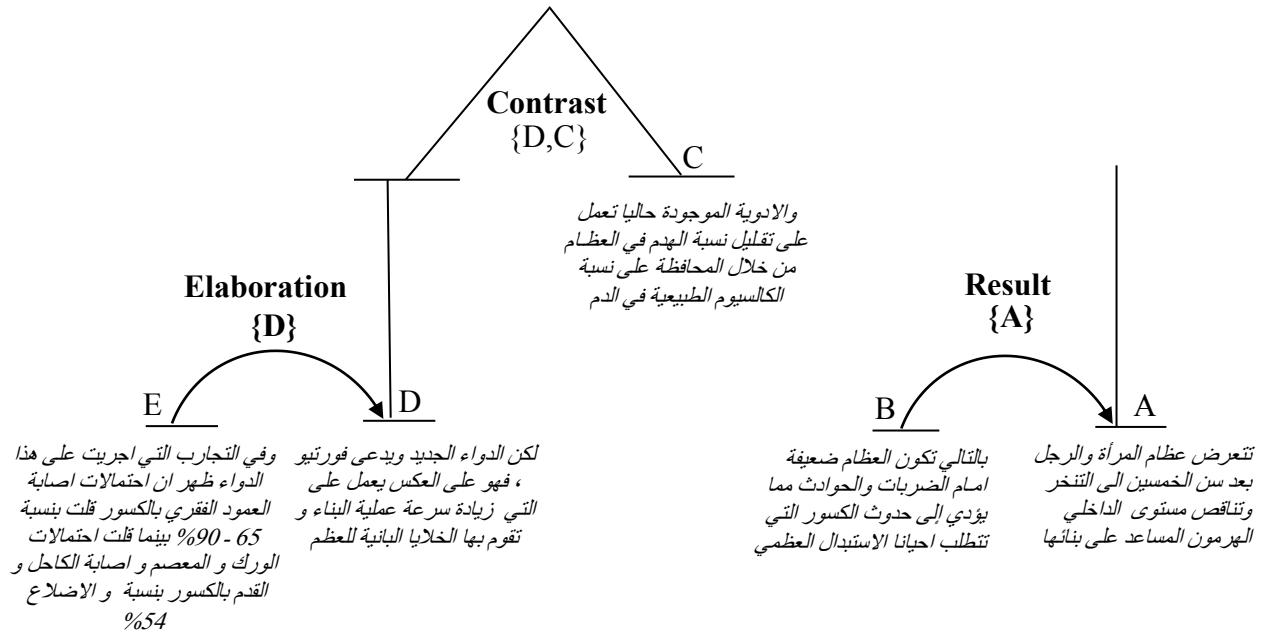


Figure 8: Sub-tree after applying the Contrast relation.

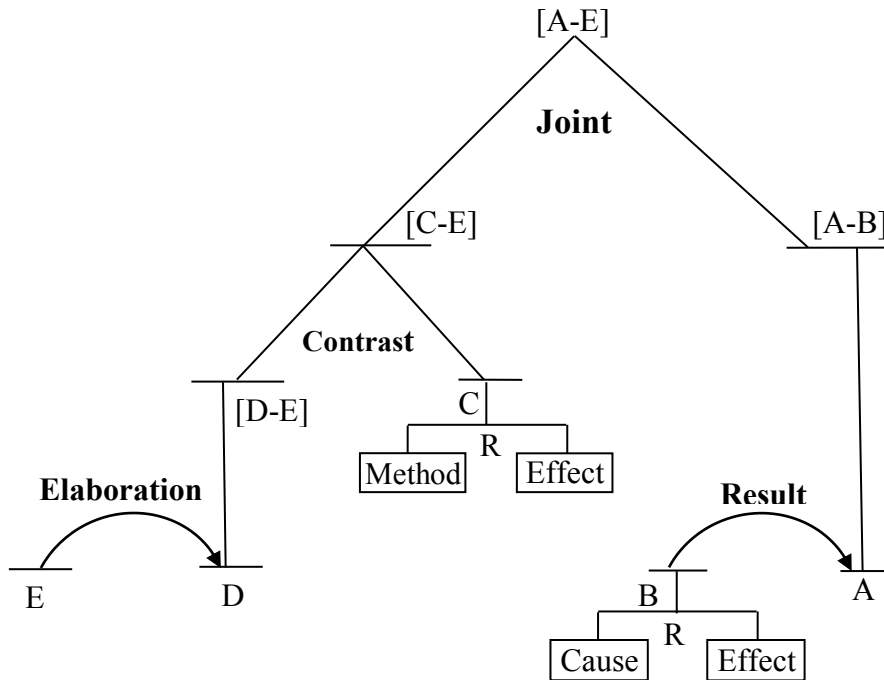


Figure 9: The generated Tree of text (8).

