Tunisian dialect agglutination processing with finite transducers

Roua Torjmen[1] and Kais Haddar[2]

[1]Faculty of Economic Sciences and Management of Sfax, Miracl Laboratory, University of Sfax, Sfax, Tunisia
rouatorjmen@gmail.com
[2]Faculty of Sciences of Sfax, Miracl Laboratory, University of Sfax, Sfax, Tunisia
kais.haddar@yahoo.fr

Abstract. The abundance of agglutination phenomenon in the Tunisian dialect requires their treatment. In this context, we want treat this phenomenon. The proposed approach is based on the construction of morphological grammars using a set of finite transducers. These carry great flexibility in the construction of morphological grammars and especially to maintain and reuse them. Our goal is to create a set of linguistic resources allowing the treatment of the phenomenon of agglutination in the Tunisian dialect. NooJ linguistic platform with new technologies makes it possible to elaborate and experiment our resources. The obtained results are ambitious and highlight our proposed method.

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

1 Introduction

Agglutination is a widespread phenomenon in the Tunisian Dialect (TD). This reveals the importance of this linguistic phenomenon and makes their 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 offer 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, they can be integrated in many applications such as automatic annotation of TD corpus, POS-Tagging, TD speech synthesis and automatic translation from TD to MSA and vice versa.

The processing of TD raises other issues, such as 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. Nevertheless, the lack of linguistic resources in linguistic platforms like NooJ for the TD exacerbates the problems of building tools and robust applications.

In this paper, our main objective is to construct morphological grammars based on finite transducers that allow the processing of agglutination in TD. In order to reach

this objective, we carry out a linguistic study on the phenomenon of agglutination. Subsequently, we need to set up a set of lexical resources using NooJ linguistic platform.

This document is divided into six sections. In the second section, we present related work dealing with the morphological analyzer for the Arabic dialect and modern standard Arabic (MSA). In the third part, we exhibit our linguistic study. In the fourth part, we explain our lexical resources related to the phenomenon of agglutination by explaining the designed dictionary and grammars. In the fifth part, 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.

Many works morphologically deal with MSA such as the Buckwalter Arabic Morphological Analyzer (BAMA) [2], the Standard Arabic Morphological Analyzer (SAMA) [5], the morpho-syntactic analyzer Alkhalil [1], MADAMIRA [10] and the Arabic morpho-syntactic analyzer using the NooJ linguistic platform [4, 9].

On the one hand, other works envisage the approach that deals with treatment of Arabic dialects using tools designed for MSA. We quote, for example, the work of [12] that is based on SAMA and BAMA to accept prefixes and suffixes of the Egyptian dialect. In addition, the authors [7] have developed an Algerian morphological analyzer based on the BAMA and Al-Khalil analyzers. Moreover, the Analyzer for Dialectal Arabic Morphology (ADAM) [11] allows the morphological processing of three dialects (Egyptian, Levantine and Iraqi). These three dialects have similar morphological characteristics such as the negation verb, the propositions and the indirect object complement.

On the other hand, other works have chosen another approach that works directly on the dialect. Among these works, the authors [14] have created a bilingual Moroccan dialect electronic dictionary (MDED) in order to develop a Moroccan morphological analyzer. Besides, the authors [3] have proposed a machine learning method to extract Egyptian morphological lexicons from morphologically annotated corpora, such as inflection classes and associated lemmas. Regarding the Tunisian dialect, the authors [6] are interested in creating a morphological analyzer using the Morphological Analyzer and GEnerator for Arabic Dialects (MAGEAD). This analyzer treats only verbs. The authors [8] have suggested a Tunisian morphological analyzer using aebWordNet, Tunisian lexical dictionary and 22 predicate rules. This created system does not treat the standardized Tunisian dialect. In addition, the authors [15, 16] have aimed to create a morphological analyzer processing the Tunisian dialect using the NooJ linguistic platform.

In fact, the second approach usually gives better morphological analyzers than the first one at the quality level because it requires hand-written rules. That is why, this present work is based on [15, 16].

