

ARABIC ANAPHORA RESOLUTION USING HOLY QUR'AN TEXT AS CORPUS

Khadiga Mahmoud Seddik, Dr.Ali Farghaly

*Faculty of Computer and Information, Cairo University , Ahmed Zewel St., Giza, Egypt
Kh_seddik88@yahoo.com, Ali.Farghaly@oracle.com*

Dr.Ali Ali Fahmy

*Faculty of Computer and Information, Cairo University , Ahmed Zewel St., Giza, Egypt
a.fahmy@fci-cu.edu.eg*

Keywords: Anaphora Resolution, Arabic Language, Qur'anic text, Statistical Approach

Abstract: This paper presents work in progress about Anaphora resolution for Arabic texts. Anaphora Resolution is a complicated problem in NLP and has attracted the attention of many researchers. Anaphora is a linguistic relation between two textual entities which is defined when a textual entity (the anaphor) refers to another entity of the text which usually occurs before (the antecedent). The process of determining the antecedent of an anaphor is called anaphora resolution.. In this paper, we present a general survey about anaphora resolution and discuss our suggested work on Arabic anaphora resolution. The suggested technique follows the statistical approach to AR and applies AR on Qur'an script.

1 INTRODUCTION

Anaphora Resolution is one of the challenging tasks of natural language processing. It is specifically concerned with matching up particular entities or pronouns with the nouns or names that they refer to.

The poor performance of current Machine Translation systems from Arabic to other languages in terms of anaphora resolution, and the fact that anaphora resolution is an understudied issue in Arabic Natural Language Processing (ANLP) (Elghamry et al., 2007) are the main motivations for this work.

Most of current machine translation systems dealing with Arabic texts, such as Sakhr MT system (www.ajeeb.com) and Google MT systems (<http://translate.google.com>), face problems because of lack of reliable Anaphora Resolution. (Elghamry et al., 2007)

The main reason for Anaphora resolution errors made by such systems are due to the difference between Arabic and English pronominal systems in syntax, morphology and semantic.

There exist a lot of researches on anaphora resolution for English such as MUC-7 (Hirschman, 1997), GNOME (Poesio, 2004), etc., and for French ARCADE (Tutin et al., 2000), FReeBank (Salmon-Alt et al., 2004), etc. Nevertheless, this is not the case for the Arabic language which has few works in this field (Hammami et al., 2009). To the best of our knowledge there is only one research work on Arabic anaphora resolution done by Mitkov and Belguith.

The Mitkov's approach was initially developed and tested for English and was later adapted for Arabic by (Mitkov and Bilguith, 1998).

The proposed work makes use of statistical approach for resolving only pronominal anaphora on Holy Qur'an script.

This paper is structured in 7 sections. In the next section, we present basic notions and terminologies,. Section 3 presents the types of anaphora and the difference between anaphora in Arabic language and other languages. Section 4 presents the challenges we face in Arabic anaphora resolution task. Section 5 presents the process of anaphora resolution,

existing techniques, and our proposed technique. In section 6, we present literature review and previous work. In section 7, we present a statistical study of different types of anaphors and discuss the suggested work. Section 8 presents a conclusion.

2 BASIC NOTIONS AND TERMINOLOGY

The classical definition for anaphora is given by (Halliday and Hasan, 1976): anaphora is cohesion which points back to some previous item. (Mitkov, 1999)

Anaphora is a linguistic relation between two textual entities which is defined when a textual entity refers to another entity of the text which usually occurs before. The "pointing back" (reference) is called an anaphor and the entity to which it refers is its antecedent. If both antecedent and the anaphor have the same referent in the real world, they are called Co-referential.

