A^{3}C: Arabic Anaphora Annotated Corpus

Abdelhalim Hafedh Dahou, **Mohamed Abdelmoazz** and **Mohamed Amine Cheragui**
Mathematics and Computer Science Department
Ahmed Draia University
Adrar, Algeria
dahou.halim1995@gmail.com, m.abdelmoazz@yahoo.com, m_cheragui@univ-adrar.edu.dz

Abstract

In this paper, we describe the different steps taken to build our annotated corpus which aims to treat a known linguistic phenomenon in Arabic texts called Anaphora. The objective behind the creation of this corpus[1] is to fill the lack of resources concerning the resolution anaphora (especially pronominal and verbal) in the Modern Standard Arabic language and this is by creating a newly annotated corpus that we have called A^{3}C which contains the anaphoric relations. To satisfy this objective, we created A^{3}T, an anaphoric annotating tool that uses linguistic and statistical rules to automatically detect anaphors and their referents. After that, we resort to human specialists to verify and correct our A^{3}T annotation's errors for the corpus's credibility. This study discusses novel features that can aid in determining the best reference, as well as the problem of the lack of resources for verbal anaphora.

1 Introduction

A corpus is considered today as a fundamental piece in natural language processing, due to the role that it plays in both the resolution and the testing phases. The building of annotated corpus in terms of number and size has known a real ascension in the last decades, in particular since the appearance of statistical and machine learning approaches (Beseiso and Al-Alwani, 2016), allowing, from textual resources, the development of resolution models for different linguistic phenomena such as anaphora.

Anaphora is typically defined as references to items mentioned earlier in a discourse or "pointing back " reference as described by (Mitkov, 99). In addition, the process of determining the referent of an anaphora and establishing the relationship between them is known as anaphora resolution. Anaphora still a very challenging linguistic phenomenon, where its identification and resolution can increase the performance of several NLP applications, such as: sentiment analysis (Cambria, 2016), question-answer systems (El-Said Nada et al., 2018), machine translation (Madhura and Satish, 2019), text summarization (Antunes et al., 2018), information extraction (Matysiak, 2007), language generation and dialog systems (Vinay et al., 2019).

Our motivation behind this work is to enhance anaphora resolution in Arabic text by building an anaphoric annotated corpus that can contribute to future works that tackle anaphora in the Arabic language.

This paper is structured in 6 sections. Section 2, describe the anaphoric typology in Arabic language. Section 3, gives an overview of existing anaphoric corpora (case of Arabic). Section 4, presents the challenges we face in Arabic anaphora resolution. Section 5, outlines the different phases of building of our A^{3}C corpus. Section 6, some observations noted during the building process of our corpus. The last section gives a conclusion and future work.

2 Varieties of Anaphora in Arabic text

What makes the anaphora resolution mechanism complex in natural language processing in general and in Arabic, in particular, is the fact that it can

[1] The Corpus is available for the community in : https://dahouabdelhalim.github.io/Anaphora-Corpus/

manifest in different forms (linguistic categories: lexical and grammatical), but also requires knowledge at different levels, as well as an "understanding" of the context. There are many varieties of anaphora in the Arabic text, we will only mention the most frequent ones.

2.1 Verbal anaphora

Verbal anaphora is used to describe or represent various movements or actions by using the verb (did - فعل -) and the different conjugation variants to minimize writing and avoid repetition (Trabelsi et al., 2016; Hamouda, 2014).

Translation: "Students who are about to graduate must complete their administrative file, if they do not, they will not receive their certificates."

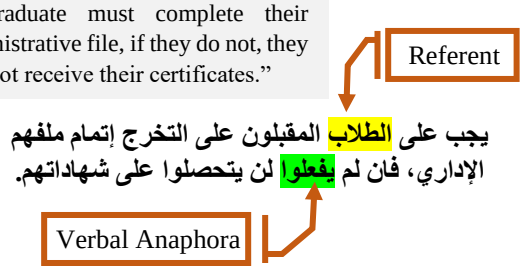

Figure 1: Example of verbal anaphora.

2.2 Lexical anaphora

Lexical anaphora occurs when the referent is designated by definite descriptions representing the same concept (the anaphora), or concepts that are semantically close (Hammami, 2009). Usually, this form of anaphora adds more information to the sentence and increases cohesion, and can take several forms (synonym, generalization / hypernymy, or specialization /hyponymy) (Seddik and Farghaly, 2014).

Translation: "Ibn Sina was born in Bukhara (in present-day Uzbekistan), one of the most famous writings of the scientist *The Canon of Medicine*".