3 Linguistic study on agglutination phenomenon

The TD is strongly agglutinated. Indeed, agglutination is the association of several grammatical categories in the same word. An agglutinative word has either proclitics or enclitics, or both. The proclitics are located before the inflected 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 several models. Regarding the proclitics of a verb, there exists 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.

As proclitics, enclitics are also found in TD verb. there are several types of enclitics: the interrogation adverb (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. In fact, the adverbs of negation and interrogation cannot be together in the same verb.

Table 1. IOC and DOC in Tunisian Dialect

	DOC	**IOC**
1st singular person	ني/ي '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)
3nd masculine singular person	هو/ه/و '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)

The structure of a TD verb can be defined by the following regular expression:

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

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

The nouns possess many models of agglutination. Concerning the proclitics, the defined nouns are preceded by the defined article (PREF) « ال » ‘il’ (the). The prepositions (PREP) that appear before the nouns are as follows: « ب » ‘b’ (by), « ل » ‘l’, « ك » ‘k’ (as) , « م » ‘m’ (from) which is the abbreviation for « من » ‘min’, « ع » ‘’a’ (on) which is the abbreviation for « على » ‘’alaa’ and « ف » ‘fi’ (in) which is the abbreviation for « في » ‘fii’. The combination between the preposition « ل » and the definite article « ال » produces the proclitics « لل ». The demonstrative pronoun « ه » ‘ha’ (this) stands as a proclitic only in front of definite nouns. Enclitics are only found for undefined nouns in the form of an annexation compound. These enclitics are noun suffixes (NSUFF) as presented in Table 2.

Table 2. Noun suffix in Tunisian dialect

	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)
3nd masculine singular person	و/ه/هو ‘huu’ (his)
3rd feminine singular person	ها ‘haa’ (her)
3rd plural person	هم ‘hum’ (their)

In fact, the structure of a TD definite noun can be defined by the following regular expression:

CONJ? PREP? DEM? PREF? Definite_Noun

In addition, the structure of a TD indefinite noun can be defined by the following regular expression:

CONJ? PREP? Indefinite_Noun NSUFF?

For example, the word « وبهالمشروع » ‘wibhalmachruu’a’ (and by this project) is the longest structure for a Tunisian definite noun and « وبمشروعهم » ‘wbimachruu’ahum’ (and by their project) is the longest structure for a Tunisian indefinite noun. The first word is composed of the conjunction « و » ‘wa’ (and), the preposition « ب » ‘bi’ (by), the demonstrative pronoun « ه » ‘ha’ (this), the definite article « ال » ‘al’ (the) and the noun « مشروع » ‘machruu’a’ (show). The second word is composed of the conjunction « و » ‘wa’ (and), the preposition « ب » ‘bi’ (by), the noun « مشروع » ‘machruu’a’ (show), 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, prepositions also have proclitics and enclitics and their regular expression is:

CONJ? ADV? preposition NSUFF? (NEG|INTERR)?

For example, the word « ومامعهش » ‘wmaam’ahimch’ (And not with it) 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 « ه » ‘h’ (it), and finally the negation adverb « ش » ‘ch’ (not).

Furthermore, the structure of a Tunisian personnel pronoun can be defined by the following regular expression:

CONJ? ADV? Personal_Pronoun (NEG|INTERR)?

For example, the word « وماهيش » ‘wmahich’ 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 structure of a Tunisian demonstrative pronoun can be defined by the following regular expression:

CONJ? Prep? Demonstrative_Pronoun

For example, the word « وبهذا » ‘wbihathaa’ is the longest structure for a Tunisian demonstrative pronoun. This word is composed of the conjunction « و » ‘wa’ (and), the preposition « ب » ‘bi’ (by) and the demonstrative pronoun « هذا » ‘hathaa’ (this).

The conjunction « و » ‘wa’ (and) marks a great presence with all grammatical categories. Moreover, the adverb « ما » ‘maa’ is frequently found in words without or with the form of interrogation or negation.

