Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2018) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

The 4th International Conference on Arabic Computational Linguistics (ACLing 2018), November 17-19 2018, Dubai, United Arab Emirates

Learning Causality for Arabic - Proclitics

Jawad Sadek[a], Farid Meziane[b]

[a] *National Institute for Health Research Innovation Observatory – Newcastle University, Newcastle NE14BF, UK*
[b] *School of Computing Science and Engineering – Salford University, Salford M54WT , UK*

Abstract

The use of prefixed particles is a prevalent linguistic form to express causation in Arabic Language. However, such particles are complicated and highly ambiguous as they imply different meanings according to their position in the text. This ambiguity emphasizes the high demand for a large-scale annotated corpus that contains instances of these particles. In this paper, we present the process of building our corpus, which includes a collection of annotated sentences each containing an instance of a candidate causal particle. We use the corpus to construct and optimize predictive models for the task of causation recognition. The performance of the best models is significantly better than the baselines. Arabic is a less–resourced language and we hope this work would help in building better Information Extraction systems.

Keywords: Causal Relations; Arabic Causality Extraction; Discourse Relation Recognition; Arabic Annotated Corpus

1. Introduction

Causal relations occur between an event (the cause) and a second event (the effect) in which the second event is understood to be the consequence of the first. In recent years, causal relations have become increasingly important for applications related to Natural Language Processing (NLP) such as Machine Translation, Text Generation and why-QA systems [1][2]. Arabic causality can be expressed using different expressions and linguistic elements. A significant one is a set of particles that are always attached to words and playing a key role in indicating causation. We refer to this group as proclitics and include: Purpose Lām (لام التعليل) Causation Fa'a (فاء السببية) and Causation Ba'a (باء السببية). Authors use proclitics substantially to express causation in Modern Standard Arabic (MSA). Proclitics are one of the most complicated and ambiguous particles in Arabic language, as they have multifunctional roles and many semantic properties; some grammarians counted more than 30 different purposes for them.

In a previous work, we hand crafted a set of lexico-syntactic patterns to detect and extract cause-effect relations from Arabic texts [3]. We also proposed three separate rule-based algorithms in order to discover the causal role indicated by proclitics. Our experimental results revealed that adopting these algorithms had boosted the efficiency by a large margin, improving the overall recall measure for Health and Science texts by 29%. However, this improvement came at the cost of precision which declined by 16%. In fact, 67% of relations returned by the algorithms were misclassified. This reduction in precision highlights the ambiguity associated with proclitics.

There are a number of annotated corpora available for Arabic language, however, these resources are either 'low-level' (e.g. syntactical or morphological) annotated or they have been labelled with Causal relations while annotating other semantic relations. Causation is a complex phenomenon and needs to have annotators to be trained and focus in particular on Causal relations. On the other hand, the syntactic patterns of the Arabic Causal relations are rather complex and no general annotated corpus can provide the diversity of Causal relations. Hence, we cannot build on top of any pre-annotated corpus but have to create a dedicated one for this type of relations.

In the current work we review the first stage towards building the Salford Arabic Causal Bank (SACB) [4]. This stage has been conducted with the goal of collecting and annotating independent sentences where instances of proclitics occurred without regard for other causal indicators. As noted above, these proclitics are ambiguous and there is a need to collect enough data to train machine learning models so they can reasonably capture the relationships that exist between input and output features. In the second part, we employ the developed corpus to build supervised classifiers that assign words starting with Lām, Fa'a or Ba'a, (henceforth, target word) into 'casual' or 'non-causal' classes.

2. Building the Corpus

We used GATE framework [5] throughout all phases of creating our corpus. Two annotators were engaged in the annotation process. First annotator (annotator A) is a graduate student in the faculty of Arabic literature. The second (annotator B) is a teaching assistant who has been educated in Arabic.

2.1. The Data

We extracted our instances from an untagged MSA corpus called *arabiCorpus*[*]. The corpus can be freely searched and downloaded. It has a large collection of resources classified into different categories. We selected the Newspapers group since it covers a wide range of topics. Searching arabiCorpus for instances containing target word yields several millions of text fragments. However, proclitics's coverage in Arabic texts are highly skewed e.g. most occurrences of Fa'a do not indicate causation. For highly skewed data, most classifiers would be biased toward the major class to obtain overall accuracy. Moreover, the natural distribution is often not the best distribution for learning a classifier [6][7]. There are different strategies by which classifiers can combat biased dataset. However, deferring handling of class imbalanced issue to the model design stage implies the annotation of a large number of instances. This makes the building of the corpus a very costly process.

