International Journal of Engineering & Technology, 7 (3.4) (2018) 204-208

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Morphological Analysis by Surface Patterns and by Graph

Iazzi Said[1]*, Yousfi Abdellah[2], Bellafkih Mostafa[3], Aboutajdine Driss[4]

[1,4]LRIT Associated Unit to the CNRST-URAC No29, Faculty of Sciences, Mohammed V University, Rabat, Morocco
[2]Team ERADIASS, FSJES, Rabat, Morocco
[3]Dept. Telecommunications Systems, Networks and Services, STRS Laboratory,
National Institute of Posts and Telecommunications, Rabat, Morocco
**Corresponding Author E-mail address: iazzisaid@yahoo.fr*

Abstract

In this article, we propose a comparison between our two morphological analyzers, which we have developed in recent years. The first is based on surface patterns Arabic words, the second is an analyzer which combines Buckwalter approach and the approach of morphological analysis in base graph. The comparison is made on a corpus of 1400 Arabic words that generalize all cases of Arabic derived words. The results obtained show the interest and the advantages of each analyzer.

Keywords: *Derived words, surface patterns, Morphological Analysis, Viterbi Algorithm, IBN-GINNY Analyzer I.*

1. Introduction

The morphological analysis of words is a very important step in the field of automatic language processing. Several works in this axis have been developed in recent years, and are generally based on one of the following approaches: the two-level approach, the concatenation approach, the finite state automaton approach, the rule-based approach, or approaches that combine these different approaches.
For the Arabic language, several analyzers have been developed, we mention for example that of: Beesly (Beesley, 1996), Buckwalter (Buckwalter, 2004), Hegazi et Elsharkawi (Hegazi et Elsharkawi, 1986), Al-Fedagi et Al-Anzi (Al-Fedagi et Al-Anzi, 1989), Al-Shalabi (Al-Shalabi et al., 2003), Sabri et yousfi (Sabri et yousfi, 2006), Soudi (Soudi, 2001, 2007), attia (Attia, 2006).

2. Page Layout Presentation of Our Morphological Analysis Based on Surface Patterns

This analyzer is based on surface patterns of Arabic words (Said and Yousfi, 2013), it is based mainly on the construction of the database of surface patterns. This morphological analyzer determines one or more possible patterns of a given word to find all the possible analyzes of the word. Patterns can effectively model morphological variations within words and detect the root of a word. Different works have been developed and use the root-schemas approach, among which we mention: (Darwish, 2003) et (Khoja, 1999), (Al-Fedaghi et Al-Anzi, 1989), (Hegazi.N, El-Sharkawi.A., 1986), (Xerox, 1998), (AlKhalil, 2010), (Sadik Bessou et Mohamed Touahria; 2011). All these works use the classical patterns of Arabic words; in our analyzer we use another adapted pattern that we have called the surface pattern.
Example: The classic pattern of the word (جادوا) is (فعلوا), but its surface pattern is (فالوا).

The algorithm of construction of surface patterns from a word :
Either a word $w = l_1 l_2 ... l_n$ (l_i Character of the word) and R its root.

The surface patterns of w is $p = f_1 f_2 ..... f_n$ with:

$$\begin{cases} f_i \quad est l'une des trois lettres ""ف،ع،ل \; il_i \in R \; et \; l_i \notin \{ا،و،ي\} \\ f_i = l_i \quad si \; l_i \; n'est \; pas \; dans R \; ou \; l_i \in \{ا،و،ي\} \end{cases}$$

The surface pattern of the word "قائلون" of the root "قال" is "فائلون".

The surface pattern of the word "نقول" is:

- "نفول" from the root "قال".
- "فعول" from the root "نقل".