3.4 Effect of agglutination on certain letters

Agglutination phenomenon has an effect on certain letters at the end of the word. The letter correlated tā « ة » becomes « ت ». For example, the noun « كرهبة » ‘karhbah’ (a car) after being agglutinated becomes « كرهبتها » ‘karhbatha’ (her car). Moreover, the letter shortened ʾalif « ى » ‘aa’ also undergoes a transformation and becomes the letter « ي » ‘yi’. For example, the preposition « على » ‘’alaa’ (on) becomes « عليها » ‘’aliiha’ (on her). Furthermore, the letter hamzah « ء » is transformed to the letter yāʾ hamzah « ئ ». For example, the noun « اصدقاء » ‘asdikaa’ (friends) becomes « اصدقائها » ‘asdikaiha’ (her friends).

4 Proposed method

Our proposed method for treating the agglutination phenomenon begins with the automatic extraction of all non-repetitive words from the study corpus. Afterwards, the filtering phase allows the elimination of the same words that are written in different inflected forms. Then, the choice of a canonical form allows the presentation of the collected words. Finally, the enrichment of this canonical form is achieved by the addition of morphological, lexical and syntactic characteristics.

All these phases are established thanks to the dictionary, inflectional and morphological grammars in NooJ linguistic platform [13].

4.1 Dictionary and inflectional grammars

Dictionary is a set of entries that are composed of a canonical form, a lexical category and an inflectional grammar if it is necessary. Fig. 1 presents an example of dictionary entries.

```
ال,PREF+NW
ة,DEM+NW
هذا,DEM+FLX=DEM
ع,PREP+NW
على,PREP
ك,NSUFF+2+s+NW
كُمْ,NSUFF+2+p+NW
مَـاك,V+FLX=VERBE2
بِـشْكُـلَـةْ,N+FLX=MFP5
تَـجْرُبَـةْ,N+FLX=MFP5
```

Fig. 1. Example of dictionary entries

The dictionary entries of 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 whereas its abbreviation « ه » ‘ha’ is not a real word but it is a proclitic. Similarly, the preposition « ع » ‘’a’ (on) is not a real word whereas « على » ‘’alaa’ is real word.

In fact, inflectional grammars generate all inflected forms of the dictionary entry. As mentioned in Fig. 1, a set of nouns uses an inflectional grammar called “MFP5”. This latter is illustrated in Fig. 2.

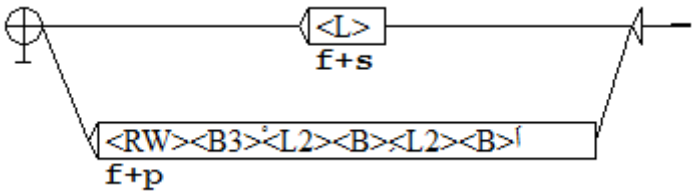

Fig. 2. Example of transducer for nouns

This grammar is dedicated to feminine nouns having scheme « فعللة » ‘fa’alalah’ in the singular and taking the scheme « فعالل » ‘fa’alil’ in the plural. For example, the canon-

ical form « بسكلة » ‘bisklah’ kept by the first path (bicycle) becomes « بساكل » ‘bsaakil’ (bicycles) by the second path.

In addition, verbs in Fig. 1 are hollow verbs and they use an inflectional grammar called VERBE2 illustrated in Fig. 3.