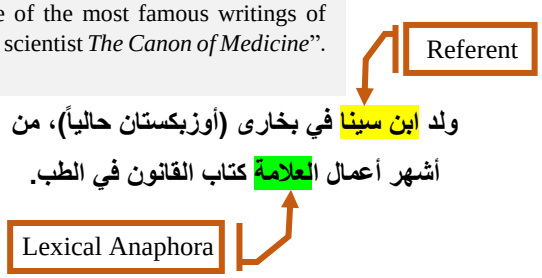

Figure 2: Example of lexical anaphora.

2.3 Comparative anaphora

This type of anaphora is manifested by the introduction of lexical modifiers (e.g., أخر / other, أخرى / one, وحدة) or comparative adjectives (أكبر من / greater than , أحسن من / better than) (Hammami, 2009). This variety of anaphora indicates a relation like: such as set-complement, similarity and comparison between the anaphora and the referent (Mahmoud Seddik and Farghaly, 2014).

Translation: "Departure of Vessels transporting more than 50,000 tons of agricultural nutrients (urea) and various other from the industrial port of AL Jubail in the Kingdom to the Port Sudan in Khartoum."

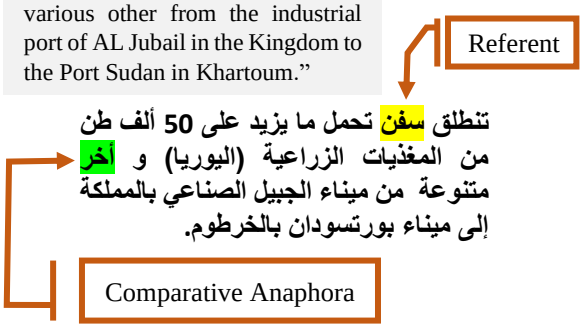

Figure 3: Example of comparative anaphora.

2.4 Pronominal anaphora

Based on statistical studies done by (Hammami, 2009) it shows that the pronominal anaphora is the most frequent variant in Arabic texts. Pronouns form a special class of an"aphora because of their empty semantic structures; they have a meaning independent of its referents and usually refer to names or noun phrases (Beseiso and Al-Alwani, 2016). However, not all pronouns are anaphoric.

Pronominal Anaphors can be divided into three categories, each category can be subdivided into subcategories according to several parameters, such as gender, number, etc.

3rd personal pronouns (ضمائر الغائب): In the Arabic, not all personal pronouns are anaphoric, so the 1st person (انا و نحن) and 2nd person (أنت انتما .., etc.) pronouns are not (they specify the communication partners and their meaning goes back to their specific uses), except the 3rd person pronouns which have this characteristic. These pronouns can be subdivided into two categories: disjoint pronoun (Example: هي / she ،هو/ he) and joint pronoun (Example: ـه، ا، ن) (El-Said Nada et al., 2018):

Translation: "Soumaia has sewn her sister's wedding dress and she is very excited".

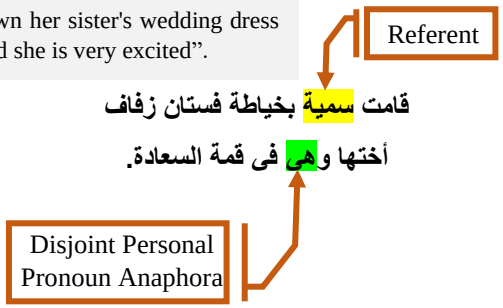

Figure 4: Example of disjoint personal pronoun anaphora.

In some cases the pronouns "ـه" and "ـها" are not anaphoric since they are not interpreted as related to an expression (referent). In this case we will call them pleonastic pronouns.

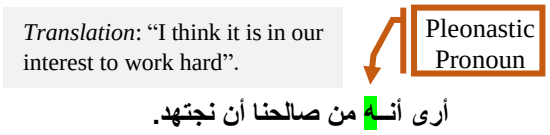

Figure 5: Example of pleonastic pronoun.

Relative pronouns (الأسماء الموصولة): Relative pronouns in Arabic have the characteristic of being always anaphoric, in addition they have only one possible referent (Trabelsi, 2016) and refer to the immediate nominal phrase mentioned before (Bouzid et al., 2014) which they agree in gender and number.

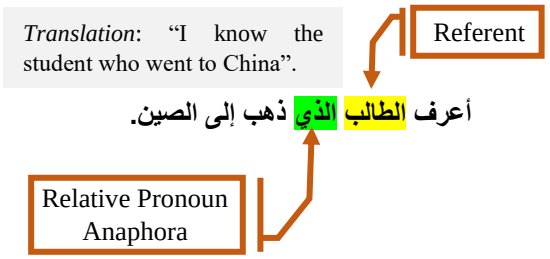

Figure 6: Example of relative anaphora.

The use of relative pronouns is possible if the referent denotes a process or situation, and here the anaphora denotes some of these lexical meanings. They refer to persons, places or things that are close or distant, the table below illustrates this type of pronouns.