The surface pattern of the root $R = g_1 g_2 ... g_k$ (gi is a character) is $P' = f'_1 f'_2 ... f'_k$ with:

$$\begin{cases} f' = l' \quad une des trois lettre "ف،ع،ل \; sig_i \; est une lettre non variante dans R \\ f'_i = g_i \quad sinon \end{cases}$$

Note: a non-variant letter in a root R, is a letter that remains the same when generating words from this root.
To perform the morphological analysis of a word w by the approach of surface patterns, it involves the following steps:

- Search for the surface patterns of the word w by applying the following function:

$$f(m; w) = \sum_{i=1}^{N} 1_{[m_i ; w_i]}$$

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

And we keep the surface patterns with $f \neq 0$.

- Extraction of root surface patterns from the surface patterns of the word w.
- Construction from the surface patterns of the roots, the roots associated with the word w, and verification is what these roots exist in the root base or not.

Example:
Either the word W = "نقول",
- The search for the surface patterns corresponding to W, one finds: P1 = "نفول"; p2 = "فعول".
- Extraction of the root surface patterns of P1 and P2, we find SR1 = "فال" and SR2 = "فعل".
- The construction of roots from P1, P2, SR1, SR2 and W, we find the roots R1 = "قال" and R2 = "نقل".

To test our approach, we have constructed all the surface patterns of Arabic derived words (nouns and verbs), this step was carried out by linguists, and they used a set of Arabic references (Mustapha, 1999 ; Bahrak, ; Hanafi et al., 1914 ; Zanjani, 1343).

Table 1: Extracted from the basis of the surface pattern of Arabic derived words.

Pattern of the Radical	Nature and Number of the Pronoun	Type Names	Surface Pattern of Derived Names
افتعل	جمع-مذكر	اسم-الفاعل	مفتَعلونَ
فعى	مفرد-مذكر	اسم-الفاعل	فَاعٍ
فال	مفرد-مذكر	اسم-الفاعل	فَائع
	مفرد-مذكر	اسم-الفاعل	مُفاوع
وعى	مفرد-مذكر	اسم- الفاعل	وَافٍ
	مفرد-مذكر	اسم-المفعول	مَوْفِيّ
	مثنى-مؤنث	اسم-المفعول	مُسْتَفاعَتَانِ
	مفرد-مذكر	المصدر	مأفعاً
	مفرد-مذكر	اسم-المفعول	مَفِيع
	مفرد-مذكر	الصفة المشبهة	فَائع

3. Presentation of the Graph-Based Analyzer IBN-GINNY I

The IBN-GINNY I analyzer (Yousfi, 2013) is a morphological analyzer based on graphs, in which a word in the Arabic language is represented by a path in this graph. To do the analysis of a word, the IBN-GINNY I analyzer goes through the following two steps:
- The construction of the global network of all Arabic words.
- Research in this global network, possible solutions for the analysis of a given word.

This system is based on very restricted dictionaries and seeks solutions in the global network using the Viterbi algorithm, and every word is modeled by a path, whose radical letters are presented by a state that loops on itself, and the affixes are presented by the characters forming these affixes.

Example:
The words 'فجامعها' ,'فداخلها' ..., are presented by the following path:

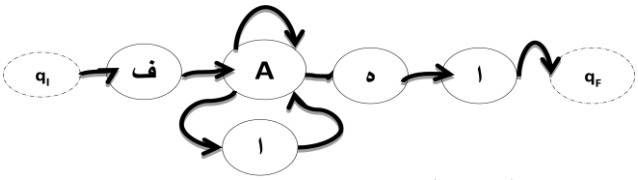


Fig. 1: The graph of the words 'فجامعها' ,'فداخلها'.

Based on all the prefixes, the infixes and suffixes of the Arabic language, the global state network is constructed with a single input state = q_I and a single output state = q_F.
Our global network is defined entirely by:

- The set Q of all the states, it consists of all the characters composing the affixes (suffixes, prefixes and infixes), of the state A, the initial state q_I and the final state q_F:

Q={q_I,q_F, A,"ف","و","ي","ل",…,"ه","م","ت",…}

- The set of all possible transitions connecting the characters of suffixes, prefixes and infixes with states A, q_I and q_F.

