

Tunisian Dialect Agglutination Processing with Finite Transducers

Roua Torjmen¹, Kais Haddar²

¹ University of Sfax,
Faculty of Economics and Management of Sfax,
Tunisia

² University of Sfax,
Faculty of Sciences of Sfax,
Tunisia

rouatorjmen@gmail.com, kais.haddar@yahoo.fr

Abstract. The agglutination is a widespread phenomenon in the formation of words in the Tunisian dialect. So, it requires a treatment at the morphological level. In this context, we present our way of dealing with this phenomenon for the different grammatical categories in Tunisian dialect. The proposed method is based on the construction of morphological grammars using a set of finite state transducers. These transducers offer great flexibility in the construction of morphological grammars and allow their maintenance and reuse. Our goal is to create a set of linguistic resources allowing the treatment of the phenomenon of agglutination in the Tunisian dialect. The NooJ linguistic platform with new technologies makes it possible to elaborate and experiment our constructed resources. The obtained results are ambitious and highlight our proposed method.

Keywords. Agglutination phenomenon, finite transducer, morphological grammar, Tunisian dialect.

1 Introduction

The agglutination is a widespread phenomenon in the Tunisian Dialect (TD). This reason reveals the importance of this linguistic phenomenon and makes its treatment a necessity. The construction of morphological grammars by relying on a set of finite transducers solves the treatment of agglutination and facilitates lexical analysis. Furthermore, finite transducers provide flexibility for maintenance and reuse.

In addition, the treatment of agglutination via a morphological analyzer allows the recognition of TD words and their different parts. Besides, this morphological analyzer can be integrated in many applications such as automatic annotation of TD corpus, POS-Tagging, TD speech synthesis and automatic translation from TD to Modern Standard Arabic (MSA) and vice versa.

The processing of TD raises other issues, such as the lack of standard spelling because it is never taught in educational institutions. In addition, there are dialectal differences from one city to another. This diversity creates the appearance of non-existent letters in MSA and the existence of words of different origins: French, Maltese, Turkish, etc.

Moreover, the lack of linguistic resources in language platforms like NooJ for TD exacerbates the problems of building robust tools and applications. In this paper, our principal objective is to build morphological grammars based on finite transducers that allow the processing of agglutination in TD.

In order to achieve this goal, we carry out a linguistic study on the phenomenon of agglutination. Subsequently, we need to implement a set of linguistic resources using the NooJ linguistic platform. The present paper is divided into six sections.

Table 1. IOC and DOC in Tunisian Dialect

Persons	DOC	IOC
1st singular person	ني 'nii' (me)	لي 'lii' (to me)
1st plural person	نا 'naa' (us)	لنا 'lnaa' (to us)
2nd singular person	ك 'k' (you)	لك 'lik' (to you)
2nd plural person	كم 'kum' (you)	لكم 'lkum' (to you)
3rd masculine singular person	هو / و / ه 'huu' (him/it)	لو 'luu' (to him/it)
3rd feminine singular person	ها 'haa' (her)	لها 'lhaa' (to her)
3rd plural person	هم 'hum' (them)	لهم 'lhum' (to them)

In the second section, we present related work dealing with the morphological analyzer for the Arabic dialects and MSA. In the third section, we exhibit our in-depth linguistic study. In the fourth section, we explain our linguistic resources related to the phenomenon of agglutination by explaining the designed dictionary and grammars.

In the fifth section, we experiment and evaluate our constructed grammars and dictionary. Finally, our paper ends with a conclusion and some perspectives.

2 Related Work

The agglutination treatment is done by the construction of morphological analyzers. In what follows, we introduce some works dedicated to MSA and Arabic dialects. Numerous works deal morphologically with MSA such as the Buckwalter Arabic Morphological Analyzer (BAMA) [1], the Standard Arabic Morphological Analyzer (SAMA) [9], the morpho-syntactic analyzer Alkhalil [8], the tool for morphological analysis and disambiguation MADAMIRA [10] and the Arabic morpho-syntactic analyzer using the NooJ linguistic platform [4, 7].