The process of determining the antecedent of an anaphor is called anaphora resolution. Anaphoric expressions could be one of four linguistic categories: pronouns, verbs, nouns and definite descriptions. (Hammami et al., 2008)

Note that, not all varieties of anaphora are based on referring expressions such as verb anaphora in example 1

The girl went to school as she does every day

3 TYPES OF ARABIC ANAPHORA

3.1 Pronominal Anaphora

The most frequent type of anaphora is the pronominal anaphora, which is identified by anaphoric pronouns. Pronouns form a special class of anaphors because of their empty semantic structure; they do not have an independent meaning from their antecedent. Moreover, not all pronouns are anaphoric, such as /I, /you, /we.

Also, pronouns have lots of types in Arabic. Each type is divided into many categories as shown in figure1, 2, 3 and 4.

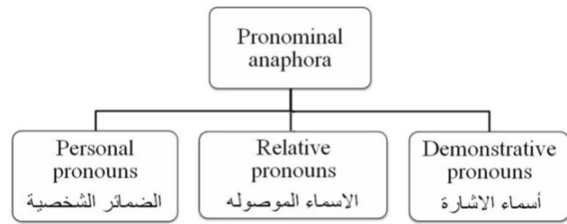


Figure 1: Pronominal anaphora.

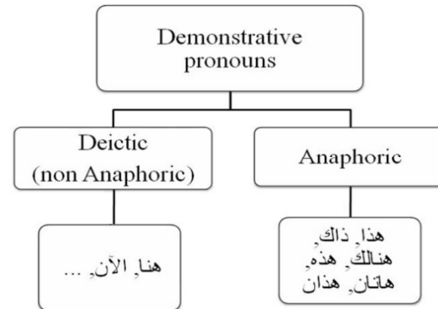


Figure 2: Demonstrative pronouns.

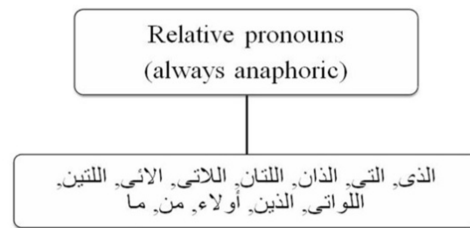


Figure 3: Relative pronouns.

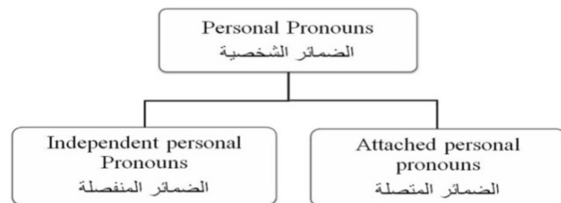


Figure 4: Personal pronouns.

3.2 Lexical Anaphora

Also called "definite noun phrase anaphora" (Mitkov, 1999). It is realized when the referring expressions are definite descriptions or proper nouns, such as example 2. The antecedent is referred by a definite noun phrase representing either same

concept or semantically close concepts (e.g. synonyms, superordinates), such as example 3.

(2) كَذَّبَتْ قَبْلَهُمْ قَوْمُ نُوحٍ ۖ فَكَذَّبُوا عَبْدَنَا ۖ وَقَالُوا مَجْنُونٌ وَازْدُجِرَ ۖ
الْقَمَر (9)¹

The people of Nûh (Noah) denied (their Messenger) before them, they rejected our slave, and said: "A madman!" and he was insolently rebuked and threatened. [Al Qamar (9)]

(3) حضر علماء الكمبيوتر من جميع دول العالم الاجتماع. و قد وجد الحاضرون صعوبة في مجاراة سرعة العرض.

Computer Scientists from many different countries attended the meeting. The participants found it hard to cope with the speed of the presentation.