7. EVALUATION

Our QA system was implemented using the Java programming language. It first applies the set of linguistic patterns constructed by Sadek (2013) to extract relations within sentences. The *Text Parser* then discovers rhetorical relations among sentences. Next, all possible relations together with the asked question are tokenized. This includes performing normalization, stemming and stop words removal. The system applies the Vector Space Model to compute the similarity between the question and the appropriate part of each of the relations as discussed in Section 2. The corresponding part of the relation is then returned as a candidate answer. Finally, answers are ranked according to the similarity value.

We did an experiment similar to the one conducted by Verberne (2007) to measure mean reciprocal rank (*MRR*) and *Recall*. *MRR* is calculated as follows: for each question, the reciprocal rank (RR) is equal to 1 divided by the rank of the highest ranked correct answer or 0 if none of the responses contained a correct answer. *MRR* is then the average of RR over all questions as shown in Formula below where rank_i is the rank position of the first correct answer for the i th question.

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank } i}$$

We have used as a baseline our previous QA system (Sadek et.al. 2012) to compare the performance of the system developed in this study. The baseline was designed to handle short texts for the Arabic language. It also employs RST to extract text structures; however, relation recognizer and EDUs determinator are solely based on cue phrases.

We selected texts of 870-2138 words each. The texts were extracted from the contemporary Arabic corpus belonging to the Health and Science & Technology categories. Five subjects were involved in this evaluation and all are native speakers of Arabic. We asked them to read some of the texts and formulate “*why*” and “*how to*” questions for the answers that could be found in the text, the subjects were also asked to pick sentences out of texts that would formulate answers to each of their questions. As a result we collected a total of 90 question-answer pairs.

We ran our system on the collected questions, and then compared the answers found by the system to the user-formulated ones; if the system’s answer matches the answer formulated by the subjects then we judged the answer as correct. The system correctly identified the answers for 61 questions (67.7% of all questions) with *MRR* of 0.62. The results are shown in Table 3 and Table 4 indicating that the current system yields significantly better performance over the baseline for both types of questions.

Considering the failed questions, we distinguish two categories. First, questions for which there are no explicit relations between the textual units representing the question and the textual unit of the answers. This category comprises five questions (18% of the questions did answer correctly). The questions belonging to this category are connected to the answers spans with relations expressed implicitly in the text and are inferred by the reader using general knowledge.

For example question (9) refers to sentence (10) in the source text. This question corresponds to the string: “حصل على منحة بمقدار مليون دولار من قطاع صناعة القطن في كاليفورنيا” “*got a grant worth one million dollars from the Cotton industry sector in California*” which is embedded in the subject-formulated answer: “لكي يحاول تعديل فراشة القطن الزهرية وراثيا” “*to genetically modify the pink cotton butterfly*”. Although the Causal relation is not explicitly indicated in sentence (10), the reader has no difficulty inferring that Miller has been granted million dollars for the purpose of conducting his research.

(9) لماذا حصل ميللر على منحة قدرها مليون دولار ؟
Why did Miller get a grant worth of one million dollars?

ويحاول ميللر، الذي حصل على منحة بمقدار مليون دولار من قطاع صناعة القطن في كاليفورنيا، تعديل فراشة القطن الزهرية وراثيا (10) لكي تكون نشطة جنسيا ولكن غير قادرة على التناسل بالطريقة المناسبة
Miller, who's got a grant worth one million dollars from the Cotton industry sector in California, genetically modifies the pink cotton butterfly to be sexually active but unable to reproduce in a proper way.

The other category (24 questions, 82%), consists of cases where particular kinds of linking words are supported by neither the linguistic patterns nor the *Text Parser*. Some of these items are seldom used for indicating such relation. For example, question (11) refers to the *Causal* relation in sentence (12) which is indicated by the expression “ما لم ندرك” “unless we become aware that”.

Figure 10 illustrates the distribution of the questions answered correctly (the first two columns) together with the failed questions (the second two columns). Nearly 55% of the questions were answered correctly based on the indication of intrasentential relations, whereas correct answer for 13% of the questions correlate to the presence of rhetorical relations between sentences.

(11) لماذا تتسم عمليات التجميل بمحدودية الفائدة ؟
Why cosmetic surgeries are of limited value?