On the one hand, other works consider the approach that treats Arabic dialects using tools designed for MSA. The work of [12] which leans on SAMA and BAMA analyzers to recognize the prefixes and suffixes of the Egyptian dialect. Besides, the authors [5] have developed an Algerian morphological analyzer based on the BAMA and Al-Khalil.

Moreover, the Analyzer for Dialectal Arabic Morphology (ADAM) [13] allows the morphological processing of three dialects (Egyptian, Levantine and Iraqi). These three dialects have many similar morphological characteristics such the negation verbs, propositions and indirect object complements.

On the other hand, other works have chosen another approach that works directly on the dialect. Among these works, the authors [11] have created an a Moroccan dialect electronic dictionary (MDED) in order to develop a Moroccan morphological analyzer.

Besides, the authors [2] have proposed a machine learning method to extract Egyptian morphological lexicons from morphologically annotated corpora, such as inflection classes and associated lemmas.

Concerning TD, the authors [3] are interested in the creation of a morphological analyzer using the Morphological Analyzer and GEnerator for Arabic Dialects (MAGEAD).

This analyzer only processes verbs. The authors [6] have suggested a TD morphological analyzer using aebWordNet, Tunisian lexical dictionary and twenty two predicate rules. This created system does not deal with the standardized TD.

In addition, the authors [15, 16] have sought to create a morphological analyzer processing TD using the NooJ linguistic platform. In fact, the second approach generally gives better morphological analyzers than the first approach in terms of quality because it requires handwritten rules. Thus, this paper is based on [15, 16].

Table 2. Noun suffix in Tunisian dialect

Persons	Noun suffix
1st singular person	ي 'ii' (my)
1st plural person	نا 'naa' (our)
2nd singular person	ك 'k' (your)
2nd plural person	كم 'kum' (your)
3rd masculine singular person	ه / و / هو 'huu' (his/its)
3rd feminine singular person	ها 'haa' (her)
3rd plural person	هم 'hum' (their)

ال , PREF+NW
 ة , DEM+NW
 فدا , DEM+FLX=DEM
 ع , PREP+NW
 على , PREP
 ك , NSUFF+2+s+NW
 كم , NSUFF+2+p+NW
 مات , V+FLX=VERBE2
 بيشكلة , N+FLX=MFP5
 تجربة , N+FLX=MFP5

Fig. 1. Example of dictionary entries

3 Linguistic Study on Agglutination Phenomenon

TD is an agglutinative dialect. Indeed, the agglutination is the association of several grammatical categories in the same word. An agglutinated word has either proclitics, or enclitics, or both. The proclitics are located before the inflected or canonic form and the enclitics are after. In the following, we list the forms of agglutination at the level of verbs, nouns and particles.

3.1 Agglutinated Verbs

The verb is the most complicated grammatical category in terms of agglutination because it has multiple patterns. Regarding the proclitics of a verb, there is the conjunction (CONJ) which is frequently و 'wa' (and), the adverb of interrogation (INTERR) ش 'ch' (what) and the adverb (ADV) ما 'maa'.

For example, the word وشماخرج 'wachmakharraj' (and what does it take out) gathers all proclitics. Like proclitics, enclitics are found in TD verb. There are several types of enclitics: the adverb of interrogation which concerns yes/no questions (INTERR) شي 'chii', the adverb of negation (NEG) ش 'ch' (not), direct object complements (DOC) and indirect object complements (IOC) as shown in Table 1.

Indeed, negation and interrogation adverbs cannot be together in the same verb. The TD verb structure is defined by the regular expression 1:

CONJ? ADV? Verb DOC? IOC? (NEG|INTERR)? (1)

For example, the word وماوراوهوليش 'wmaawarrawhuuliich' (they did not show it to me) is the longest structure for a Tunisian verb that represents an entire sentence. This word is composed of the conjunction و 'wa' (and), the adverb ما 'maa', the verb وراو 'warraw' (show), the DOC هو 'huu' (it), the IOC لي 'lii' (to me) and finally the negation adverb ش 'ch' (not).