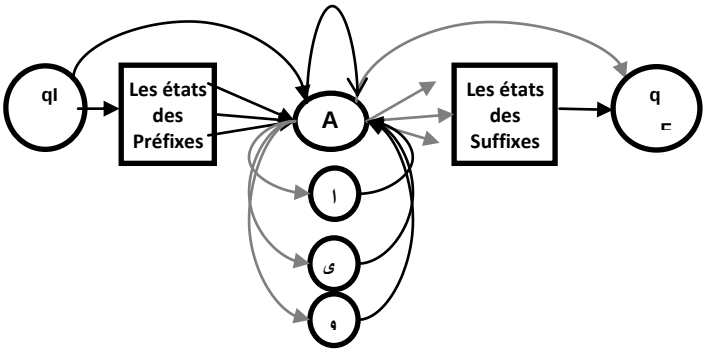


Fig. 2: Diagram of our global network

Searching for possible paths of a word to analyze:
To analyze a given word W, we look in the global network for the different possible paths associated with W. All these paths are given by:

$$S = \{ \xi \in B / P_r(w/\xi) \neq 0 \} \quad (1)$$

B: the set of all possible paths in our network and are the same length as w.
Solutions are all the paths that make it possible to send the word with a non-zero probability. To facilitate and reduce the calculation in formula (1), we adapted the Viterbi algorithm in the following format:

$$\delta_t(c_j) = \underset{c_i \to c_j}{\mathrm{NL}} (\delta_{t-1}(c_i) \cdot a_{ij} \cdot 1_{c_j}(w_t))$$

NL (x) is the non-zero value of x. We are looking for C_i states that give non-zero values $\delta_{t-1}(c_i) \cdot a_{ij} \cdot 1_{c_j}(w_t)$.

$\delta_T(q_F)$: is the maximum probability of sending the word w from a given path. By a recursive calculation we recover all the possible paths that give non-zero values (T the length of the word w).
With:

$c_j \in Q$

a_{ij} : The transition probability from state C_i to state C_j, with:

$$a_{ij} = \begin{cases} 1 & \text{si la transition est possible.} \\ 0 & \text{sinon.} \end{cases}$$

w_i: t^{th} character of the word w.

$$1_{c_j}(w_t) = \begin{cases} 1 & si \quad C_j = w_t \\ 0 & \text{sinon} \end{cases}$$

We take: $1_A(w_t) = 1$

Initialization :

$$\delta_0(c_i) = \begin{cases} 1 & si\ c_i = q_I \\ 0 & \text{sinon} \end{cases}$$

To evaluate our approach, we first created the different dictionaries of suffixes, prefixes, and radicals. Then from the list of prefixes, suffixes and infixes, we have generated a global network of states as previously reported without the use of lexical dictionaries (this is the great advantage of our analyzer over the Buckwalter analyzer and the state automata based analyzer).

to construct the dictionary of radicals, we make the difference between two types of roots: healthy roots and defective roots (معتلة), the presentation of healthy roots in our dictionary is done by the roots themselves, but for the defective verbs the radicals of Buckwalter are kept [5]. The root "قال" is presented by the 5 radicals in our dictionary: The root "قال" is presented by the 5 radicals in our dictionary: قائل, قيل, قول, قل, قال.

But for healthy root "كتب" they are presented only by "كتب".

With this presentation, was decreased the dictionary size radicals of 62.5% compared to Buckwalter [5].

Example: For the word "فيسكتهم", the possible paths on the global network are:

Table 2: Les chemins possibles du mot "**فيسكتهم**" dans le réseau global avec leur racines proposées

Suffixe	Préfixe	Chemins Possibles	Racines Proposes	N
تهم	0	AAAA ت ه م	فيسك	1
هم	0	AAAAA ه م	فيسكت	2
0	0	AAAAAAA	فيسكتهم	3
تهم	ف	فـAAA ت ه م	يسك	4
هم	ف	فـAAAA ه م	يسكت	5
0	ف	فـAAAAA	يسكتهم	6
تهم	في	ف يAA ت ه م	سك	7
هم	في	ف يAAA ه م	سكت	8
0	في	ف ي AAAA	سكتهم	9

After verification, only the solution is left: "هم" "سكت" "في".