3.3 Comparative Anaphora

The comparative anaphora is realized when the anaphoric expressions are introduced by lexical modifiers (e.g. آخر ، أخرى) or comparative adjectives (e.g. أكبر ، أحسن). Such as example 4

(4) قَدْ كَانَ لَكُمْ آيَةٌ فِي فِئَتَيْنِ الثَّقَاتِ ۖ فِئَةٌ إِتَّقَى اللَّهَ ۚ فَأَمَّا الْآخَرُ ۖ فَكَافِرَةٌ ۚ [ال عمران (13)]

Here has already been a sign for you (O Jews) in the two armies that met (in combat i.e. the battle of Badr). One was fighting in the Cause of Allâh, and as for the other (they) were disbelievers. [Aal-'Imran (13)]

4 CHALLENGES IN ARABIC ANAPHORA

Beyond the traditional challenges of natural language processing for English, there are unique complexities for the Arabic language, which make the work on Arabic anaphora resolution more difficult than other languages. Here, we will present

some difficulties we face in Arabic anaphora resolution task.

4.1 Lack of Diacritical Marks

Unlike many other languages, Arabic text is usually presented without vowels, which are substituted by diacritical marks placed above or below the character. Diacritics are not included in printed or electronic Arabic text. Therefore, understanding Arabic text and the part-of-speech of Arabic word is difficult and sometimes ambiguous because of the missing of diacritics marks.

For example, the word "كتب" could be active verb (فعل مبني للمعلوم) such as example 6, or passive verb (فعل مبني للمجهول) such as example 7, or plural noun (اسم جمع) such as example 8:

(6) كَتَبَ أَحْمَدُ الْمَقَالَ امس

Ahmed wrote the article yesterday

(7) الْمَقَالَ كُتِبَ امس

The article is written yesterday

(8) اشْتَرَى أَحْمَدُ كُتُبًا جَدِيدَةً

Ahmed bought new books

Without diacritics marks, we must rely on semantic based parsing which is very challenging task.

4.2 Complex Sentence Form Structure

Unlike other languages, Arabic has a unique structure form, which may combine the verb, subject, and object in one word such (انجيناهم) which means "we saved them". We have to break up the phrase, and recognize the subject, before starting the process of anaphora resolution.

4.3 Free Word Order

The Arabic sentence constituents can be swapped without affecting structure or meaning such as example 9. This adds more syntactic and semantic ambiguity, and requires more complex analysis.

(9) ذَاكَرَ الطِّفْلُ الدَّرْسَ بِاتِقَانٍ (VSO)

The boy studies the lesson carefully

الطِّفْلُ ذَاكَرَ الدَّرْسَ بِاتِقَانٍ (SVO)

The boy studies the lesson carefully

ذَاكَرَ الدَّرْسَ الطِّفْلُ بِاتِقَانٍ (VOS)

The boy studies the lesson carefully

4.4 Ambiguity of the Antecedent

In some cases, the anaphora resolution's algorithm fails to identify the correct antecedent because of the

¹ The Noble Quran has been translated into the modern English Language by Dr. Muhammad Taqi-ud-Din Al-Hilali, Ph.D. & Dr. Muhammad Muhsin Khan (<http://www.dar-us-salam.com/TheNobleQuran>)

ambiguity of the antecedent. In these cases, external knowledge about the context is required to identify the correct antecedent (Baker, 2002) such as example 10.

(10) اكلت الطفلة¹ الموزة² لانها¹ جائعة بالرغم انها² لم تنضج بعد
The girl ate the banana because she is hungry although it isn't ripped yet

In each case the pronoun (ها) refers to something different, in 1 الطفلة/the girl, in 2 الموزة/the banana. Each "ها" is going to be interpreted correctly if we use the knowledge that the girl, being human, are likely to be hungry and bananas, being fruit, are likely to be ripe. (Wilks, 1973; 1975).

4.5 Hidden Antecedents

In some cases, especially in Qur'anic texts, the pronoun may refer to something which is not presented in the text, such as example 11. The pronoun (هو) refers to "Allah", which isn't presented in the text. The human mind can determine the hidden antecedent by using the knowledge that Allah is the only one who knows the unseen. But for the anaphora resolution algorithm, it may fail.