3.2 Agglutinated Nouns

Nouns have many agglutination patterns. Regarding the proclitics, definite nouns are preceded by the definite article (PREF) ال 'il' (the). The prepositions (PREP) that appear before the nouns are as follows:

ب 'b' (by), ل 'l' (to), ك 'k' (as), م 'm' (from) is the abbreviation for من 'min', ع 'a' (on) is the abbreviation for على 'alaa' and ف 'fi' (in) is the abbreviation for في 'fii'.

The combination between the preposition ل 'l' and the definite article ال 'il' produces the proclitic لل 'lil' (to the). The demonstrative pronoun ه 'ha' (this) stands as a proclitic only before definite nouns.

Enclitics are found only for indefinite nouns in the form of an annexation compound. These enclitics are noun suffixes (NSUFF) as shown in Table 2. Formally, The TD definite noun structure is defined by the regular expression 2:

CONJ? PREP? DEM? PREF? Definite_Noun (2)

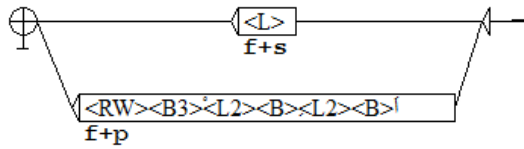


Fig. 2. Example of transducer for nouns

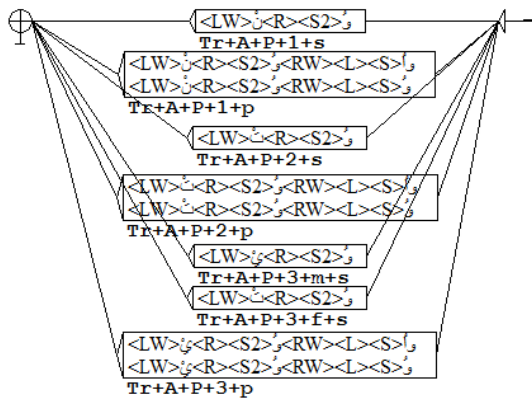


Fig. 3. Extract of transducer for Hollow verbs

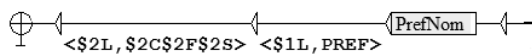


Fig. 4. Transducer for definite noun

In addition, the TD indefinite noun structure is defined by the regular expression 3:

$$\text{CONJ? PREP? Indefinite_NounNSUFF?} \quad (3)$$

For example, the words *وبالمشروع* 'wibhalmachruu'a' (and by this project) and *وبمشروعهم* 'wbimachruu'ahum' (and by their project) are respectively the longest definite noun and indefinite noun structures in TD.

The first word is composed of the conjunction *و* 'wa' (and), the preposition *ب* 'b' (by), the demonstrative pronoun *هـ* 'ha' (this), the definite article *ال* 'il' (the) and the noun *مشروع* 'machruu'a' (project).

The second word is composed of the conjunction *و* 'wa' (and), the preposition *ب* 'b' (by), the

noun *مشروع* 'machruu'a' (project), the noun suffix *هم* 'hum' (their).

3.3 Agglutinated Particles

Particles are composed of several grammatical categories. Most of them have the phenomenon of agglutination. First of all, prepositions also have proclitics and enclitics. The TD preposition structure is defined by the regular expression 4:

$$\text{CONJ? ADV? Preposition NSUFF? (NEG|INTERR)?} \quad (4)$$

For example, the word *ومامعهاش* 'wmaam'ahaach' (And not with her) is the longest structure for a Tunisian preposition. This word is composed of the conjunction *و* 'wa' (and), the adverb *ما* 'maa', the preposition *مع* 'm'a' (with), the noun suffix *ها* 'haa' (her), and finally the negation adverb *ش* 'ch' (not). Furthermore, the TD personal pronoun structure is defined by the regular expression 5:

$$\text{CONJ? ADV? Personal_Pronoun (NEG|INTERR)?} \quad (5)$$

For example, the word *وماهيش* 'wmahich' (She is not) is the longest structure for a Tunisian personal pronoun.