4. Comparison Between the Two Analyzers

Our personal contribution is the development of two approaches to the morphological analysis of the Arabic language. The first is based on the surface pattern of all Arabic derived words, and the second approach, appeals to the graphs for the morphological analysis of a given word. In this article, we propose a comparison between these two approaches. Before presenting the comparison in terms of precision and recall, we present the advantages and disadvantages of each approach.

Among the advantages of these two approaches, we cite as an example:

- Both approaches have the characteristics of reducing the size of the dictionaries used in the other analyzers: The first approach uses the basic surface patterns in addition to the root base to model all morphological variations of words derived. As for the second approach, it is articulated only on the basis of roots.
- In the second approach we generated a global network of states without the use of lexical dictionaries (This is the great advantage of our analyzer over the Buckwalter analyzer and the finite state automaton analyzer).
- The graph-based approach gives all the morphological analyses of a given word, but the approach based patterns, gives only the analyses associated with existing surface patterns in the database patterns.
- The second approach uses no linguistic knowledge base to make the morphological analysis of a word.
- The coverage rate of these two approaches for Arabic derived words is very high compared to other analyzers.

Unfortunately, both approaches have some disadvantages:

- The patterns-based approach processes only derivative words, while the graph-based approach can be used even for non-derived words.
- The second approach does not take into account all the rules of compatibility between the affixes (infixes, prefixes, and suffixes) and the roots. As a result, this approach gives errors because of this negligence.

4.1. Evaluation of the Patterns Based Approach

For the analyzer based on surface patterns, we used:

- A lexicon of 6216 surface patterns. This lexicon contains all morphological classes of Arabic derived words.
- A dictionary of roots, which contains 1200 roots.
- A dictionary of stems, which contains 6000 stems.

The evaluation is carried out on two corpora, the first contains 10063 derived words, and it contains almost all categories of derivative words, and the second contains 4600 derived words and it is used to compare between the two approaches. The following table represents the error rate, the analysis based on surface patterns for each category of derivative words.

Table 3: Represents the error rate

Derived Word Type	Number of Words	Error Rate
فعل	2964	7,02%
اسم-الفاعل	1603	9,11%
اسم-المفعول	1661	3,91%
اسم-الآلة	348	2,59%
اسم-التفضيل	250	4,00%
اسم-الزمان	700	0,86%
اسم-المكان	715	1,26%
التصغير	73	8,22%
المرة	210	0,95%
المصدر	315	2,86%
المصدر-الصناعي	109	1,83%
المصدر-الميمي	641	0,31%
الهيئة	191	1,05%
صيغة-المبالغة	237	5,49%
الصفة-المشبهة	46	0,00%
Total	10063	4,86%

In general, we note that the analysis error rate, for each type of derivative words, is close to 4.86%. The lowest rate is obtained for the type "الصفة-المشبهة" with an error rate of 0%, while the highest rates are that of "اسم-الفاعل" and "التصغير" with a percentage of 9 , 11% and 8.22%. The average error for all of these types, it is of the order of 4.86%. Most of these errors come mainly from the fact that compatibility rules between affixes and roots are not taken into account.

4.2. Comparison between the Two Approaches:

To compare between the performances of each approach, we used the following body:

The test set, it contains 4000 Arabic derived words, and they represent almost all cases of Arabic derived words. It is used to evaluate both analyzers.