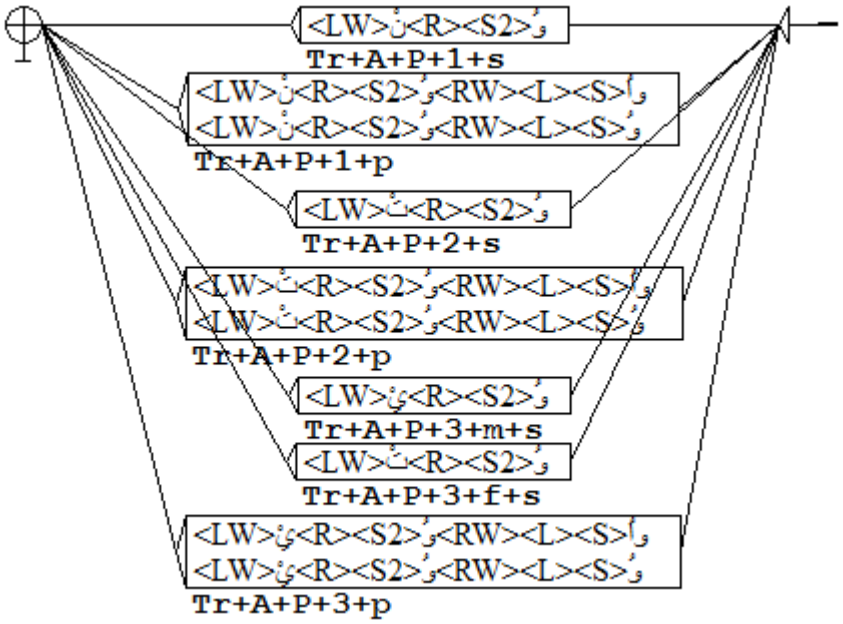

Fig. 3. Extract of transducer for Hollow verbs

In fact, this grammar makes it possible to generate hollow verbs whose origin of the second letter is « و » ‘w’ and is transformed into the letter « ا » ‘a’. For example, the Tunisian canonical form « مات » ‘maat’ (to die) becomes, after conjugation in the present tense (P) with the third masculine singular person (3+m+s), the word « يموت » ‘ymuut’ (he dies).

4.2 Morphological grammars

To treat the agglutination phenomenon, we establish two morphological grammars based on a set of recursive finite transducers. Regarding the noun agglutination, we construct several transducers, among which; the recursive transducer presented in Fig. 4 solves the issue of definite noun.

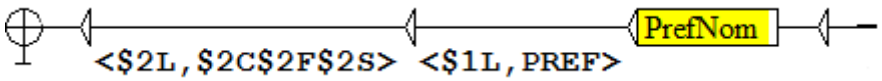

Fig. 4. Transducer for definite noun

The recursive transducer of Fig. 4 contains a sub-graph called PrefNom and two nodes. The first one recognizes the first lemma ($1L) as 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). The fig. 5 presents the transducer of PrefNom and their sub-graph PREF.

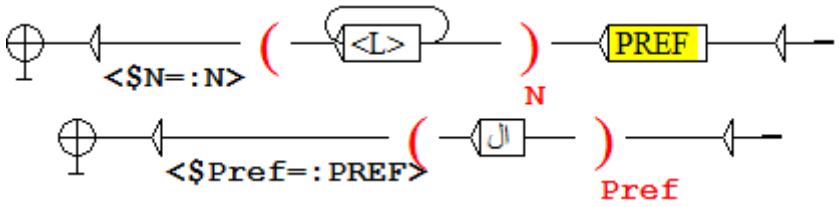

Fig. 5. PrefNom and Pref transducers

In the first transducer, the noun is stored in the variable ($N) indicating that the loop (<L>) means a sequence of letters. Thus, the content of the variable ($N) is verified by a search in the dictionary. With the same variable principle, the second transducer recognizes the definite article « ال » 'al' (the).

Regarding verbs, we also establish a several transducers, for example, the set of transducers in Fig. 6 solves the negation form of verbs.

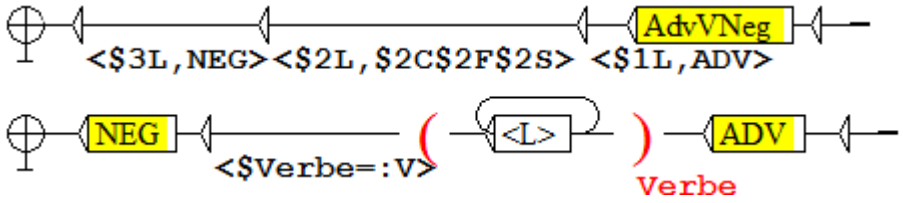

Fig. 6. Transducer for negation form of verbs

The first transducer is the main graph contains a sub-graph called AdvVNeg and three nodes. The first one recognizes the first lemma ($1L) as an adverb (ADV). The second node is explained before. 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 respectively by its two sub-graphs ADV and NEG.