Demonstrative pronouns (أسماء الإشارة): They are linguistic elements that accompany a designation gesture in order to coordinate the attention of the interlocutors when they are speaking (Jarbou, 2018). Generally, demonstrative pronouns are cataphoric and in some cases they can be anaphoric and even deixis (Bouzid et al., 2014). Demonstratives agree in person, gender and number with their referent. In addition, there are pronouns, which are considered demonstratives, and which designate time and place (Example: هذا / this, هنا / here).

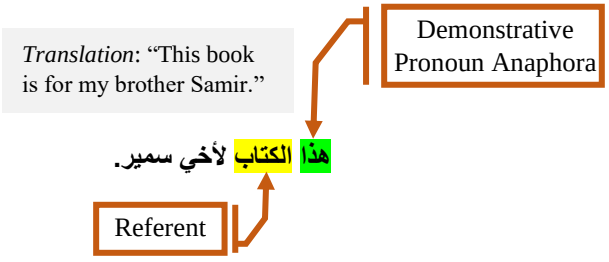

Figure 7: Example of demonstrative anaphora.

3 Related Work

For Arabic language, a considerable effort has been made concerning the anaphoric phenomenon during the last two decades, which is reflected by several studies aiming in their majority to solve the problem of the pronominal anaphora. The objective of this section is to present an overview of works dedicated to building annotated corpus (anaphora identification and referent determination).

Corpus	Size	Anaphoric Resolution Category
AnATAr (Hammami, 2009)	18895 words 2722 pairs of anaphor /referent.	Pronominal anaphora
(Hadder, 2000)	200 Sentences	Zero Anaphora
Holy Qur'an Corpus (Farghaly and Fahmy, 2015)	127,795 words 24,653 personal pronouns	Pronominal Anaphora
QAC (Sharaf and Atwell, 2012)	128,000 words 24,679 Pronouns	Pronominal Anaphora

Table 1 : Existing corpora concerning the Arabic anaphora.

4 Ambiguities and anaphoric resolution

The aim of this section is to present the main factors, which affect anaphoric resolution.

4.1 Ambiguities and lack of diacritics

Without diacritics marks, an Arabic text is extremely unclear (morphologically and grammatically). According to (Debili and Achour, 1998), 74% of Arabic words might potentially take several lexical diacritization, making it difficult to determine if the anaphoric phenomenon or referent is the case.

Word	Word + Diacritics	Translation
كـتـب	كَـتَـبَ	he wrote
	كُـتُـبٌ	books
	كُـتِـبَ	Written
	كُـتِّـبَ	was caused to write
	كَـتَّـبَ	To make someone to write

Table 2: Example of ambiguities due to the lack of diacritics.

4.2 Agglutination phenomenon

The Arabic script is characterized by the agglutination phenomena, which is explained by the fact of combining numbers of words in just one. Compared to French or English, an Arabic word can sometimes correspond to a full sentence (Bouzida and Zribi, 2020).

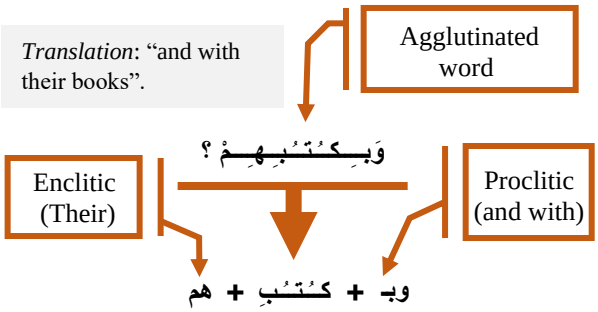

Figure 8: Example of agglutination.

4.3 Syntactic flexibility (Words free order)

Arabic is a nearly free-order language. This order causes artificial syntactic ambiguities, since the grammar should provide all the possible combination rules for reversing the order of words in the sentence. For anaphora resolution, this type of flexibility is a problem for referent localization (Beseiso and Al-Alwani, 2016; Fotiadou et al., 2020).

Sentences	English Translation	Order
قرأ محمد الكتاب	Mohamed Read the book	VSO
محمد قرأ الكتاب	Mohamed, he read the book	SVO
الكتاب محمد قرأه[2]	The book Mohamed read it	OSV
قرأه[3] الكتاب محمد	The Book was reading by Mohamed	VOS

Table 3: Words free order in Arabic sentences.

4.4 Ambiguity of the referent

This difficulty occurs when the referent is ambiguous (due to the presence of two or more referents for the same anaphora). In this case, external knowledge of the context is necessary to identify the correct referent (Brunner et al., 2002).