For the comparison between these two analyzers, we used the following indicators: precision, recall and execution time. The results found are shown in the following table:

Table 4: Comparison between the two approaches

Analyzers	Precision	Recall	Execution Time
A basic patterns	97.08%	94,20%	24 ms
Based on Graph	95.97%	98.58%	14 ms

The analyzer based on surface patterns gives an accuracy of 97%; the remaining 3% comes from the analysis errors of this analyzer. These errors come mainly from not taking into consideration the rules of correspondence between prefixes, suffixes and radicals. For the recall, our system was able to return 94.20% of the correct analyzes, among all possible analyzes, the remaining 5.8% of the recall is due to the inadequacy of the lexicon of surface patterns and roots that do not cover 100% of the 4000 words.

The graph-based analyzer gives an accuracy of 95.97%, and the remaining 4.03% represents the number of false analyzes among returned analyzes. The recall, it is of the order of 98.85% that is to say that this analyzer represents a deficiency of 0.15% in the lexicon of the radicals used to analyze the 4000 words.

For the comparison between these two analyzers, the surface pattern analyzer returns more valid solutions than the graph based one. While the graph-based analyzer returns more analysis in a time almost by half compared to that at base surface patterns.

5. Conclusion:

In this paper, we proposed a comparison between the two morphological analyzers based surface patterns and basic graph. This comparison allowed us to detect the strengths and weaknesses of each analyzer. Besides, it will be very interesting to combine these two approaches, in a single analyzer to increase the precision and recall the same time, keeping an execution time very close to that of the second analyzer.

References

[1] Al Fedaghi.S and Al-Anzi, 1989 : A new application to generate Arabic Root-Pattern Forms, Procedings of the 11th National Computer Conference and Exhibition, March, Dahran, Saudia Arabia, 391-400.

[2] Al-Hamlawy A. 1957. Shaza Al-Orf in the art of morphology. (by) Dar Al-Kiaan, Riadh, KSA, (Arabic book).

[3] ALESCO, "Arabic Language Derivation and morphological System," Published by the Arab League Educational, Cultural and Scientific Organization, http://www.reefnet.gov.sy/ed4-2.htm, Last Visited 2007.

[4] Al-Ghalayyni. 2005. جامع الدروس العربية"Jami' Al-Duroos Al-Arabia". Saida - Lebanon: Al-Maktaba Al-Asriyiah "المكتبة العصرية".

[5] Al-Rajhi A. 1979. The application of morphology. (by) Dar Al-Nahdha Al-Arabia Beirut, (Arabic book).

[6] Al-Sughaiyer, I. A. and Al-Kharashi, I. A. 2004. Arabic morphological analysis techniques: A comprehensive survey. Journal of the American Society for Information Science and Technology 55(3): 189-213.

[7] Al-Shalabi et al., 2003 : New approach for extracting Arabic roots. Paper presented at the International Arab Conference on Information Technology (ACIT'2003), Alexandria, Egypt.

[8] M.Al- Ǧalāyīnī, 2000; جامع الدروس العربية, المكتبة العصرية بيروت, صيدا لبنان, 2000.

[9] Alexia Blanchard. Analyse morphologique des réponses d'apprenants en environnement d'Apprentissage Assisté par Ordinateur. Université Stendhal-Grenoble III,UFR des Sciences du Langage.

[10] Attia, M. 2006 : An Ambiguity-Controlled Morphological Analyzer for Modern Standard Arabic Modelling Finite State Networks. The Challenge of Arabic for NLP/MT Conference, the British Computer Society, London.

[11] Audebert C, Jaccarini A. 1988. De la reconnaissance des mots outils et des tokens. Annales islamologiques 24, Institut francais d'archeologie orientale du Caire.

[12] Awajan.A, 2011: "Multilayer Model for Arabic Text Compression", The International Arab Journal of Information Technology, Vol. 8, No. 2, April 2011

[13] Bahrak. (-869) , بحرق، جمال الدين محمد بن عمر بن مبارك الحميري الحضرمي فتح الأقفال وحل الإشكال بشرح لامية الافعال. 930هـ

[14] Beesly.KR 1998. Arabic Morphology Using Only Finate-State Operations, Proceedings of the Workshop on Computational Approaches to Semetic languages. Montreal, Quebec, pp 50-57.