In addition, we dedicate transducers to particles. For example, the set of transducers of Fig. 7 deals with a type of agglutination for prepositions.

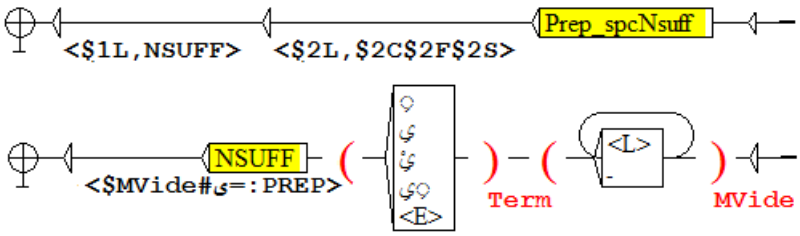

Fig. 7. Transducer for agglutinated preposition

The first transducer is the main graph contains a sub-graph called Prep_spcNsuff and two nodes. This sub-graph allows to recognize the prepositions that undergo a transformation of letter shortened ʾalif « ى » 'aa' into the letter « ي » 'yi'. This issue is solved by using two variables. The first variable ($MVide) stores the unchanged part and the second stores the modified part. Thus, the code ($MVide#ى= :PREP) makes it possible to add the letter « ى » to the first variable and then checks its existence in the dictionary. Moreover, the sub-graph called NSUFF recognizes the suffix noun.

In conclusion, we construct 95 transducers; among which 23 main transducers for agglutinated verbs, 13 main transducers for agglutinated nouns and 18 main transducers for agglutinated particles.

5 Experimentation and evaluation

To experiment 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" that 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. 8. In addition, morphological grammars allowing agglutination resolving and saved in the file "agglutination.nom" implemented by finite transducers.

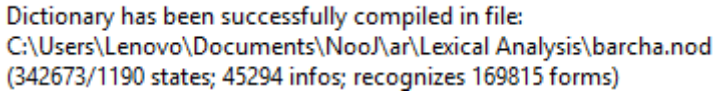

Fig. 8. Dictionary compilation result

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.

Table 3. Summarizing the obtained metrics for all words

Corpus	Recall	Precision	F-measure
18680	0.89	0.95	0.91

More precisely, Table 4 shows the results obtained for nouns, verbs, particles and adjectives. The obtained results of the prototype application are presented in Table 4.

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

Table 4 shows that our prototype recognizes 89% of total words. 93% of nouns are recognized. Some unrecognized nouns are compound proper, city or company names. In addition, our designed prototype detects 86% of verbs, particles and adjectives. Moreover, a set of unrecognized words belongs 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 and also the inflection of the verb. Another example, the adjective « الحائرين » ‘ha’iriin’ (worried) it is not considered a Tunisian adjective because the correct writing is « الحايرين » ‘hayiriin’. Among the undetected words, these contain a repetitive series of letters such as « برشاااا » ‘barchaaaaaa’ (very much). In fact, our prototype detects all demonstrative, relative and personal pronouns and interrogative adverbs. Furthermore, the agglutinated words are well recognized in different grammatical categories.

For example, the linguistic analysis of the Tunisian sentence presented in Fig. 9: « ما قالولهاش عالحقيقة » ‘maa kaaluulhaach ‘alhkikah’ (they do not tell her about the truth) is as follows. We obtain that the word « ما » ‘maa’ is recognized as an adverb (ADV). Moreover, the recognized word « قالولهاش » ‘kaaluulhaach’ (tell) is a verb "V" conjugated to the past tense (I) with the third person (3) plural (p) having a recognized IOC enclitics in the form « ل » ‘l’ (to) as a preposition and « ها » ‘ha’ (her) as a suffix noun. Finally, the recognized word « عالحقيقة» (about the truth) is a singular (s) feminine (f) noun (N) that is preceded by the definite article (PREF) « ال » ‘al’ (the) and also the preposition (PREP) « ع » ‘’a’ (about).