This word is composed of the conjunction *و* 'wa' (and), the adverb *ما* 'maa', the personal pronoun *هي* 'hiya' (she) and finally the negation adverb *ش* 'ch' (not). In addition, the TD demonstrative pronoun structure is defined by the regular expression 6:

$$\text{CONJ? Prep? Demonstrative_Pronoun} \quad (6)$$

For example, the word *وبهذا* 'wbihathaa' (and by this) is the longest structure for a Tunisian demonstrative pronoun. This word is composed of the conjunction *و* 'wa' (and), the preposition *ب* 'b' (by) and the demonstrative pronoun *هذا* 'hathaa' (this).

The conjunction *و* 'wa' (and) marks a great presence in all grammatical categories in TD. In addition, the TD adverb *ما* 'maa' is frequently found in words without or with the form of interrogation or negation.

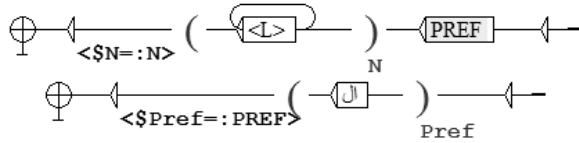


Fig. 5. PrefNom and Pref transducers

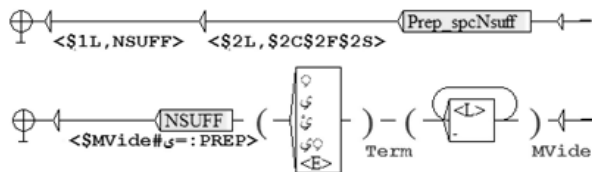


Fig. 6. Transducer for agglutinated preposition

Dictionary has been successfully compiled in file:
 C:\Users\Lenovo\Documents\NooJ\ar\Lexical Analysis\barcha.nod
 (342673/1190 states; 45294 infos; recognizes 169815 forms)

Fig. 7. Dictionary compilation result

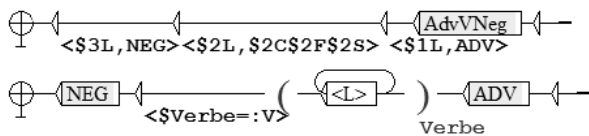


Fig. 8. Transducer for negation form of verbs

3.4 Effect of Agglutination on Certain Letters

The agglutination phenomenon has an effect on some letters at the end of the word. The correlated letter tā *ط* becomes *ت*. For example, the noun *كرهبة* ‘*karhbah*’ (a car) after the agglutination becomes *كرهبتها* ‘*karhbatha*’ (her car). Moreover, the letter shortened alif *ي* ‘*aa*’ also undergoes a transformation and becomes the letter *ي* ‘*yi*’.

For example, the preposition *على* ‘*alaa*’ (on) becomes *عليه* ‘*aliih*’ (on it). Also, the letter hamzah *ء* becomes the letter *yā* hamzah *ئ*. For example, the noun *اصدقاء* ‘*asdikaa*’ (friends) becomes *اصدقائها* ‘*asdikaiha*’ (her friends).

4 Proposed Method

The method we propose to deal with the agglutination phenomenon starts with the automatic extraction of all non-repetitive words from the study corpus. Afterwards, the filtering phase allows to eliminate the same words which are written under different inflected forms.

Then, the choice of a canonical form allows to present the collected words. Finally, this canonical form is enriched by adding morphological, lexical and syntactic features. All these phases are established thanks to the dictionary, inflectional and morphological grammars of the linguistic platform NooJ [14].

4.1 Dictionary and Inflectional Grammars

The dictionary is a set of entries that consist of a canonical form, a lexical category and an inflectional grammar if necessary. Fig. 1 shows an example of dictionary entries. The dictionary entries presented in Fig. 1 contain different lexical categories such as definite article (PREF), demonstrative pronoun (DEM), preposition (PREP), noun suffix (NSUFF), verb (V) and noun (N).

Entries with the NW (non-word) code should not be analyzed as real words because they are either proclitic or enclitic. For example, the demonstrative pronoun *هذا* ‘*hathaa*’ (this) is a real word while its abbreviation *هـ* ‘*ha*’ is not a real word but it is a proclitic.

Similarly, the preposition *ع* ‘*a*’ (on) is not a real word while *على* ‘*alaa*’ is real word. In fact, inflectional grammars (FLX) generate all the inflected forms of the dictionary entry. As mentioned in Fig. 1, a set of nouns uses an inflectional grammar called “MFP5”.