[15] Beesley KR 1996. Arabic Finite-State Morphological Analysis and Generation. Proceedings of the 16th conference on Computational linguistics, Vol1. Copenhagen,Denmark: Association for Computational Linguistics, pp 89-94.

[16] Beesley, Kenneth, Tim Buckwalter, and Stuart Newton, "Two-Level Finite-State Analysis of Arabic Morphology." Proceedings of the Seminar on Bilingual Computing in Arabic and English, Cambridge, England, 1989.

[17] Beesley, Karttunen; Finite-State Non-Concatenative Morphotactics 2000.

[18] Beesley, 2001, Finite-State Morphological Analysis and Generation of Arabic at Xerox Research: Status and Plans - Beesley – 2001.

[19] Beesley, Karttunen ; Finite-State Morphology: Xerox Tools And Techniques - (Show Context) 2003.

[20] Berri,j, Zidoum,H, & Atif, Y: Web-based Arabic Morphological analyser. In Gelbukh, A (Ed.): CICLing 2001, LNGS 2004, pp. 216-225. Springer-Verlag Berlin Heidelberg.

[21] Boudlal, A., Lakhouaja, A., Mazroui, A., Meziane, A., Bebah, M. O. A. O. and M.Shoul. 2010. Alkhalil Morpho Sys: A Morphosyntactic analysis system for Arabic texts. IJCSI International Journal of Computer Science Issues.

[22] Buckwalter.T 2002. Buckwalter Arabic Morphological Analyzer. Version 1.0. Linguistic Data Consrtium, catalog. Number LDC2002L49 and ISBN 1-58563-257-0.

[23] Buckwalter, T. 2004. Buckwalter Arabic Morphological Analyzer Version 2.0. Linguistic Data Consortium, catalog number LDC2004L02 and ISBN 1-58563-324-0.

[24] Chomsky, Noam (1957). Syntactic Structures. The Hague: Mouton and Co.

[25] Cohen.D, 1961/1970: Essai d'une analyse automatique de l'arabe. Dans: David Cohen. Etudes de linguistique sémitique et arabe. Paris:Mouton, p. 49-78, (1970).

[26] Dahdah, A. 1987. A Dictionary of Arabic Grammer in Charts and Tables "معجم قواعد اللغة العربية العربية –في جداول ولوحات". Beirut, Lebanon: Librairie du Liban publisher.

[27] Dahdah, A. 1993. A dictionary of Arabic Grammatical nomenclature Arabic – English "معجم لغة النحو العربي عربي-انكليزي". Beirut, Lebanon: Librairie du Liban publishers.

[28] Darwish K. (2002). Building a Shallow Morphological Analyzer in One Day. Proceedingsof the workshop on Computational Approaches to Semitic Languages in the 40th Annual Meeting of the Association for Computational Linguistics (ACL-02). Philadelphia, PA, -USA.

[29] Dichy. J, S. Ammar, Les Verbes Arabes, Bescherelle, Paris, 1999.

[30] Dukes, K. and Habash, N. 2010. Morphological Annotation of Quranic Arabic. Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta,19-21 May 2010.: European Language Resources Association (ELRA).

[31] El-Sadany.T.A and Hashish.M.A 1989. An Arabic Morphological System. IBM Systems Journal. Vol.28, No.4, 600-612.

[32] Farghaly. A, K. Shaalan. Arabic Natural Language Processing: Challenges and Solutions, ACM Transactions on Asian Language Information Processing (TALIP), the Association for Computing Machinery (ACM). TALIP Vol 8, Issue 4, December 2009.

[33] Gaubert C. Analyse morphologique d'un texte par ordinateur – Résultats et évaluation. AnIsl 29 (1996), IFAO, p. 283-311

[34] Goldsmith and John.A (2001). Unsupervised learning of the morphology of a natural language. Computational Linguistics, 27(2), 153-198.