We aim to create a reasonably confident corpus in which instances are independent and well-represented. To this end we did a multistage sampling, by first splitting the data into multiple groups based on the length of target words. Each group is then divided into sub-groups that share a common syntactical feature. Finally, we performed judgment sampling to force the harvested instances being represented between causal and non-causal labels. A native speaker was required to read all sub groups and to randomly pick out an equivalent number of instances that may/may not express causation. All collected sentences were passed to an NLP pipeline of the following components: tokenization, sentence-splitting and POS tagging. The last process was implemented using the Arabic model of Stanford POS tagger.

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

2.2. Annotation Guidelines

There's a lot of uncertainty surrounding the decision about when two events are causally linked [8][9]. In this work, we build our corpus with the aim of supporting various NLP applications, thus we adopt a broad definition embracing the following principles: would event B (effect) have occurred if event A (cause) had not occurred? If A is a sufficient though not a necessary condition for B to occur, we conclude that A caused B. We didn't limit cause or effect to certain types of entities. An effect can be a fact, an event, a method, and a cause can refer to a motivation, reason, human action, technological causation etc.

We requested the annotators to make a reliable interpretation of the writer's intention and then decide whether the target word expresses causation bearing in mind that both cause and effect constitutes an independent clause (i.e. they don't overlap). And the effect part has to be explicitly the result of the cause. For example, we label the proclitic Fa'a in Sentence (1) as non-causal since that piece of information which made the writer reaches his conclusion is not specified.

(1) لقد قرأت ذات يوم كتابا يقول كيف تصبح مليونيرا فلما انتهيت منه ادركت انني لن اصبح مليونيرا.

"I once read a book titled How to Become a Millionaire and when I finished it, I realized that I would never become a millionaire."

After the target word's function was identified, the annotators override all POS tags within a window of five words surrounding the target word with new fine-grained ones i.e. to assign different POS tags on sub-word level. To accomplish this, we expanded the Stanford POS tag-set. For example, we added TIM (ظرف زمان) "adverb of time" – LOC (ظرف مكان) "adverb of place" - PRPY (ضمير متصل) "inseparable pronoun". The annotators were also requested to assign a new tag referring to the "morphological pattern" "الوزن الصرفي" of the target word. Most arabic words are derived by applying a set of morphological patterns to their consonantal roots. These patterns are abstractions which can be considered as an indicator of the common concept of the word such as tool an event place/time and instrument. This classification is a valuable feature when recognizing the role of certain proclitic as it will be discussed later in Section 3.1.

2.3. Adjudication

As previously stated, the topic of causation is a matter of debate among experts belonging to this field. It is inevitable that the annotators disagree about the classification of some instances. We reconcile the disagreement between annotators by accepting the instances where both annotators agree on whether a proclitic indicates causation; subsequently we eliminated approximately 300 instances. Table 1 presents the observed agreement between Annotators. Table 2 presents the main aspects of the final dataset: number of instances (N), number of annotated text elements (Tokens), number of instances labelled with causal status (causal), number of instances labelled with non-causal status (¬causal). We automatically corrected all minor mistakes made by annotators (e.g. annotation words' length) using a Groovy script. The output of GATE annotation tool is document files formatted according to the GATE XML style. We converted the documents using another Groovy script so that all annotated instances are encoded in a lightweight XML file. We plan to make the corpus an open access repository for the research community.

Table 1. Inter-annotator agreement.

Proclitic	Observed Agreement		
	POS	Template	Status
Lām	0.91	0.80	0.88
Fa'a	0.89	0.78	0.83
Ba'a	0.90	0.81	0.90

Table 2. The distribution of all annotated instances.

Proclitic	N	Token	causal	¬causal
Lām	984	31564	439	545
Fa'a	577	20097	247	330
Ba'a	601	17912	290	311
Total	2162	69573	976	1186

3. Predicting Causal Relations

In this section we describe the process of constructing predictive models for the problem of building binary classifiers on whether a target word indicates a causation relationship. We employ the label information learned from the annotated corpus we have created in this work. We used Scikit-Learn module in python [10] throughout the task of causation recognition.