...ولكن السر الذي يخفى على جميع السيدات وخاصة ذوات البشرة الزيتونية هي أن جميع هذه المستحضرات والعلاجات ذات فائدة (12) محدودة ما لم ندرك دور أشعة الشمس الخليجية في افشال الكثير من هذه العلاجات.
“But the secret, which is hidden from most women especially those with olive complexion, is that all of these products and treatments are of limited effect unless we become aware that the Gulf sunlight spoils most of such treatments”

Table 3
 Results for the *Recall*

	Baseline	Current System
Why	38%	70%
How to	45%	65%
Overall	40%	67.7%

Table 4
 Results for MRR

	Baseline	Current System
Why	0.37	0.61
How to	0.45	0.65
Overall	0.39	0.62

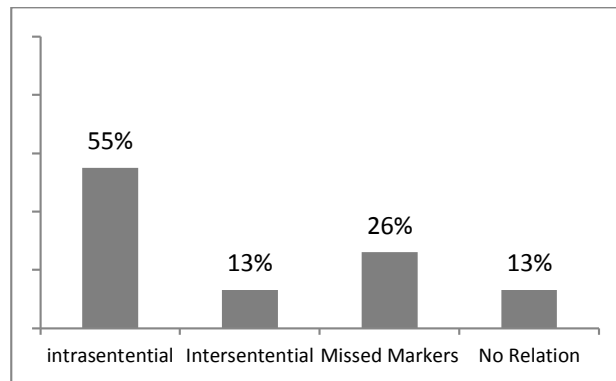


Figure 10: the distribution of the questions test.

8. RELATED WORK

A number of studies tackling the problem of automatic discourse parsing have been performed in recent years. A fair number of the developed parsers were eventually applied to text summarization. Mathkour (2008) presented an early attempt at automatic derivation of Arabic discourse structure using the methodology introduced by Marcu (2000b). However, they restricted the scope of their study to small size articles stating that “*for large texts the result might take hours to be generated*”. In another study conducted by Ibrahim and Elghazaly (2012), the authors gathered a number of Arabic rhetorical relations. They first verified the collected relations using an Arabic corpus tagged with articles structure; in the second phase they identified the most significant paragraphs. their method was applied for text summarization and achieved F-measure of 29%. Marcu (2000b) proposed a shallow, surface-based approach to decompose a free unrestricted text into EDUs and hypothesizes rhetorical relations that hold among textual units based on the appearance of cue phrases and then, produces all binary rhetorical trees compatible with the hypothesized relations. Soricut and Marcu (2003) developed their automatic sentence-level parsing of discourse (SPADE) system based on a Treebank annotated with discourse structure called RST-DS. SPADE uses two probabilistic models to accomplish the task of sentence segmentation into non-overlapping discourse units along with linking these units with labelled hierarchical structures. However, their discourse parser was restricted to building sub-trees spanning over individual sentences. Theijssen (2007) employed machine learning techniques in order to find relations between multi sentential discourse units (MSDU) within the same paragraph. She adopted five different learning algorithms with the aim of automatically extracting values for each of the potential relevant features that can lead to detecting whether a text span is rhetorically related to the preceding or the following MSDU. Theijssen (2007) pointed out that the performance of the classification algorithm was disappointing due to the small data set and large number of features. In their discourse parser, Feng and Hirst (2012) used two classifiers for discourse tree building. A binary structure classifier to decide whether two consecutive text units should be merged to form a new subtree, and a multi-class classifier to evaluate which discourse relations are the most likely to hold between the new subtree. They measured the performance of their parser under three discourse conditions: *Within-sentence*, *Cross-sentence* and *All level*. Their experimental results showed that the parser was relatively poorer on cross-sentences than that on within-sentence which, the authors stated, indicates “*the difficulty of text-level discourse parsing*”.

Most attention in QA community was paid to factoid questions fostered by TREC annual conferences; few studies were dedicated to dealing with “*why*” and “*how to*”. Verberne (2007) intensively worked on finding answers to “*why*” questions for the English language. In (Verberne et al., 2007) authors approached the answer extraction problem as discourse analysis task by employing