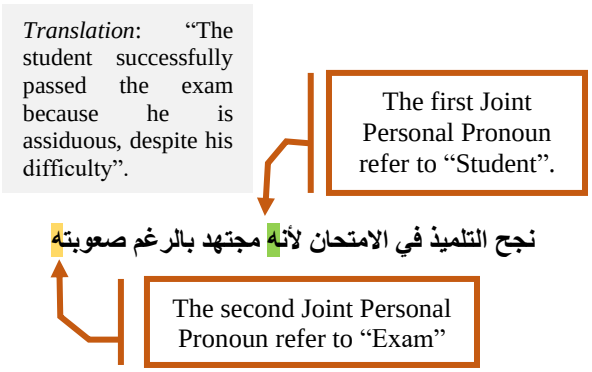

Figure 9: Example of ambiguity of the referent.

4.5 Hidden referent

This case occurs when the anaphora refers to something, which is not present in the sentence or text. The Qur'anic text is an example where this phenomenon persists (Seddik and Farghaly, 2011), so in the example below the pronominal anaphora (هو / he) refers to (الله /Allah) which is not present in the "Aya". The human through his knowledge and reasoning system can easily make the connection between the pronominal anaphora (هو / he) and (الله /Allah). However, for anaphoric resolution systems the task is complicated.

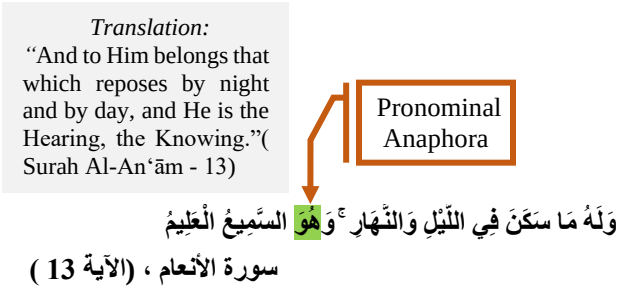

Figure 10: Example of hidden referent.

5 Building the A3 C

As mentioned above, the main objective is to provide an annotated resource that can be used in the automatic Arabic anaphora resolving systems. We decided to create an operational tool with a friendly interface that would help computer scientists and linguists to develop such resources.

In this section, we'll go over the steps involved in building our corpus A^{3}C and annotating it with our A^{3}T system. We thought about breaking down the creation of our work environment into three

[2] Joint Personal Pronoun « ه » are anaphoric.

[3] Joint Personal Pronoun « ه » are cataphoric

(03) phases: data collection, anaphora resolution system, corpus annotation and verification. Each phase consists of essential modules that take place to accomplish the phase's purpose.

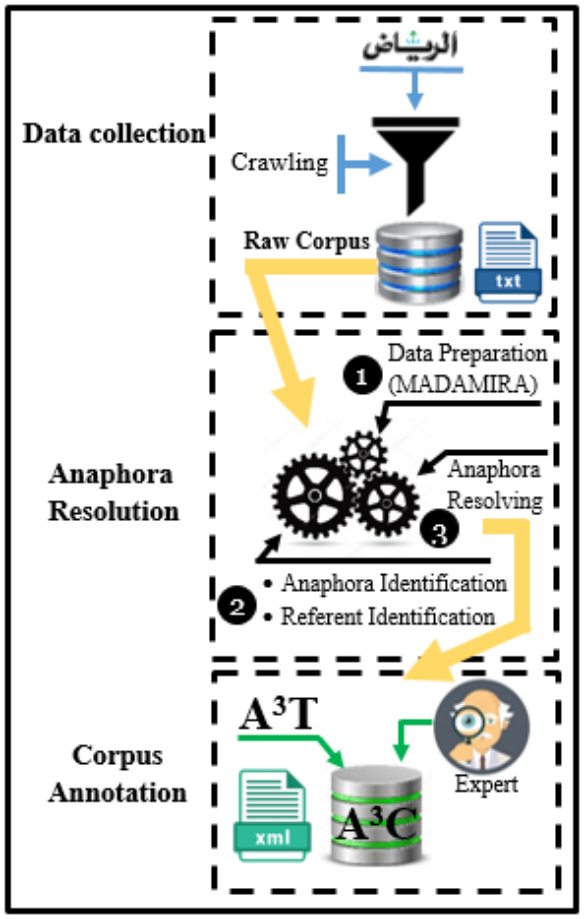

Figure 11: General architecture of building A³C.

5.1 Data collection

Our purpose is to build a corpus of texts from different fields to cover two types of anaphora, pronominal and verbal anaphora. The texts in our corpus are taken from the **Alriyadh newspaper**[4], a daily Arabic newspaper, and they are divided into five categories: culture, sports, politics, economy, and miscellaneous. The choice of those categories is made after an analysis of different categories of texts in terms of the number and diversity of anaphora types. On the other hand, the choice of this newspaper is due to the volume of information, good structure of articles and diversity of categories. To attend to this objective, we developed a crawler system that takes as an input the URL of the category page and the limited number of articles, then returns as an output a cleaned text file in (.txt) format.