(11) وَعِنْدَهُ مَفَاتِيحُ الْغَيْبِ لَا يَعْلَمُهَا إِلَّا هُوَ [الانعام (59)]
And with Him are the keys of the Ghaib (all that is hidden), none knows them but He [Al An'am (59)]

4.6 The Lack of Arabic Corpora Annotated with Anaphoric Links

Still the major problem is the lack of Arabic Corpora annotated with anaphoric relations; however, it is needed in the most of NLP systems. The annotation task of anaphoric relations is very time consuming and requires a huge effort from the human annotator. To the best of our knowledge there is no available resource for Arabic except (Hammami et al., 2009). Arabic annotated resource is much needed to encourage works on Arabic anaphora resolution.

5 THE PROCESS OF ANAPHORA RESOLUTION AND DIFFERENT TECHNIQUES USED

The process of anaphora resolution goes as follow: (1) the first step must be the identification of the search scope. Most approaches set their search scope to the current and one previous sentence. (2) After identifying the search scope, we must identify all

NPs within the search area as candidates for antecedent for specific anaphor. (3) Some anaphora resolution factors are applied on those candidates to select the correct antecedent.

The two main questions for any similar classification problem are the features used for classifier and the classification technique.

As for the first question, we have to clarify the types of factors. Eliminating factors those which exclude certain NPs from the set of possible candidates and Preferential factors those which give more preference to certain candidates and less to others (Mitkov, 1999).

For the second question, we have to presents the anaphora resolution techniques to be used in this step. The difference between each approach in anaphora resolution relies on the factors used in each one.

5.1 Different Techniques for Anaphora Resolution

5.1.1 Rule-based Approach

Rule based approach consists of building a knowledge base, which is written not only by computer scientists but also by language experts. Rule based has some characteristics: (1) It has a strict sense of well-formedness in mind, (2) It imposes linguistic constraints to satisfy wellformedness, (3) It is based on the use of heuristics (such as a verb cannot be preceded by a preposition), and (4) It relies on hand-constructed rules that are to be acquired from linguists rather than automatically trained from data.

The advantages of this approach are that it is easy to incorporate domain knowledge into the linguistic knowledge. Also it can deeply analyze at syntax and semantic levels.

The disadvantage of this approach is the requirement of huge linguistic knowledge and very large number of rules to cover all the features of a language, also there are a lot of exceptional cases which don't follow any rules. Another disadvantage, that there are a lot of ambiguous cases in which the sentence has more than one meaning and can follow more than one rule or can't follow anyone at all.

5.1.2 Statistical Approach

Instead of using huge set of hand coded rules in the traditional rule-based approach, statistical approach uses large set of correctly annotated corpora for training, and algorithm to analyses those corpora

automatically to learn such rules from those trainable model. Thereafter, the model can automatically solve the anaphora in other texts. This approach has many advantages over the traditional approach. Here are some advantages of statistical approach:

1- System based on statistical approach can give more accurate result by just adding more input data (i.e. increasing the training set supplied) of features. On the other side, to increase the accuracy of results for the rule based approach, the only way is to increase the complexity of the rules, which is very difficult task.

2- Statistical approach can make use of statistical inference algorithms which produce strong and robust systems able to handle unfamiliar input, misspelled words, and never seen before structures. That is very helpful for Arabic language as we mentioned before that Arabic has a free word order.

In our work, we suggested to use the statistical approach. The reasons for our choice are mainly because the Statistical approaches are robust, fast, the ability to generalize generated model and it does not require much human interaction.

6 LITERATURE REVIEW

Most of anaphora resolution approaches use the same features, but the major difference among those approaches rely on the way the antecedent is tracked down and computed.

Most traditional approaches to anaphora resolution rely heavily on linguistic and domain knowledge. While alternative approach use the statistical models, e.g., neural networks, a situation semantics framework, or the principles of reasoning with uncertainty (Mitkov and Stys, 1997).

In the rest of this section, previous works for traditional and new statistical approaches will be presented. All these works are on English and other Latin languages except (Mitkov, 1998), which – to the best of our knowledge – the only research work on Arabic anaphora resolution.