RST Discourse Treebank. This Treebank was manually annotated with discourse relations. Verberne (2007) shifted the “*why*” QA task towards paragraph retrieval rather than a textual span. The last (third) version of the JAVELIN system that was originally implemented for factoid English language has been extended to accept non-factoid question including “*why*” and “*how*” questions for the Japanese language (Shima and Mitamura, 2007). In its third edition the system used an annotated database with various information such as morpheme text chunks, POS and named entities along with predicate-argument analysis. The adoption of machine learning technique was incorporated with hand crafted cue words that may identify the type of relation sentences. The results obtained from the system showed that the performance was less efficient than the versions created for factoid questions. One reason for that is the small number of the examples available for the training phase. Another system that made use of machine learning is presented by Higashinaka and Isozaki (2008) with the aim of ranking a given set of candidate answers for Japanese why-questions. The study based on the assumption that answers are of a one sentence or paragraph long and to be extracted from top-N documents returned by a document retrieval module. The features (causal expressions, causal relation and content similarity) were mainly based on causal expressions extracted from semantically tagged corpora. The answer candidate ranker obtained *MRR* of 0.305 for top-5. Surdeanu et al. (2008) took advantage of the abundant content provided by Yahoo! Answers for developing an answer ranking engine for “*how to*” questions. The authors selected as a baseline the output of the answer retrieval model with BM25 similarity formula (ranking function based on binary independence model); their system achieved a 14% improvement in *MRR* over their baseline. Akour et al. (2011) introduced the QArabPro system for the Arabic language based on a set of separate rules for each type of question. The authors used the same method to handle all question types including “*why*” questions. However, many studies demonstrated that knowledge about discourse relations is crucial to answer this type of question. For example, in their work they marked the word “حيث” as stop word that has to be omitted out of query/document processing while it is used in contemporary Arabic language to indicate *Causal* relations. Moreover authors stated that they handled the question type “كيف” “*how to*”. Whilst they actually treated the type (how much/many) “كم”.

9. CONCLUSION AND FUTURE WORK

In the current research, we aimed at improving the performance of Arabic QA system. Obviously, considering relations spanning over only individual sentences one at a time is more computationally efficient than considering the whole text. Furthermore, linguistic patterns incorporated in the developed parser had a fundamental role in discovering causation and explanation within sentence; in most cases such information cannot be captured using the rhetorical relations of RST when handling small text fragments.

The experiment in this study focused on the evaluation of the QA system. As an extension to this work, we plan to evaluate trees constructed by our *Text Parser* with the participation of trained judges. An automated learning algorithm could also test different scores for each relation and DM in order to determine whether a better set of scores exists than the one currently in use.

The test data were collected through elicitation i.e. subjects had access to the text, which implies that the formulated questions might have influenced by the same linguistic cues used by the text producer. Future work should investigate query expansion techniques since users dealing with real QA systems will not have access to the documents.

REFERENCES

- Akour, Mohammed, Abufardeh, Sameer, Magel, Kenneth and Al-Radaideh, Qasemm. 2011. QArabPro: A Rule Based Question Answering System for Reading Comprehension Tests in Arabic, *American Journal of Applied Science* Vol.8 (6), PP:652-661.
- Al Kohlani, Fatima. 2010. The Function of Discourse Markers in Arabic Newspaper Opinion Articles, PhD thesis, Georgetown University, Washington.
- Bernardi, Raffaella, Valentin, Jijkoun, Gilad, Mishne and Maarten, De Rijke. 2003. Selectively Using Linguistic Resources throughout the Question Qnswering Pipeline . In *Proceedings of the 2nd CoLogNET*, PP:50-60.
- Blakemore, Diane. 2003. Discourse and Relevance Theory, *The Handbook of Discourse analysis*, Oxford: Blackwell, PP: 100-115.
- Breck, Eric, Burger, John, Ferro, Lisa, Greiff, Warren, Light, Mani and Rennie, Jason. 2000. Another Sys Called Qanda, In *Proceedings of the Ninth Text REtrieval conference, NIST Special Publication 500-246*, Maryland, PP: 369-379.
- Corston-Oliver, Simon. 1998. Computing Representations of the Structure of Written Discourse, PhD thesis, University of California, Santa Barbara.
- Fareh, Shehdeh and Hamdan, Jihad. 1999. The Translation of Arabic ‘Wa’ into English: Some Problems and Implications, *Dirasat: Human and Social Science*, Vol.26, PP:590-603.
- Feng, Vanessa and Hirst, Graeme. 2012. Text-level Discourse Parsing with Rich Linguistic Features. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-2012)*, Korea.
- Fraser, Bruce. 1996. Pragmatic markers, *Pragmatics*, Vol. 6 (2), PP: 167-190.
- Higashinaka, Ryuichiro and Isozaki, Hideki. 2008. Corpus-based Question Answering for why-Question, In *Proceedings of the 3rd International Joint Conference on Natural Language Processing*, India, Vol 1, PP: 418-425.
- Ibrahim, Ahmed, and Elghazaly, Tarek. 2012. Arabic text summarization using Rhetorical Structure Theory. In 8th International Conference on Informatics and Systems (INFOS), IEEE, PP:34-38.
- Kupice, Julian. 1999. MURAX: Finding and Organizing Answers form Text Search”. *Natural Language Information Retrieval*, Netherlands, PP: 311-332.
- Mann, William and Thompson, Sandra. 1988. A Rhetorical Structure Theory: Toward a Functional Theory of Text Organization, *Text-Interdisciplinary Journal for the study of Discourse*, Vol. 8 (3), PP: 243-281.
- Marcu, Daniel. 2000a. The Rhetorical Parsing of Unrestricted Texts: a surface-based approach, *Computational Linguistics*, Vol.26 (3), PP: 395-448.
- Marcu, Daniel. 2000b. The Theory and Practice of Discourse Parsing and Summarization, MIT Press London, England.
- Mathkour, Hassan, Touir, Ameer and Al-Sanea, Waleed. 2008. Parsing Arabic Texts Using Rhetorical Structure Theory. *Journal of Computer Science*, Vol.4 (9), PP:713-720.
- Sadek, Jawad and Meziane, Farid. 2016. Extracting Arabic Causal Relations Using Linguistic Patterns. *Journal ACM Translations on Asian and Low-Resource Language Information Processing*, Vol. 15 (3), Paper 14
- Sadek, Jawad. 2013. Automatic Detection of Arabic Causal Relations, In *Proceedings of the 18th International Conference on Application of Natural Language to Information Systems*, (NLDB), UK, PP: 400-403.
- Sadek, Jawad, Chakkour Fairouz and Meziane, Farid. 2012. Arabic Rhetorical Relations Extraction for Answering “Why” and “How to” Questions, In *Proceedings of the 17th International Conference on Application of Natural Langugae to Information Systems, NLDB*, The Netherlands, PP: 385-390.