[4] https://www.alriyadh.com/

5.2 Co-reference Resolution

We all know how effort and time consuming it is to manually resolve anaphora and annotate a text corpus. As a result, we created the A3T (Arabic Anaphora Annotating Program), a tool that manages resolution and annotation in an automatic way, while also providing a user-friendly interface to modify the results. The resolution process was divided into two sub-modules:

Data Preparation: To help us address the anaphora problem, the text corpus must go through three processes. The first step is to break each text file into sentences using a sentence splitter mechanism based on the punctuations. Secondly, organizing these sentences in a specific input structure to prepare them for the POS and morphological analysis (Figure 11). Finally, determine which grammatical category a given word belongs to and other morphological features such as gender, number, state, voice. The MADAMIDA tool was chosen for our purposes because of its 95.9% precision and high-quality word-level disambiguation as mentioned in (Pasha et al., 2014). The word-level disambiguation functionality will help us in the identification of the attached pronouns.

Translation: "A pilot attacks a building with his plane."

```
< preprocessed SEN 1 ="طيار يقتحم مبنى بطائرته" >
  <word id="1" word=" طيار ">
    <morph_feature_set pos="noun" vox="na"
      mod="na" gen="m" num="s" stt="c" />
  <word id="2" word=" يقتحم ">
    <morph_feature_set pos="verb" vox="a" mod="u"
      gen="m" num="s" stt="na"/>
  <word id="3" word="مبنى">
    <morph_feature_set pos="noun" vox="na"
      mod="na" gen="m" num="s" stt="i"/>
  <word id="4" word="بطائرته" >
    <morph_feature_set pos="noun" vox="na"
      mod="na" gen="f" num="s" stt="c" />
      <tok id="0" form0="ب+" form1="PREP+" />
      <tok id="1" form0="طائرة" form1="NN" />
      <tok id="2" form0="+ه" form1 = " + POSS
      PRON 3MS" />
</preprocessed>
```

Figure 12: POS by MADAMIRA.

Anaphora Resolution System: A³T allows the expert to select text to automatically detect and resolve anaphora. Once selected, the following three steps are applied to detect and resolve the problem :

- Anaphora identification: Anaphora is identified by referring to their grammatical code, which is based on the MADAMIRA tag set. The output here is a list of all anaphora in the text with additional information like Id, Name, Gender, Number, and Sentence number. For the pronominal anaphora, the process differs from one type to another, for example, the POS tagging for pronominal attached anaphora doesn't have a tag for gender, number, and person because the output is in the attached form, we should apply a split mechanism to place each of them in their proper tag as illustrated in (Figure 13). On the other hand, for verbal anaphora identification, we combine all of the elements used for pronominal anaphora identification, such as gender, number, and so on, with a new feature that will aid in the resolution which is the voice feature (active or passive form). Tables 4 and 5 illustrate the distribution of the various types of anaphora in our corpus after applying this process.

Category	Pro. Ana[5]	POSS[6]	DEM[7]	REL[8]
Economy	18580	40.16%	36.25%	23.59%
Education	28540	47.44%	32.5%	20.06%
Politics	13210	48.73%	27.04%	24.29%
Sport	10953	51.75%	28.5%	19.75%
Miscellany	15069	43.78%	32.21%	25.01%

Table 4: Statistics about the A^{3}C corpus (A).

Category	Verbal Anaphora
Economy	1455
Education	1294
Politics	924
Sport	1370
Miscellany	1932

Table 5: Statistics about the A^{3}C corpus (B).

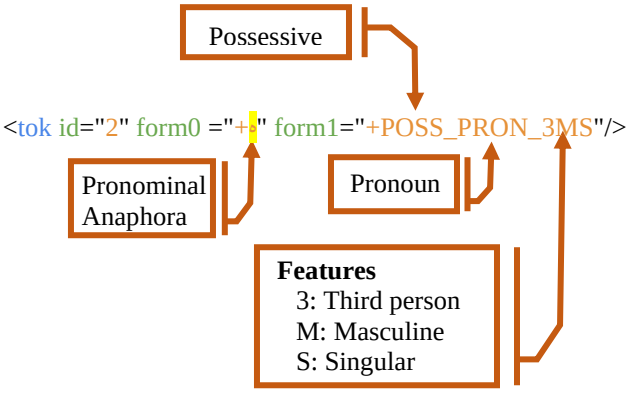

Figure 13: Example of pronominal anaphora identification

- Identification of referent candidates: Referents are chosen based on their POS (nouns, NPs and proper noun) and a specific search scope is adjusted based on some tests and previous research (Mitkov, 99). The search scope is still not fixed in the case of anaphora, but based on analysis, a high number of references occurs on the two previous sentences. In our case, we took two sentences before and as a special case for the demonstrative anaphora, we took the same number after. For the case of verbal anaphora, in the active form, we took two sentences after the verb and for the passive or unknown form; we took two sentences before the verb. The selection considers all of a candidate's features, including gender, number, voice, definiteness, and sentence number.