6.1 Traditional Approach

6.1.1 D. Carter's Shallow Processing Approach

Carter's Shallow approach mainly depends on heavily use of syntax, semantic, and local focusing as much as possible regardless of the domain

knowledge presented. He said in his PhD thesis (Carter, 1986), that the relying on the large amount of world knowledge is hard to process accurately.

Carter's algorithm resolves anaphora resolution and other linguistics problems using English stories as corpus. The algorithm output paraphrase for each sentence in the story. Carter's program combines many existing theory such as (Sidner, 1979) theory and (Wilks, 1975) theory.

6.1.2 J. Carbonell and R. Brown's Multi-strategy Approach

(Carbonell and Brown, 1988) approach assumed that the anaphora resolution process will be more accurate and best done if we combine more than one strategy

The authors used multiple knowledge sources such as sentential syntax, case-frame semantics, dialogue structure and general world knowledge.

6.1.3 R. Mitkov: Combination of Linguistic and Statistical Methods

The main idea is combining the traditional approach with a new statistical approach for centering tracking. (Mitkov, 1994a; Mitkov, 1996a)

The system uses syntax, semantic, and different types of knowledge. The results show an improvement in resolving anaphors when traditional linguistic approaches are combined with the proposed statistical approach for tracking center (Mitkov, 1999).

6.2 New Statistical Approach

6.2.1 An Uncertainty-reasoning Approach (Mitkov, 1995b)

The approach presented by (Mitkov, 1995b) based on uncertainty reasoning based on the following two lemmas:

1- Natural language understanding problems always suffer from the incomplete information and the missing of some data. The anaphora resolution program has to find the antecedent for the anaphor under the circumstance of the incomplete and missing input data.

2- If we put in our minds that the constraints and preference are originally determined by humans, then, we have to regarded them as uncertain facts (Hammami et al., 2009; Mitkov, 1997).

6.2.2 Two-engine Approach (Mitkov, 1997a)

Mitkov's two-engine approach assumed that combining two engines, each one separately, work successfully with anaphora resolution, will be much better in generating more power and confidence in the search for the antecedent (Mitkov, 1997).

The first engine follows the integrated approach presented in (Mitkov, 1994a), while the second engine follows the uncertainty reasoning approach presented in (Mitkov, 1995b), which make an interactivity between the traditional and new statistical approach.

The results of using both approaches on English text are more accurate.

6.3.3 Multilingual Robust Anaphora Resolution Approach (Mitkov, 1998)

Finally, we will present a short description for (Mitkov and Bilguith, 1998), which – to our knowledge – the only research work on Arabic anaphora resolution.

The approach used in this work consists of a modification of Mitkov's approach (Mitkov, 1998a), and operates on texts pre-processed by a part-of-speech tagger. The system checks input against number and agreement of antecedent indicator. A score is set to each candidate according to each indicator, and the candidate with highest score is returned as antecedent.

The approach initially developed and tested for English, but it has been subjected to some modifications to work with Polish and Arabic.

The approach resolves pronoun anaphora resolution without using linguistic or domain knowledge, or even parsing. Instead, it makes use of corpus-based NLP techniques.

The evaluation for Arabic showed a very high "critical success rate" as well. The robust approach used without any modification scored a "critical success rate" of 78.6%, whereas the improved Arabic version scored 89.3%.

7 STATISTICAL STUDY AND DISCUSSION

In the suggested work, we decide to work on the pronominal anaphora, specifically third personal pronouns. Pronominal anaphora is the most common type of Anaphors; also it's very difficult and complicated.

We made our statistics on a sample of Qur'an text to compare among the appearance of the different types of Anaphors, we calculated the

number of pronouns compared with comparative and verb anaphors in surat AlBaqarah (the longest surah in Qur'an), and we found:

- 1685 pronoun anaphors
- 43 comparative anaphors
- 6 verb anaphors

The statistics covered 6 types of pronouns (According to the classification of (Hammami et al., 2009) for pronouns), 8 types of comparative anaphors, and verb anaphor.