3.1. Features

Many studies showed that lexical and syntactic information is very useful for the detection of semantic relations task [11][12][13][14]. We selected a list of shallow linguistic features based on the observations we made in [3]. The list came up while building the rule-based classifiers which allowed us to have adequate understanding of the characteristics of the proclitics.

Morphological patterns: As described in Section 2.2, morphological patterns can be indicators of the meaning of words. For instance, applying the pattern (**مفعال**) to the root (**ف ت ح**) "open" we get the (**اسم الالة**) *noun of Instrument* **مفتاح** (key). In this context, a proclitic can be classified as non-causal if the target word belongs to a set of nominal patterns e.g. **اسم الفاعل** '*present participle'*, **جمع التكسير** '*irregular plural'* or **اسم مكان** '*noun of place'*. Sentence (2) illustrates this fact where the target word "**لمعمل**" factory is a *noun of place*. Among these patterns, one particular category which is "**المصدر**" Al-Masdar has been employed by previous studies as a critical indicator for discourse relations [15] [16].

(2) سمحت الوزارة لمعمل الالبان باستئناف نشاطه.

"The Ministry allowed the dairy factory to resume its operations."

Negative words: is binary feature that indicates the presence of negation words in the argument preceding the target word. This feature can help to classify an instance as non-causal if its value is set to 'Yes'. To detect negation, we manually built a lexicon of Arabic negation words derived from **الادوات الجازمة** 'jussive tools'- **الادوات الناصبة** 'subjunctive tools' such as **لا** 'no' and **لن** 'won't'. Note that the negative polarity could be shadowed if certain function words occurred in the same argument. For example, the exception noun **سوى** 'but' in Sentence (3) revokes the negation status indicated by '**لا**' 'no', ergo, the sentence can be labelled with causation.

(3) لا حل للقضية الكردية سوى بالتفاوض مع بغداد التي خسرت الحرب.

"There is no solution to the Kurdish question but by negotiation with Baghdad which lost the war".

Part Of Speech: We included all POS tags occurring in a window of five words surrounding the target word. All POS are fine-grained gold standard tags obtained from human annotators as described in Section 2.2. We also included POS *sequences of three* inside the target word boundaries. We gain more information about morphological patterns by regarding POS trigrams. For example, if the first token of the word is tagged as VBD (verb, past tense) and followed by PRPY tag, the target word is mostly classified as non-causal e.g. **فحصه** "examined him".

Token length: We added a separate feature of integer value to capture the number of letters of the target word, one word before and the word after. This feature is very useful when classifying words of few-letters. In the case of the words prefixed with *Lām* for example, the majority of instances with three-letters and less are assigned the non-causal class e.g. فوز "victory" – لمس "touch".

Word string: We used bag of word representation of the words comprising the whole instance. This representation captures enough information about lexical cues that may associate with certain class. For example, the occurrence of word اما "as for" or بالنسبة "regarding" before the proclitic *Fa'a* indicates a non-causal function. Likewise, the occurrences of interrogative pronouns e.g. كم "how" - لماذا "why" in the argument precedes the target word, implies that the instance holds the context of inquiry and the non-causal label seems to be well suited. Adopting more complex models of word n-grams increases the dimensionality of the feature space without bringing more information.

Stop word: We added this binary feature to check if the target word is contained in the stop-words list. While this feature may be useful for one proclitic, it may be the opposite for the others. It is most likely to assign the non-causal class for instances containing stop-word prefixed with *Lām*. However, some stop-words indicate causation if prefixed with *Fa'a* e.g. جميع "all" - بعض "some".

Punctuations position: An Integer feature to record the position of the first punctuation appears near the target word in both directions (before and after). Thus, if the word followed by comma the feature assigns the value 1. It takes the value 0 if no punctuation was found in the sentence. This feature is most valuable when recognizing the syntactical function of *Fa'a*.

Morphological syllables: We include this feature to encode the number of morphological syllables the target word is comprised of. Each target is assigned an integer number corresponding to the how many tokens spotted in the word.