- Anaphora resolving: The goal is to choose the most appropriate referents from among the most likely candidates for each anaphora. We used morphological filters to remove unsuitable candidates by comparing gender, number, and existing sentence (search scope). To find the suitable referent, we used a collection of preferential factors that favor certain candidates over others, as shown in Table 6. Each rule has a score that is fixed after a series of experiments that took into account previous work (Abolohom and Omar, 2017). Each candidate was given a score for each rule, and the one with the highest overall score was recommended as

[5] Pro. Ana: Pronominal Anaphora

[6] POSS: Possessive
[7] DEM: Demonstrative
[8] REL: Relative

the best referent. We chose the one that came closest to overcoming the score similarity.

Linguistic rules	Description
Description	A score of 1 is given if an NP is definite and of 0 if not.
Recency	A score of 1 is assigned to the recency (nearest one) NP to the anaphora and 0 if not.
Referential Distance	A score of 2 is assigned to NPs in the previous sentence or two sentences and further than those are given 0.
First Noun Phrases	A score of 1 is issued to the first NP of each sentence and 0 if not.
NPs in the title	A score of 1 is issued to the existing NP in title and 0 if not
Grammatical function	Scores of 1 are given to an NP that has the same morpho-syntactic features as the anaphora and 0 if not.
Frequency of NP in text	A score of 2 is assigned to the most frequent NP in text and 0 if not.

Table 6: The linguistic preferences and their respective Scores.

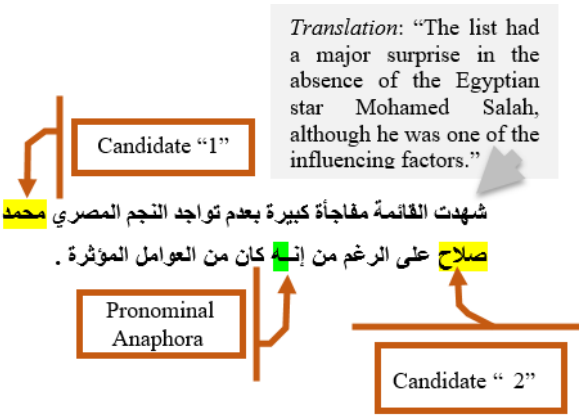

Figure 14: Score similarity (example in pronominal case).

Algorithm: Anaphora Resolution

```
1: input: Anaphora list A, Candidate list C, score function S A, C)
2: initialize score = 0, results = []
3: for t = 1 to N do
4:     for b = 1 to Z do
5:       if Cb [ID] < At [ID] do          ▶ compare if the candidate before or
                                            after the anaphora
6:           score = S (At, Cb)           ▶ S function will calculate every rule
                                            and join the score
7:           results.ADD (At, Cb, score) ▶ store the pairs with their final score
8:       end if
9:       else
10:          break
11:      end else
12:    end for
13: end for
14: results = results.GroupBy(At[ID]).max(score))  ▶ take the pairs with
                                                     maximum score
                                                     for each anaphor
15: output results list with best referent for each anaphor
```

Figure 15: Anaphora resolution heuristics.

5.3 Corpus annotation and verification

This phase aims to annotate the text document using the obtained information from the previous phase, which is a list of pairs of anaphora and their appropriate referent, along with features like Gender, Number, Type, and POS. We used our tool A^{3}T to make the annotation process simpler and fast.

The tool offers a user-friendly interface to linguistic experts, allowing them to check and, if possible, change the connections between anaphora and its referent, resulting in a reliable corpus that can be used in other studies.

More specifically, the interface displays the annotated text in the center, while all of the couples anaphora/candidates are displayed on the right, with the system's chosen couple.

In this case, all the expert has to do is check whether the anaphora tag's number of referent matches the correct one, if not, he may adjust the number of referent to the correct one from the other suggested couples or create a new one if the system doesn't find out the correct antecedent.

In the final part, the tool will add automatically the following tags for the referent and the anaphora: the first will be marked with the <Referent> tag. The remaining elements (anaphora) will be marked with <Anaphor >. We also include the features listed above in each referent and anaphora tag. Finally, the A^{3}T will generate an XML file that contains the text with anaphoric relationship tags as shown in Figure 15.