This grammar is presented by a transducer in Fig. 2. This transducer is dedicated to feminine nouns having the scheme *فعللة* ‘*fa’alalah*’ in the singular and transforming into the other scheme *فعالل* ‘*fa’alil*’ in the plural. For example, the canonical form *بسكلة* ‘*bisklah*’ (bike) remains the same in the first path and becomes *بساكيل* ‘*bsaakil*’ (bikes) in the second path.

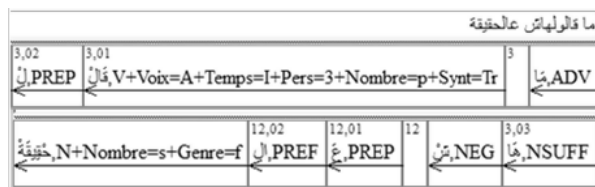


Fig. 9. Example of linguistic analysis

After	Seq.	Before
ق مسحت دموعي	وقلت	زعمة زعمة نسيت
زيت يدي لربي و	قلت	انشالله بوها و امها
تتفرغ عليك كان	ماقوللهاش	يا بطية عرفت وحد
كوها عليه نعاود	نقول	راهر هالديويات ه
ما في مخي اخيرا	نقولها	ونأكد انو مانيش نذا

Fig. 10. Result excerpt of <قال> location

Table 3. Summarizing the obtained metrics for all words

Corpus	Recall	Precision	F-measure
18680	0.89	0.95	0.91

In addition, the TD verb (V) in Fig. 1 is a hollow verb. Thus, it uses an inflectional grammar called "VERBE2" presented by a transducer in Fig. 3. Indeed, this transducer allows to generate hollow verbs whose root origin having the second letter is 'w' and transforming into the letter 'a'.

For example, after conjugation in the present tense (P) with the third person masculine singular (3+m+s) of the Tunisian canonical form مات 'maat' (to die), it becomes the conjugated verb يموت 'ymuut' (he dies).

4.2 Morphological Grammars

To deal with the phenomenon of agglutination, we establish two morphological grammars based on a set of nested finite transducers.

For TD noun agglutination, we construct several transducers, among which the nested transducer shown in Fig. 4 solves the definite noun problem. Indeed, the nested transducer in Fig. 4 contains a subgraph called "PrefNom" and two nodes. The first one recognizes the first lemma (\$1L) as a definite article (PREF) and the second node

recognizes the category (\$2C), the inflectional feature (\$2F) and the semantic and syntactic feature of the second lemma (\$2L). Fig. 5 shows the PrefNom transducer and its PREF subgraph. In the first transducer, the name is stored in the variable (\$N) indicating that the loop <L> means a sequence of letters.

Thus, the contents of the variable (\$N) are verified by a dictionary lookup. With the same variable principle, the second transducer recognizes the definite article ال 'il' (the). For verbs, we also establish several transducers, for example, the set of transducers in Fig. 8 solves the negation form of verbs.

The first transducer is the main graph contains a subgraph called "AdvVNeg" and three nodes. The first one recognizes the first lemma (\$1L) as an adverb (ADV). The second node is explained above. The third node recognizes the third lemma (\$3L) as a negation particle (NEG). In addition, the second transducer recognizes the adverb and the negation particle by its two subgraphs called "ADV" and "NEG" respectively.

In addition, we dedicate a set of finite state transducers to deal with TD particles. For example, the two finite transducers illustrated in Fig. 6 treat a specific type of agglutination for prepositions explained below. The first transducer is the main graph that contains a subgraph called "Prep_spcNsuff" and two nodes.

This subgraph which is the second transducer is used to recognize prepositions that undergo a transformation from the shortened letter alif 'aa' to the letter 'yi'. This problem is solved by using two variables. The first variable (\$Mvide) stores the unchanged part and the second stores the changed part. Thus, the code (\$Mvide#_y=:PREP) adds the letter 'aa' to the first variable and then checks its existence in the dictionary.