4. Learning Predictive Models

As most machine learning models expects numerical vectors with a fixed size, the initial processing step is to convert our categorical features into numerical variables suitable for feeding into a predictive model. This transformation resulted in a matrix with over 10000 one-hot-variables. Including redundant or irrelevant variables can be misleading to the classifier, for example, Instance-based methods such as k-nearest neighbour select a small number of the closest neighbours in the feature space to produce class membership predictions. These predictions can be greatly skewed by redundant variables, thus, searching for the best subset of variables in the dataset improves the prediction performance.

We used Recursive Feature Elimination (RFE) method to reduce the size of features space and identify which combination of variables contributes the most to predicting the target class. RFE considers the selection of a set of variables as a search problem. It uses statistical model to evaluates, compares and assign a score to different combinations; it then recursively removes attributes based on the model accuracy.

Like other data preparation and cleaning procedures, variable selection must be restricted to a separate data set to avoid any data leakage, over fitting or model biases situations. We split our data into three random subsets where 20% of the instances used for the development stage i.e. variables selection; 55% designated for learning the predictive models and 25% held out for performance evaluation.

As for machine learning models, we used Scikit-Learn implementations of Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Multinomial Naïve Bayes (MNB), Random Forest (RF) and K-Nearest Neighbors (KNN). We aim to discover which of these models will be able to better generalize the training examples and achieve the highest learning performance. However, each model has some hyper-parameters whose values can directly affect its predictive performance. The experimental results show that tweaking the values of hyper-parameters can have a direct impact on the performance. Even when presenting a low average improvement over all datasets, it is statistically significant in most of the cases [17][18]. In order to address this issue, we used Grid Search (GS) algorithm as an optimization technique to find a set of suitable configurations for each model. However, searching over many different parameters simultaneously may be computationally expensive; we picked the parameters that are considered most important for each model.

We constructed a pipeline to chain the sequence of data processing stages into one task. The pipeline was applied on each of the three proclitics dataset separately. It starts with the Vectorization transformation process where the training examples are turned into numerical features vectors. Then RFE with the LR algorithm is applied on the development set to select the top variables of the features space. Next, The GS tuning algorithm is presented with a set of hyper-parameters to determine the optimal values. GS proceeds on the training set to train and evaluate each model for every combination of the parameters. We set up the GS algorithm to use tenfold cross validation splits where models are first fitted to training set within each fold after which they are maximized in electing the hyper-parameters over the validation set. Finally, the optimized learning models are applied on the held back test set for performance analysis.

5. Experimental Results

We run the pipelines over all instances in each dataset i.e. *Lām*, *Fa'a* and *Ba'a*. The experiments used the same GS parameters in the tuning process. The results are compared with three baselines; the first one (B1) is obtained from the rule-based approach proposed in [3]. The second baseline (B2) relates to assigning the most frequent class for every instance in each dataset. The last baseline (B3) is based on evaluating whether the target word's category is Al-Masdar. As we discussed in Section 3.1, some researchers argued that proclitics attached to Al-Masdar normally signal causation. We acquired this baseline from the corpus developed in this study where Al-Masdar is manually annotated with VB tag. Tables 3-5 summarize the experimental results obtained by the constructed models in terms of accuracy and micro-averaged F1-score. Results under *All Variables* segment are obtained by the models constructed based on the entire features vectors. The ones under *Selected Variables* attribute to the models trained on a subset of feature space, where N is the number variables that was found to be optimal for each model. The instances are balanced across target classes; thus the F1-score and the accuracy are almost equal.

For most of the classifiers (15 of 18), the results show that the models trained by combining feature selection step produced statistically significant improvement over the ones trained using the complete set of the feature space. We observe that SVM, LR and RF models obtain the best performance for *Fa'a*, *Lām* and *Ba'a* with F1-score of 0.78, 0.81, and 0.77 respectively. Table 6 compares the best predictive models (BM) with the three baselines. Appendix A presents the hyper-parameters values which attribute to the optimized models.

Considering the instances for which the models produced false positive predictions, it was noted that the target words in most cases occurred within a clause constitutes a basic component of the argument. That is to say, the preceding part cannot stand alone as an *effect* or *cause* argument on its own. As an example, the target word "لخفض" "cut down" in Sentence (4) functions as the predicate of the dependent clause "ان هناك اتجاها" "there is an intention" which provides a necessary context to the first part of the sentence. In fact, determining the syntactic function of a proclitic can be achieved if syntactical trees annotations are incorporated in the feature space. Another area of errors attributes to instances which appear in a semantic context that is rather ambiguous. For example, the word "بدافع" "prompted by" in Sentence (5) is annotated as causal particle, however, we may alternatively select the word "لمشاهدة" "to watch" as the causal indictor.