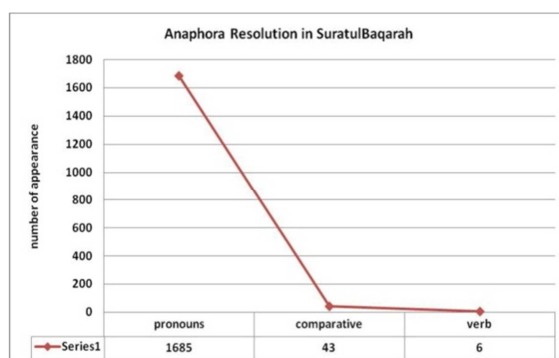


Figure 5: Statistic shows the appearance of different types of anaphora

The statistic shows how the pronominal anaphora frequently appeared more than other types of anaphors.

On the other hand, we mentioned that the pronouns in Arabic have many categories. A statistics made on the different categories of pronouns and we found:

- 1269 Third Personal pronouns
- 369 Relative pronouns
- 47 Demonstrative pronouns

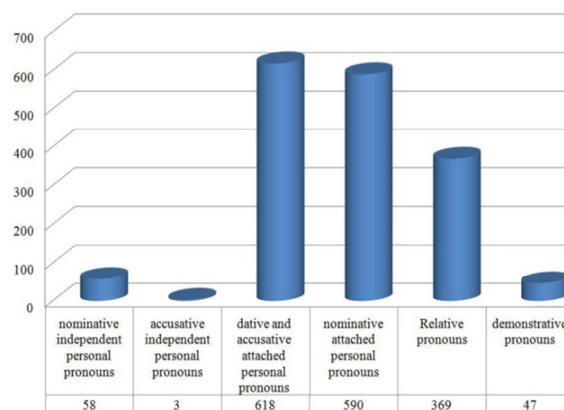


Figure 6: Statistic shows the different types of pronouns anaphor

It is very clear that the personal pronouns are the most common pronouns, specially the attached personal pronouns, whether it was Dative and Accusative, or Nominative ones.

The result of the suggested technique follows the evaluation criteria presented by MUC (Hirschman, 1997). MUC suggested the measure "recall" and "precision" to be used in evaluating the performance of Anaphora resolution systems. Also F-measure, precision rate and success rate are used. These criteria is already used by (Azzam et al., 1998), (Baldwin, 1997) and others.

There are many factors that affect the accuracy of the result such as the inefficient scope of search for antecedent (the correct antecedent exists outside search scope), errors related to POS tagging, the size of the corpora, and others.

8 CONCLUSION

We presented in this paper, a general survey about Anaphora resolution, especially, Anaphora resolution in Arabic texts. As we have seen in the paper, how the Arabic language differs from other languages because of its complex structure and the fact that it's a rich languages. The major motivation for this work is the lack of Arabic NLP works especially in anaphora resolution and corpora annotation.

Also, a suggested work is presented in this paper, which is Anaphora resolution system using Qur'an text as corpus. The algorithm suggested follows a statistical approach to Anaphora Resolution because of its advantages over the traditional approach mention before. The proposed work is still in progress.

ACKNOWLEDGEMENTS

We would like to express the gratitude to Eng.Mohamed Elarnaoty, a Teaching Assistant at Faculty of Computers and Information, Cairo University for his help in preparing and reviewing the paper.