In addition, the subgraph called "NSUFF" recognizes the noun suffix in TD. In conclusion, we construct 95 finite state transducers to solve the different forms of the agglutination phenomenon for all grammatical categories in TD; among which 23 main transducers for agglutinated verbs, 13 main transducers for agglutinated nouns and 18 main transducers for agglutinated particles.

Table 4. Obtained results for all words

	Noun	Verb	Particle	Adjective	Total
Corpus	8450	3260	5740	1140	18680
Correct recognized word	7940	2820	4910	990	16660

5 Experimentation and Evaluation

To experiment with our constructed linguistic resources on the collected test corpus, we have implemented our lexical resources in the NooJ linguistic platform. In fact, the dictionary is edited and saved in the file “barcha.dic” which is extended by the file “barcha.nod” after compilation.

Up to now, our NooJ morphological analyzer generates, from 4422 entries, 169815 forms as presented in Fig. 7. Moreover, the morphological grammars allowing the resolution of agglutination are stored in the file “agglutination.nom” and are implemented by finite transducers.

As already indicated, to evaluate our resources, we have collected a corpus from Tunisian dialect novels and social networks such as Facebook and Twitter. The test corpus contains 3300 sentences and 18680 words. The evaluation of our NooJ prototype is based on the recognition of TD words. Thus, we used the known metrics: recall, precision and f-measure.

We obtain the following results presented in Table 3. More precisely, Table 4 shows the results obtained of the prototype application. It shows how well our grammars recognize nouns, verbs, particles and adjectives. Table 4 shows that our prototype recognizes 89% of the total words.

93% of nouns in the corpus are recognized. Some unrecognized nouns are compound proper names, city names or company names. Moreover, our prototype detects 86% of verbs, particles and adjectives in the corpus. In addition, a set of unrecognized words belong entirely to MSA.

For example, ستعودين ‘sata’udina’ (you will come back) is not a Tunisian verb because the proclitic س ‘sa’ (will) does not belong to the Tunisian dialect as well as the flexion of the verb. Another example, the adjective حائرين ‘ha’iriin’ (worried) it is not considered as a Tunisian adjective because the correct writing is حائرين ‘hayiriin’.

Among the undetected words, these contain a repetitive series of letters such as برشاللا ‘barchaaaa’ (many). In fact, our prototype detects all demonstrative, relative and personal pronouns as well as interrogative adverbs. Moreover, agglutinated words are well recognized in different grammatical categories in TD.

For example, the linguistic analysis of the Tunisian sentence shown in Fig. 9: ما قالولهاش عالْحَقِيْقَة: ‘maa kaaluulhaach ‘alhkikah’ (they do not tell her about the truth) is as follows. We get that the word ما ‘maa’ is recognized as an adverb (ADV).

Moreover, the recognized word قالولهاش ‘kaaluulhaach’ (tell) is a verb (V) conjugated in the past tense (I) with the third person (3) plural (p) having recognized enclitics: ل ‘l’ (to) as a preposition and ها ‘haa’ (her) as a noun suffix which are an enclitic IOC and ش ‘ch’ (not) as an adverb of negation (NEG).

Finally, the recognized word عالْحَقِيْقَة ‘alhkikah’ (about truth) is a singular (s) feminine (f) noun (N) that is preceded by the definite article (PREF) ال ‘il’ (the) and also by the preposition (PREP) ع ‘a’ (about).

Thanks to the NooJ linguistic platform, we can locate patterns in the test corpus and detect all the different morphological, inflected, and agglutinated forms of different grammatical categories. For example, to locate a specific verb, we simply write <قال> in the regular expression box.

An excerpt of the result of this location is shown in Fig. 10. Among the results, there are just inflected forms like the words قلت ‘kult’ (I said), نقول ‘nkuul’ (I say), and several agglutinated forms. Precisely, there are words with only proclitics like وقلت ‘wkult’ (and I said), others with only enclitics like نقولها ‘nkuulhaa’ (I said it).

The TD words in our dictionary have different origins like French and Turkish. The unrecognized words have a typographical error or are MSA words. We consider that the results obtained are ambitious. Moreover, they can be improved by increasing the coverage of the dictionary and by adding more morphological rules.