(4) ان هناك اتجاها <u>لخفض</u> المشتريات بنسبة 5 في لمئة، واجراء خصومات على المخزون الراكد.
"There is an intention to reduce purchases by five percent and to discount the slow-moving inventory."

(5) امتلأ الملعب بالآلاف الذين جاءوا <u>بدافع</u> من الفضول لمشاهدة السباق الغريب، وكان من بينهم الاديب والشاعر.
"The stadium was crowded with thousand prompted by curiosity to watch the unique race, and among them were the writer and the poet."

Table 3. Classification results for Causation *Fa'a*.

	Selected Variables			***All Variables***	
Model	N	F1	**Acc**	**F1**	**Acc**
LR	25	.73	.72	.72	.73
DT	150	.76	.75	.76	.79
RF	80	.72	.72	.68	.67
SVM	15	.78	.80	.73	.74
MNB	20	.74	.76	.70	.71
KNN	15	.75	.77	.72	.72

Table 4. Classification results for Causation *Ba'a*.

	Selected Variables			***All Variables***	
Model	N	F1	**Acc**	**F1**	**Acc**
LR	25	.76	.77	.71	.71
DT	25	.74	.75	.65	.65
RF	20	.77	.76	.72	.72
SVM	40	.75	.75	.71	.72
MNB	10	.74	.75	.72	.73
KNN	25	.73	.74	.67	.67

Table 5. Classification results for Causation *Lām*.

		Selected Variables		*All Variables*	
Model	N	F1	**Acc**	**F1**	**Acc**
LR	1495	.81	.81	.76	.76
DT	590	.78	.77	.78	.79
RF	740	.79	.77	.76	.76
SVM	1100	.80	.81	.76	.76
MNB	1135	.80	.80	.79	.80
KNN	440	.77	.77	.70	.70

Table 6. Best models' performance against the baselines.

	Lām		***Fa'a***		***Ba'a***	
	F1	Acc	**F1**	**Acc**	**F1**	**Acc**
B1	.63	.61	.66	.65	.68	.66
B2	.55	.55	.57	.56	.52	.50
B3	.69	.69	.41	.48	.65	.64
BM	.81	.81	.78	.80	.77	.76

6. Related Work

Causality detection has been studied extensively in a wide range of disciplines [19][20]. Early attempts made use of hand-coded patterns to detect the causation knowledge in texts. In the COATIS system [21], a tool was built to identify causality links in French texts. It applies the strategy of Contextual Exploration, by targeting linguistic indicators of causality in sentences. The system takes into account the context in which the located indicators appear to confirm the presence of *Causal* relations. The author obtained a precision rate of 85%. The authors in [22] identified a set of English linguistic patterns to extract cause–effect templates from a database of medical journal articles' abstracts. They reported to have reached an accuracy of 0.41 and 0.48 for extracting the *cause* and the *effect* slots, respectively.

Machine-learning techniques were employed by a number of studies. Blancol et al. [23] focused their work on extracting patterns with the form *[VP rel C], [rel C, VP]* by which is, as the authors stated, they were able to classify causations signalled by the indicators: *because* and *since*. The system obtained an F measure of 0.89 for cause and 0.91 for not-cause cases. SVM models were built by Beamer et al. [24] for causal knowledge acquisition. The first model included a set of semantic feature represented in 18 lexico-syntactic patterns. It used the SemEval 2007 dataset for training purpose. The other model was trained over annotated texts from the *Wall Street Journal.* The first model achieved an accuracy of 77.5% in identifying cause–effect noun pairs; while the last model obtained a precision of 24.4% and recall of 79.7%. Prasad and Joshi [25] used the Penn Discourse TreeBank (PDTB). The authors aimed to exploit *Causal* relations discovery in answering "*why*" questions. They selected a Question Answer pairs from a text collection of the PDTB corpus. The authors reported that 71% of the questions were correlated with *Causal* relations. Mirza and Tonelli [26] presented an annotation framework to model causal signals (CSIGNAL) and causal relations (CLINK) between events using a web-based application called CAT tool. The documents were taken from TimeBank corpus [27]. They annotated 171 CSIGNALs and 318 CLINKs. A more recent work was presented by Mostafazadeh et.al [28] where a semantic framework was designed to capture a set of temporal and causal relations between events. The authors annotated 1600 sentences taken from ROCStories corpus.