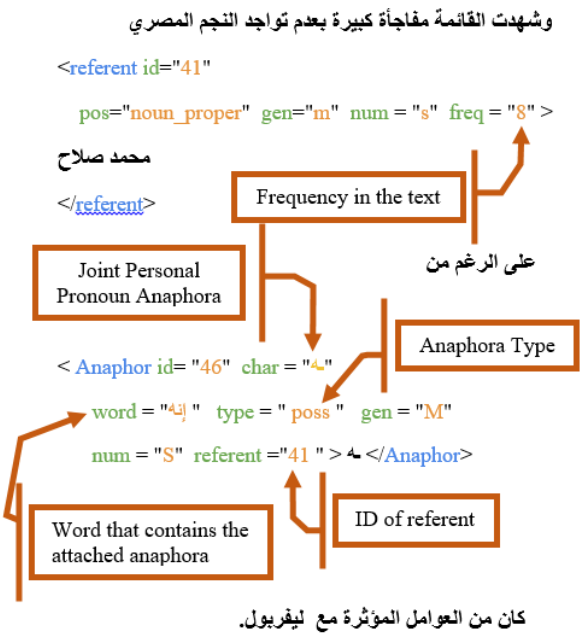

Figure 16: Score similarity (example in pronominal case)

6 Results and Discussion

In our A^{3}T system's testing, we used the "AnATAr" corpora for the evaluation for both anaphora phenomena. We used the standard accuracy metric to calculate the efficiency of the A^{3}T and Table 7 presents the results obtained. We were unaware of any prior works on verbal anaphora, so we tested our work by taking the help of a linguistics specialist and utilizing the same corpora.

Anaphora Type	Corpus	Accuracy achieved
Pronominal anaphora	**AnATAr**	**83.19%**
Verbal anaphora	**AnATAr**	**57.23%**

Table 7 : The result of anaphora resolution system A^{3}T on AnATAr corpora

After analyzing the output of our system, particularly for the verbal anaphora, we found some factors that have influenced our findings. The first factor is the word disambiguation tool limitation that can't in some cases specify the correct meaning of a word such as "سموه" which can act as verb (name; designate) and noun (Highness; grace) or "اخاه" (brother, fraternize). The second factor is the search scope, which could also lead to the best referent being excluded from the list of referents due to being out of scope. In the automatic resolution, the tool rid the references that span multiple sentences but we correct this issue in the expert verification part. The third factor is that the MADAMIRA tool can't recognize composed words like " جمهورية مصر العربية" (Arab Republic of Egypt) or even compound proper names that always occur together like "محمد صلاح" (Mohamed Salah). Finally, in some situations, the voice feature causes a faulty judgment when deciding if the better referent occurs before or after the verb anaphora.

7 Conclusion

Anaphora plays an important role in understanding text and making it coherent. At the same time, it is still a challenging task in the Arabic language due to the complexity of language, the low number of tools, and the lack of linguistic resources. Our present work will make a contribution in the field of linguistic resources for anaphora in the Arabic language and that by providing an annotated corpus that takes into consideration the pronominal and the verbal type. In terms of reducing effort and time consuming during the phase of resolution and annotating, we created A^{3}T, a tool that uses linguistic concepts to identify this phenomenon. With the help of the expert, we are sure that the A^{3}C will be very useful to use in terms of developing intelligence tools that tackle the Arabic anaphora. For the perspectives, our vision will concentrate on the amelioration of the verbal resolution mechanism by using state-of-the-art tools and methods in computational linguistics and at the same time increase the size of the A^{3}C corpus.

References

Abolohom Abdullatif and Omar Nazlia. 2017. A Computational Model for Resolving Arabic Anaphora using Linguistic Criteria. *Indian Journal of Science and Technology*. Volume 10. Issue 3.

Antunes Jamilson , Dueire Lins Rafael , Lima Rinaldo , Oliveira Hilário , Riss Marcelo , Simske Steven. 2018. Automatic cohesive summarization with pronominal anaphora resolution. *Computer Speech & Language*. Volume 52, page (s) 141–164.

Baker Kathryn, Brunner Annelen, Mitamura Teruko, Nyberg Eric, Svoboda Dave and Torrejon, Enrique. 2002. Pronominal Anaphora Resolution in the KANTOO Multilingual Machine Translation System. *Language Technologies Institute, Carnegie Mellon University*.

Beseiso Majdi and Al-Alwani Abdulkareem, 2016. A Coreference Resolution Approach using Morphological Features in Arabic. *International*

Journal of Advanced Computer Science and Applications. Volume 7. Issue 10.

Bouzida Saoussen Mathlouthi and Zribi Chiraz Ben Othmane. 2020. A Generic approach for Pronominal Anaphora and Zero Anaphora resolution in Arabic language. *In proceeding of the 24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems*.