- Sanders, Ted and Noordman, Leo. 2000. The role of Coherence Relations and Their Linguistic Markers in Text Processing. *Discourse processes*, Vol. 29 (1), PP: 37-60.
- Schiffrin, Deborah, Tannen, Deborah and Hamilton, Heidi. 2001. Discourse Markers: Language, Meaning, and Context, *Basil Blackwell*, Oxford.
- Schneuwly, Bernard. 1997. Textual Organizers and Text Types: Ontogenetic aspects in writing, *Processing Interclausal Relationships: Studies in the production and comprehension of text*, PP: 245-263.
- Segal, Erwin, Duchan, Judith and Scott, Paula. 1991. The Role of Interclausal Connectives in Narrative Structuring: Evidence from adults' interpretation of simple stories. *Discourse processes*, Vol.14, PP: 27-54.
- Shima, Hideki and Mitamura, Teruko. 2007. JAVELIN III: Answering non-factoid Question in Japanese, In *Proceedings of NTCIR-6 Workshop Meeting*, Tokyo, PP: 464-468.
- Soricut, Radu and Marcu, Daniel. 2003. Sentence Level Discourse Parsing Using Syntactic and Lexical Information. In *Proceedings of the Human Language Technology and North American Association for Computational Linguistics*, Canda, PP: 149-156.
- Surdeanu, Mihai, Giaramita, Massimiliano and Zaragoza, Hugo. 2008. Learning to Rank Answers on Large Online QA Collections, In *Proceedings of ACL-08*. USA, PP: 719-727.
- Theijssen, Daphne. 2007. Feature for Automatic Discourse Analysis of Paragraphs, Master's thesis, Radboud Universiteit Nijmegen, The Netherlands.
- Timmerman, Sander. 2007. Automatic Recognition of Structural Relations in Dutch Text, Master Thesis, University of Twente, The Netherlands.
- Verberne, Suzan. 2007. Paragraph Retrieval for Why-question Answering. In *Proceedings of the 30th Annual International ACM SIGR Conference on Research and Development in Information Retrieval* 922-927, New York, PP: 922-927.
- Verberne, Suzan, Boves, Lou and Osstdijk, Nelleke. 2007. Discourse-based Answering of why-questions. *Traitement Automatique des langues, Special issue on computational Approaches to Discourse and Document Processing* Vol.47(2), PP: 21-41.
- Wright, William and Caspari, Paul. 1896. A Grammar of the Arabic Language, Cambridge University Press. UK.
- Zaki, Mai. 2011. The Semantics and Pragmatics of Demonstratives in English and Arabic, PhD thesis, University of Middlesex.