6 Conclusion and Perspectives

In the present paper, we have created a set of linguistic resources for TD in the NooJ language platform. These resources that deal with the agglutination phenomenon are realized through a set of nested finite transducers and are based on a deep linguistic study.

All these resources allow us to construct a NooJ prototype. Moreover, we have demonstrated the efficiency of our NooJ prototype. Thus, the evaluation is performed on a set of sentences belonging to the test corpus. The obtained results are ambitious and show that several agglutinated words can be detected and resolved.

As perspectives, we will increase the coverage of our dictionaries. Furthermore, we will improve our grammars by adding morphological rules that recognize other linguistic phenomena.

References

1. **Buckwalter, T. (2004).** Buckwalter Arabic morphological analyzer: version 2.0.
2. **Eskander, R., Habash, N., Rambow, O. (2013).** Automatic extraction of morphological lexicons from morphologically annotated corpora. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1032–1043.
3. **Hamdi, A., Boujelbane, R., Habash, N., Nasr, A. (2013).** Un système de traduction de verbes entre arabe standard et arabe dialectal par analyse morphologique profonde, pp. 396–406.
4. **Hammouda, N., Torjmen, R., Haddar, K. (2018).** Transducer cascade to parse Arabic corpora. **Silberstein, M., Atigui, F., Kornysheva, E., Métais, E., Meziane, F.**, editors, Natural Language Processing and Information Systems. NLDB '18, volume 10859, pp. 230–237.
5. **Harrat, S., Meftouh, K., Abbas, M., Smaïli, K. (2014).** Building resources for Algerian Arabic dialects. pp. 2123–2127.
6. **Karmani, N. B. M., Soussou, H., Alimi, A. M. (2016).** Intelligent tunisian arabic morphological analyzer. 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), pp. 1–8.
7. **Mesfar, S. (2008).** Analyse morpho-syntaxique automatique et reconnaissance des entités nommées en arabe standard.
8. **Mohamed, B., Azzeddin, e. M., Mohamed Ould, A., Abdelhak, L., Abderrahim, B. (2017).** Alkhalil morpho sys 2: A robust Arabic morpho-syntactic analyzer. Journal of King Saud University - Computer and Information Sciences, Vol. 29, No. 2, pp. 141–146.
9. **Mohamed, M., Graff, D., Bouziri, G., Krouna, S., Bies, A., Kulick, S. (2010).** LDC standard Arabic morphological analyzer (SAMA) version 3.1. Linguistic Data Consortium, LDC catalog number LDC2009E73.
10. **Pasha, A., Elbadrashiny, M., Diab, M., Elkholy, A., Eskandar, R., Habash, N., Pooleery, M., Rambow, O., Roth, R. (2014).** MADAMIRA: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic. Proceedings of the 9th International Conference on Language Resources and Evaluation, pp. 1094–1101.
11. **Ridouane, T., Bouzoubaa, K. (2014).** Building a Moroccan dialect electronic dictionary (MDED). pp. 216–221.
12. **Salloum, W., Habash, N. (2013).** Dialectal Arabic to English machine translation: Pivoting through modern standard Arabic. pp. 348–358.
13. **Salloum, W., Habash, N. (2014).** ADAM: Analyzer for dialectal Arabic morphology. Journal of King Saud University - Computer and Information Sciences, Vol. 26, No. 4, pp. 372–378.
14. **Silberstein, M. (2005).** NooJ's dictionaries. Proceedings of LTC, Vol. 5, pp. 21–23.
15. **Torjmen, R., Haddar, K. (2019).** Construction of morphological grammars for the Tunisian dialect. Formalizing Natural Languages with NooJ 2018 and Its Natural Language Processing Applications, Springer International Publishing, pp. 62–74.

- 16. Zribi, I., Ellouze, M., Belguith, L., Blache, P. (2017).** Morphological analysis of Tunisian dialect. International Workshop on Temporal, Spatial, and Spatio-Temporal Data Mining, Vol. 11107, pp. 180–187.

*Article received on 21/02/2019; accepted on 04/01/2021.
Corresponding author is Roua Torjmen.*