Cambria Erik. 2016. Affective computing and sentiment analysis. *IEEE Intelligent Systems*. Volume 31(2). Page (s) 102–107, 2016.

Debili Fathi and Hadhémi Achour. 1998. Voyellation automatique de l'arabe. *In Proceedings of the Workshop on Computational Approaches to Semitic Languages.* Page (s). 42–49.

El-Said Nada Aya Nabil Mostafa, Saad Sameh and Al-Ansary Abou El-Magd. 2018. A Syntactic based Approach to Anaphora Resolution in Arabic. *In proceeding of The Eighteenth Conference on Language Engineering (ESOLEC'18)*.

Fotiadou Georgia, Muñoz Ana Pérez and Tsimpli Ianthi Mari. 2020. Anaphora resolution and word order across adulthood: Ageing effects on online listening comprehension. *Glossa: a journal of general linguistics.* Volume 5. Issue 1. Page (s) 1–29.

Haddar Kais.2000. Caractérisation Formelle des Ellipses de la Langue Arabe et Processus de Recouvrement de la Langue Arabe. PhD thesis, University of Tunis.

Hammami Souha, Belguith Lamia, and Ben Hamadou Abdelmajid. 2009. Arabic Anaphora Resolution: Corpora Annotation with Coreferential Links. *The International Arab Journal of Information Technology.* Volume 6, No. 5.

Hamouda Wafya. 2014. Anaphora Resolution for Arabic Machine Translation: A Case Study of Nafs. *Ph.D. dissertation, University of Newcastle Upon Tyne*.

Jarbou Samer Omar. 2018. Time frame as a determinant of accessibility of anaphoric demonstratives in Classical Arabic. *Topics in Linguistics.* Volume 19. Issue 2. Page (s) 57-71.

Madhura Phadke, and Satish Devane. 2019. Pronoun Resolution Task for Multilingual Machine Translation. *In Proceeding of the 5th International Conference on Next Generation Computing Technologies (NGCT-2019)*.

Mathlouthi Bouzid Saoussen, Zribi Chiraz Ben Othmane and Trabelsi Fériel Ben Fraj. 2017. How to combine salience factors for Arabic Pronoun Anaphora Resolution. *In the proceeding of the 4th ACS/IEEE International Conference on Computer Systems and Applications*.

Matysiak Ireneusz. 2007. Information Extraction Systems and Nominal Anaphora Analysis Needs. *In Proceedings of the International Multiconference on Computer Science and Information Technology*. Page (s) 183–192.

Mitkov Ruslan 1999. Anaphora resolution: The state of the art. *In Paper based on the COLING'98/ACL'98 Tutorial on Anaphora Resolution, (University of Wolverhampton)*.

Pasha Arfath , Al-Badrashiny Mohamed , Diab Mona , El Kholy Ahmed , Eskander Ramy , Habash Nizar , Pooleery Manoj , Rambow Owen , Roth Ryan. 2014. Madamira: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic. *In the Proceeding of LREC.* Page (s) 1094–1101.

Seddik Khadiga Mahmoud and Farghaly Ali. 2014. Anaphora Resolution. *Chapter of book, Theory and Applications of Natural Language Processing, Springer.*

Seddik Khadiga Mahmoud and Farghaly.Ali. 2011. Arabic Anaphora Resolution Using Holy Qur'an Text As Corpus. *In proceeding of Arabic Language Technology International Conference (ALTIC).*

Seddik Khadiga, Farghaly Ali and Fahmy Aly Aly. 2015. Arabic Anaphora Resolution: Corpus of the Holy Qur'an Annotated with Anaphoric Information. *International Journal of Computer Applications*. Volume 124. Issue 15.

Sharaf Abdul baquee.,and Atwell Eric. 2012. QurAna. Corpus of the Quran annotated with Pronominal Anaphora. *In the Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12).*

Sukthanker Rhea, Poria Soujanya, Cambria Erik and Thirunavukarasu Ramkumar. 2020. Anaphora and Coreference Resolution: A Review. *Information Fusion*. Volume 59.

Trabelsi Fériel Ben Fraj, Zribi Chiraz Ben Othmane and Mathlouthi Saoussen. 2016. Arabic Anaphora Resolution Using Markov Decision Process. *In the Proceeding 17th International Conference on Intelligent Text Processing and Computational Linguistics*.

Trabelsi fériel ben fraj. 2016. A Novel Approach Based on Reinforcement Learning for Anaphora Resolution. *In the proceeding of the 28th International Business Information Management conference.*

Vinay Annam, Nikhil Koditala and Radhika Mamidi. 2019. Anaphora Resolution in Dialogue Systems for South Asian Languages. *arXiv:1911.09994*.