Some works presented for Arabic with the aim of building annotated corpora with discourse relations. Al-Saif and Markert [15] produced the Arabic Discourse Treebank by extracting a list of 80 discourse connectives. They employed the connectives identify 18 adjacent discourse relations. Their final Treebank has 600 sentences annotated with Causal relation. Another annotated corpus was created by [16] based on documents collected from Discourse Arabic Treebank. They grouped the relations into: Temporal, Structural, Causal and Thematic.

7. Conclusion and Future Work

The importance and difficulty of extracting causal information suggest that additional efforts are needed in order to reliably create mature language resources. In Arabic, causal relations indicated by proclitics account for a high percentage of the total *Causal* relation in texts. In the current research we created a causation corpus annotated with instances contain words prefixed with certain proclitic. We used this corpus to build and tune a variety of machine learning models based on shallow linguistic features. A variable selection model was combined to eliminate irrelevant or partially relevant variables that might negatively impact the models' performance. In future, we will test deep linguistic features such as syntactical tree and phrase chunker, we also plan to extend the corpus to include other causal indicators.

Appendix A. hyper-parameters values associated with the optimized models

Model	Parameter	Fa'a	Ba'a	Lām
LR	C	0.1	1	1
MNB	alpha	1.5	2	0.5
	fit_prior	False	False	True
SVM	C	1	10	1000
	coef0	2	0	-
	kernel	poly	Poly	rbf
	degree	5	2	-
	gamma	-	-	0.0001
DT	criterion	entropy	gini	Gini
	splitter	random	random	best
	max_depth	14	5	5
	min_samples_split	20	20	20
KNN	P	1	2	2
	metric	minkowski	minkowski	minkowski
	algorithm	brute	brute	brute
	n_neighbors	15	22	15
RF	criterion	gini	entropy	Entropy
	bootstrap	False	True	False
	max_depth	20	24	24
	n_estimators	30	30	5
	min_samples_split	100	80	100

References

[1] Kaplan, Randy, and Berry-Rogghe, Genevieve. (1991) "Knowledge-based acquisition of causal relationships in text." *knowledge Acquisition* 3(3): 317–337.

[2] Sadek, Jawad., and Meziane, Farid. (2016) "A discourse-based approach for Arabic question answering." Journal ACM Translations on Asian and Low-Resource Language Information Processing 16, 2, Article 11.

[3] Sadek, Jawad., and Meziane, Farid. (2016) "Extracting Arabic causal relations using linguistic patterns." Journal ACM Translations on Asian and Low-Resource Language Information Processing 15, 3, Article 14.

[4] Sadek, Jawad and Meziane, Farid. (2018) "Building a causation annotated corpus: The Salford Arabic Causal Bank - proclitics." In the 3rd Workshop on Open Source Arabic Corpora and Processing Tools, 8 May 2018, Miyazaki, Japan.

[5] Bontcheva, K., Cunningham, H., Roberts, I., Roberts, A., Tablan, V., Aswani, N., and Gorrell. G. (2013) "GATE Teamware: A web-based, Collaborative Text annotation framework." Language Resources and Evaluation. 47(4).

[6] Weiss, Gary M., and Provost, Foster (2003) "Learning when Training Data are Costly: the Effect of Class Distribution on Tree Induction." Artificial Intelligence Research, 19:315-345.

[7] Chawla, Nitesh V. (2003) "C4.5 and imbalanced data sets: Investigating the effect of sampling method, probabilistic estimate, and decision tree structure." In ICML Workshop on Learning from Imbalanced Data sets, Washington, DC.

[8] Davidson, Donald. (1980) "Causal relations." *Essays on actions and events*, 149-162 .Oxford University Press.

[9] Mackie, J.L. (1980) "The cement of the universe: A study of causation." Oxford University Press.