[35] Habash, Nizar and Owen Rambow. MAGEAD: A Morphological Analyzer and Generator for the Arabic Dialects. In Proceedings of ACL, Sydney, Australia, 2006.

[36] Habert et al., 1997 Les linguistiques de corpus. U Linguistique. Paris: Armand Colin/Masson.

[37] Hamada, S. 2009b. "المحللات الصرفية للغة العربية" proposal for evaluating morphological analyzers for Arabic text. Proceedings of the workshop of morphological analyzer experts for Arabic language, organized by Arab League Educational, Cultural and Scientific Organization ALECSO, King Abdul-Aziz City of Science and Technology (KACST) and Arabic Language Academy., Damascus, Syria. 26-28 April 2009.

[38] Hamada, S. 2010. Evaluation of the Arabic Morphological Analyzers Proceedings of The Sixth International Computing science Conference ICCA, Hammamet, Tunisia.

[39] Hanafi. (1914). حنفي بك ناصف-محمد بك دياب-مصطفي طموم-محمود افندي طبعه مصر سنه .قواعد اللغه العربيه لتلاميذ المدارس الثانويه ,عمر-سلطان بك محمد 1914 اديان. علوم الدين

[40] Hegazi.N and ElSharkawi.A 1986 : Natural Arabic Language Processing, Proceedings of the 9th National Computer Conference and Exhibition, Riyadh, Saudi Arabia, 1-17.

[41] Hilal, Yahiah, 1985 : Morphological analysis of Arabic speech. In: Computer processing of the Arabic language.

[42] Iazzi, S, Yousfi, A, Bellafkih, M. Analyseur morphologique des mots arabe en utilisant le dérivé et le schème de surface. 5ème Edition de la conférence internationale sur les Technologies d'Information et de Communication pour l'Amazighe (TICAM 2012) 26 - 27 novembre 2012, Institut Royal de la Culture Amazighe (IRCAM), morocco.

[43] Iazzi, S, Yousfi, A, Bellafkih, M, Aboutajdine, D. Morphological Analyzer of Arabic Words Using the Surface Pattern. IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 1, March 2013, ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784.

[44] Iazzi, S, Yousfi, A, Bellafkih, M, Aboutajdine, D. Analyse morphologique à base de graphe. 8TH international conference on intelligent systems: theories and applications, 08-09 may 2013, rabat, morocco. SITA'13.

[45] Iazzi, S, Yousfi, A, Bellafkih, M, Aboutajdine, D. Graph-based morphological analysis. Journal of Computer Science and Engineering Volume 19, Issue 2 June 2013.

[46] Jaccarini A., (1997). Grammaires modulaires de l'arabe. These de doctorat. Universite de Paris-Sorbonne.

[47] Kadri, Y., Benyamina, A.: 1992, Un système d'analyse syntaxico-sémantique du langage arabe non voyellé", Mémoire d'ingénieur, Université d'Oran.

[48] Khoja.S and Garside.R (1999). Stemming Arabic text. Computer Science Departement, Lancaster University, Lancaster, UK.

[49] Kiraz, G.A., Computational Nonlinear Morphology with Emphasis on Sematic Languages. Studies in Natural Language Processing, ed. B. Boguraev and S. Bird. 2001, Cambridge: Cambridge Univeristy Press.

[50] Koskenniemi and Kimmo (1983). Two Level Morpology. A General Computational Model for Word-form Recognition and Production. Publication No. 11, Dep. of General Linguistics, University of Helsinki, Helsinki.

[51] Mesfar.S, (2006), Analyse lexicale et morphologique de l'arabe standard utilisant la plateforme linguistique NooJ, Actes de la 13ème Conférence sur le Traitement Automatique des Langues Naturelles RECITAL/ TALN2006, Presses universitaires de Louvain, Louvain-la-Neuve, Belgique, Avril 2006

[52] Mustapha.G (1999). .المكتبة ,الشيخ مصطفى .العربية الدروس جامع الغلاييني العصرية ،

[53] Ouersighni, R. (2001) A major offshoot of the DIINAR-MBC project: AraParse, a morpho-syntactic analyzer of unvowelled Arabic texts, In ACL 39th Annual Meeting. Workshop on Arabic Language Processing: Status and Prospect, Toulouse, pp. 66-72.