REFERENCES

- Abadi, D., 2003. "Comparing Domain-Specific and Non-Domain-Specific Anaphora Resolution Techniques". Master Thesis. MPhil, Computer Speech, Text, and Internet Technology, Churchill College, University of Cambridge
- Azzam, S., Humphreys, K., Gaizauskas, R., 1998. "Coreference resolution in a multilingual information extraction". *Proceedings of the Workshop on Linguistic Coreference*. Granada, Spain.
- Baker, K., Brunner, A., Mitamura, T., Nyberg, E., Svoboda, D., Torrejon, E., 2002. "Pronominal Anaphora Resolution in the KANTOO Multilingual Machine Translation System", Language Technologies Institute, Carnegie Mellon University
- Baldwin, B., 1997. "CogNIAC: high precision coreference with limited knowledge and linguistic resources". *Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution*, Madrid, Spain, pp. 38-45
- Barbu C., Mitkov R., 2001. "Evaluation tool for rule-based anaphora resolution methods". *ACL '01 Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics, Stroudsburg, USA, pp. 34-41
- Carbonell, Jaime G., Brown, Ralf D., 1998. "Anaphora Resolution: A Multi-Strategy Approach". *COLING '88 Proceedings of the 12th conference on Computational linguistics*, Association for Computational Linguistics, (Volume 1), pp. 96-101
- Carter, David M., 1986. "A shallow processing approach to anaphor resolution". PhD thesis. University of Cambridge
- Carter, David M., 1987. "A shallow processing approach to anaphor resolution". Computer Laboratory, University of Cambridge
- D. Liddy, E., H. Liddy, J., 2001. "An NLP Approach for Improving Access to Statistical Information for the Masses". Center for Natural Language Processing, School of Information Studies, Syracuse University
- Elghamry, K., Al-Sabbagh, R., El-Zeiny, N., 2007. "Arabic Anaphora Resolution Using Web as Corpus", *Proceedings of the seventh conference on language engineering*, Cairo, Egypt
- Fass, D., Wilks, Y., 1983. "Preference Semantics, Ill-Formedness, and Metaphor". *MIT press*, Computational Linguistics - Special issue on ill-formed input (Volume 9), Cambridge, MA, USA, pp. 178-187
- Halliday, M., Hasan, R., 1976. *Cohesion in English*, Longman
- Hammami, S., Belguith, L., and Ben Hamadou, A., 2008. "Anaphora Resolution in Arabic Language: Developing a Corpora Annotating Tool For Anaphoric Links", LARIS-MIRACL Laboratory, University of Sfax, Tunisia
- Hammami, S., Belguith, L., and Ben Hamadou, A., 2009. "Arabic Anaphora Resolution: Corpora Annotation with Coreferential Links". *The International Arab Journal of Information Technology*, (volume 6), pp. 480-488
- Hirschman L., 1997. MUC-7 Coreference Task Definition, Longman