[10] Pedregosa, Fabian., et. al. (2011) "Scikit-learn: Machine Learning in Python" JMLR 12:2825-2830.

[11] Gildea, Daniel., and Jurafsky, Daniel. (2000) "Automatic Labeling of Semantic Roles." In *Proceedings of the 38th Annual Conference of the Association for Computational Linguistics (ACL-00)*, 512-520.

[12] Blaheta, Don and Charniak, Eugene (2000) "Assigning Function Tags to Parsed Text." In *Proceedings of the 1st Annual Meeting of the North American Chapter of the Association for Computational Linguistics*, Seattle, 234–240.

[13] Fasha, Mohammad., Obeid, Nadim., and Hammo, Bassam. (2017) "A Proposed Model for Extracting Information from Arabic-Based Controlled Text Domains." In Proceedings of the New Trends in Information Technology (NTIT-2017). Amman, Jordan.

[14] Bouaziz Mezghanni, Imen and Gargouri, Faiez. (2017) "Deriving ontological semantic relations between Arabic compound nouns concepts." Journal of King Saud University, computer and Information Science. vol: 29(2), 212-228.

[15] Al-Saif, Amal, Katja Markert. (2011) "Modelling discourse relations for Arabic." The proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, (EMNLP 2011), Edinburgh

[16] Keskes, Iskandar., Zitoune, Farah. B ., and Belguith, Lāmia. H. (2014) "Learning explicit and implicit arabic discourse relations." Journal of King Saud University, computer and Information Sciences vol: 26 (4), 398–416.

[17] Mantovani, Rafael G., Horváth, Tomáš., Cerri, Ricardo., Vanschoren, Joaquin., and de Carvalho, André C.P.L.F. (2016) "Hyper-Parameter Tuning of a Decision Tree Induction Algorithm." Brazilian Conference on Intelligent Systems (BRACIS), 37-42.

[18] Chapelle, Olivier., Vapnik, Vladimir., Bousquet, Olivier., and Mukherjee, Sayan. (2002) "Choosing multiple parameters for support vector machines." Machine Learning, vol.46(1-3), 131–159.

[19] Hassani, Hossein., Huang, Xu., and Ghodsi, Mansi. (2017) "Big Data and Causality." Annals of Data Science, 1-24.

[20] Asghar, Nabiha. (2016) "Automatic Extraction of Causal Relations from Natural Language Texts: A Comprehensive Survey," *arXiv preprint arXiv:1605.07895*

[21] Carcia, Daniela. (1997) "COATIS, an NLP system to locate expressions of actions connected by Causality Links." In *Proceedings of the 10th European Workshop on Knowledge Acquisition, Modeling and Management*, 345–352.

[22] Khoo, Christopher., Chan, Syin., and Niu, Yun. (2000) "Extracting causal knowledge from a medical database using graphical patterns." In Proceedings of 38th Annual Meeting of the ACL, 336–343.

[23] Blancol, Eduardo., Castell, Nuria., and Moldovan, Dan. (2008) Causal relation extraction. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'08)*, 310–313.

[24] Beamer, Brandon., Bhat, Suma., Chee, Brant., Fister, Andrew., Rozovskaya, Alla., and Girju, Roxana. (2007) "UIUC:A knowledge-rich approach to identifying semantic relations between nominal." In *Proceedings of the 4t International Workshop on Semantic Evaluation*, 386–389.

[25] Prasad, Rashmi., and Joshi, Aravind. (2008) "A discourse-based approach to generating why-questions from text." In *Proceedings of the Workshop on the Question Generation Shared Task Evaluation Challeng.*

[26] Mirza, Paramita., and Tonelli, Sara. (2014) "An analysis of causality between events and its relation to temporal information." In proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, 2097-2106, Ireland

[27] Pustejovsky, James., Littman, Jessica., Sauri, Roser., and Verhagen, Marc., (2006) "Timebank 1.2 documentation." Technical report, Brandeis University, April.

[28] Mostafazadeh, Nasrin., Grealish, Alyson., Chambers, Nathanael., Allen, James., and Vanderwende, Lucy. (2016) "CaTeRS: causal and temporal relation scheme for semantic annotation of event structures." In Proceedings of the 4th Workshop on Events: Definition, Detection, Coreference and Representation, 51-61, California.