[54] Ryding, K. C. 2005. A Reference Grammar of Modern Standard Arabic. Cambridge University Press.

[55] Rafea.A and Aisha Rafea, 1884 : 'Understanding an Arabic word in a text automatically', Nineteenth Annual Conference on Statistics, Computer Science and Operations Research, Cairo, Egypt, 1984, pp. 52–77.

[56] Sabri, S., Yousfi, A., Bouyakhf, E. (2006b) : Système d'analyse morphologique des noms Arabes. MCSEAI'06 December 07-09, 2006, Agadir.

[57] Saliba.B and Al-Dannan.A (1989). Automatic Morphological Analysis of Arabic: A Study of Content Word Analysis. Proceedings of the First Kuwait Computer Conference, Kuwait, March, 3-5.

[58] Sam.A et Youssef D (1999). سام عمار، يوسف ديشي، مجموعة العربية الأفعال هاتيبه باريس، .بيشرال

[59] Selim. S, M. Gheith and M. Mashhour, 'A general morphological analyzer', 20th Annual Conference on Statistics, Computer Science and Operations Research, Cairo, Egypt, 1985, pp. 21–42.

[60] Smrz, O. 2009 : ElixirFM Functional Arabic Morphology: Case Studies. Proceedings of the workshop of morphological analyzer experts for Arabic language, organized by Arab League Educational, Cultural and Scientific Organization (ALECSO), King Abdul-Aziz City of Science and Technology (KACST) and Arabic Language Academy., Damascus, Syria.26-28 April 2009.

[61] Soudi, A., Violetta Cavalli-Sforza (2001). A Computational Lexeme-Based Treatment of Arabic Morphology. In Proceedings of the Association for Computational Linguistics, Arabic Processing Workshop, Toulouse, July 2001, France.

[62] Soudi A., Van den Bosch A. and Neumann G (2007). Introductory chapter to the book Arabic Computational Morphology: knowledge-based and Empirical Methods, Editors: Soudi A., Van den Bosch A. and Neumann G. In Kluwer/Springer's series on Text, Speech, and Language Technology (series editors Nancy Ide and Jean Veronis).

[63] Thalouth, B. Al-Dannan, A., 1990 : "A comprehensive Arabic Morphological Analyzer. Trujillo A., Translation Engines: Techniques for Machine Translation, Springer Verlag, 1999.

[64] Wehrli, E. 1997. L'analyse syntaxique des langues naturelles : problèmes et méthodes, Paris,Masson.

[65] Wright, W. 1996. A Grammar of the Arabic Language, Translated from the German of Caspari, and Editted with Numerous Additions and Corrections. Beirut: Librairie du Liban.

[66] William B. Frakes and Ricardo A. Baeza-Yates, editors. 1992. Information Retrieval: Data Structures ~ Algorithms. Prentice-Hall.

[67] Yousfi.A (2010). The morphological analysis of Arabic verbs by using the surface patterns. IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 3, No 11, May 2010.

[68] Yousfi.A Sabri.S and Bouyakhf.E (2006). Système d'analyse morphologique des noms Arabes. JETALA (Journées d'Etudes sur le Traitement Automatique de la Langue Arabe), 2006, Rabat 5-7 juin, 2006.

[69] Yousfi.A, Iazzi.S : "نحو محلل صرفي عربي يعتمد على أوزان الكلمة". 7th International Computing Conference in Arabic (ICCA'11). Riyadh, Saudi Arabia (May 31- June 2, 2011).

[70] Zanjani. (1343 هـ) الزنجاني . متن البناء و متن التصريف العربي .القاهرة مطبعة مصطفى البابي الحلبي ,هـ 1343