- Iida, R., Inui, K., Kobayashi, N., Matsumoto, Y., 2005. "Opinion Extraction Using a Learning-Based Anaphora Resolution Technique" *In The Second International Joint Conference on Natural Language Processing (IJCNLP)*, Japan, pp. 175-180
- Mitkov, R., Belguith, L., 1998. "Multilingual robust anaphora resolution", *In Proceedings of the 3rd Conference on Empirical Methods in Natural Language Processing*, Granada, Spain, pp. 7-16
- Mitkov, R., Evans, R., Orasan, C., Ha L.A., and Pekar, V., 2007. "Anaphora Resolution: To What Extent Does It Help NLP Applications?", *DAARC'07 Proceedings of the 6th discourse anaphora and anaphor resolution conference on Anaphora: analysis, algorithms and applications*, Springer-Verlag Berlin, Heidelberg, pp. 179 – 190
- Mitkov, R., Stys, M., 1997. "Robust reference resolution with limited knowledge: high precision genre-specific approach for English and Polish", *In Recent Advances in Natural Language Processing (RANLP-97)*, pp. 74-81
- Mitkov, R., 1994a. "An Integrated Model for Anaphora Resolution". *COLING '94 Proceedings of the 15th conference on Computational linguistics*, (Volume 2), Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 1170-1176
- Mitkov, R., 1995b. "An uncertainty reasoning approach for anaphora resolution". *Proceedings of the Natural Language Processing Pacific Rim Symposium (NLP'95)*, Seoul, Korea, pp. 149- 154
- Mitkov, R., 1996a. "Anaphora resolution: a combination of linguistic and statistical approaches". *Proceedings of the Discourse Anaphora and Anaphor Resolution (DAARC'96)*, Lancaster, UK.
- Mitkov, R., 1997. "Factors in anaphora resolution: they are not the only things that matter. A case study based on two different approaches", *In Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution*, Madrid, Spain, pp. 14-21
- Mitkov, R., 1997. "Two engines are better than one: Generating more power and confidence in the search for the antecedent". *In Recent Advances in Natural Language Processing: Selected Papers from RANLP'95*, pp. 225-234
- Mitkov, R., 1998. "Robust pronoun resolution with limited knowledge". *COLING '98 Proceedings of the 17th international conference on Computational linguistics*, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 869-875
- Mitkov, R., 1998a. "Evaluating anaphora resolution approaches". *In Proceedings of the Discourse Anaphora and Anaphora Resolution Colloquium (DAARC'2)*
- Mitkov, R., 1999. "Anaphora Resolution: The state of the art", *Technical Report based on COLING'98 and ACL'98 Tutorial on Anaphora Resolution*, School of Languages and European Studies, University of Wolverhampton
- Mitkov, R., 2001. "Outstanding Issues in Anaphora Resolution", *proceeding of CICLing '01 Proceedings of the Second International Conference on Computational Linguistics and Intelligent Text Processing*, Springer-Verlag London, UK, pp. 110 - 125
- Nivre, J., 2002. "On Statistical Methods in Natural Language Processing". *Promote IT. Second Conference for the Promotion of Research in IT at New Universities and University Colleges in Sweden*. University of Skövde, pp. 684-694.
- Oltmans, J. A. E., 2000. "A Knowledge-Based Approach to Robust Parsing". PhD thesis, University of Twente, The Netherlands
- Qinan, H., 2008. "A Corpus-Based Study on Zero Anaphora Resolution in Chinese Discourse", *International Conference on Arabic Language Technology*, Department of Chinese, Translation and Linguistics, University of Hong Kong
- Poesio, M., 2004. "The MATE/GNOME Scheme for Anaphoric Annotation," *in Proceedings of SIGDIAL, Boston*, pp. 168-175
- Salmon-Alt, S., Bick, E., Romary, L., and Pierrel, M., 2004. "La FREEBANK: Vers Une Base Libre De Corpus Annotés," *in Processions of TALN, Morocco*, pp. 19-21
- Shalaan, Kh., 2010. "Rule-based Approach in Arabic Natural Language Processing". *International Journal on Information and Communication Technologies (IJICT)*, (Volume 3), New Delhi, India, pp. 11-19
- Sharp, C. R., Kwon, S., Mitkov, R., 1995. "Anaphora Resolution in Machine Translation". *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation TMI'95*, Leuven, Belgium, pp. 87-95
- Sidner, L., 1979. "Towards a computational theory of definite anaphora comprehension in English discourse". Technical Report No. 537. M.I.T., Artificial Intelligence Laboratory.
- Stuckardt, R., 1996. "Anaphor Resolution and the Scope of Syntactic Constraints". *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 932-937
- Tutin, A., Trouilleux, F., Clouzot, C., Gaussier, E., Zaenen, A., Rayot, S., and Antoniadis G., 2000. "Annotating a Large Corpus with Anaphoric Links," *in Proceedings of the Discourse Anaphora and Reference Resolution Conference*, UK, pp. 134-137
- Wilks, Y., 1973. "Preference semantics". Stanford AI Laboratory memo AIM-206. Stanford University.
- Wilks, Y., 1975. "Preference semantics". The formal semantics of natural language ed. by E. Keenan, Cambridge University Press
- Wilks Y., 1985. "Right Attachment and Preference Semantics". *EACL '85 Proceedings of the second conference on European chapter of the Association for Computational Linguistics*, Stroudsburg, PA, USA, pp. 89-92