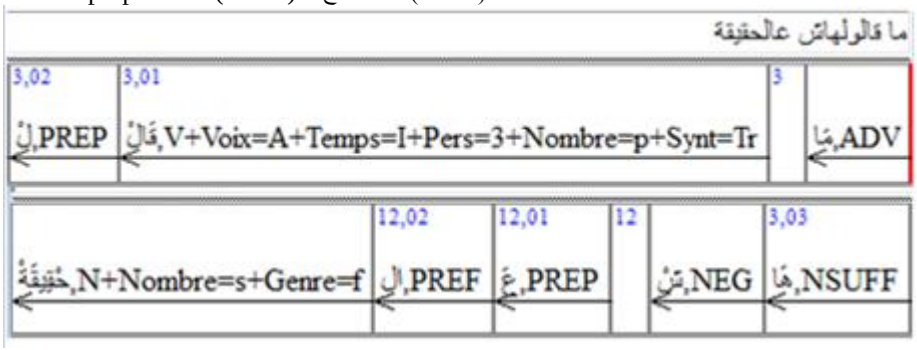

Fig. 9. Example of linguistic analysis

Thanks to NooJ linguistic platform, we can locate patterns in the test corpus and detect all the different morphological, derivative and agglutinated forms of different grammatical categories. For example, to locate a specific verb, we just write <قال> in the regular expression box. An extract of result of this location is presented in Fig. 10. Among the results, there are just inflected forms such as the words « قلت » ‘kult’ (i said), « نقول » ‘nkuul’ (we say), and several agglutinated forms. Precisely, there are words having only proclitics such as the word « وقلت » ‘wkult’ (and i said), others having only enclitics such as « نقولها » ‘nkuulhaa’ (i said it) and « قالتلو » ‘kaalitluu’ (she tells him). Furthermore, there are others verbs having both such as « ماتقوللهاش » ‘maatkuulilhaach’ (do not tell her) and « ونقولها » ‘wnkuulhaa’ (and i said it).

After	Seq.	Before
هاذم دويا ما يتسوفو وزير	تقول	للبات صفاقس كي تتسوفهم يحكيو
زعمة زعمة نسيت واول ما	وقلت	ت بوجيعت الفراق مسحت دموعي
عليه سافر على بغل حتساكم	تقولش	مسالو ويحطلنا اللي هو تاعب
انشاالله بوها و امها يتسوفوها	قلت	بعد هزيت يدي لربي و
شنوا فيلم فيه ممثلة صهيونية	قال	وومن الكلها تحكي عليه و
راهو البطيات هن الجميلات اش	قالولهم	يقعدو هكاكة خاطر برشا مرضى
راني نحكي عالديناصورات الكبار مالا	نقول	كيفاش ؟ ماكش متعذبة ؟ و نعاود
يا بطية عرفت وحدها انو	ماتقوللهاش	بطية و تتنرفز عليك كان
راهو هالدبدوبات هاذم مايتسموش ديناصورات	نقول	شي المعاق يفركوها عليه نعاود
و نأكد انو مانيش نتمسخر	نقولها	المعلومة كيما في مخي اخيرًا
أيا قوم. قالتلو: لوين؟ قاللها	وقاللها	جبد القله ورماها على الكريطة
لوين؟ قاللها : نمشيو نعديو نهار :	قالتلو	.على الكريطة وقاللها أيا قوم
نمشيو نعديو نهار في البلاد :	قاللها	وقاللها أيا قوم. قالتلو: لوين؟

Fig. 10. Result extract

In fact, TD words in our dictionary have different origins such as French, Maltese and Turkish. In fact, unrecognized words present a typographical error or are Arabic standard words. We consider that the obtained results are ambitious. Besides, they can be improved by increasing the dictionary coverage and adding other morphological rules.

6 Conclusion and perspectives

In the present paper, we have created a set of linguistic resources for TD in NooJ linguistic platform. These resources that treat the agglutination phenomenon are accomplished through a set of recursive finite transducers and are based on a deep linguistic study. All these resources allowed us to construct a NooJ prototype. In addition, we have shown the efficiency of our NooJ prototype. Thus, the evaluation is performed on a set of sentences belonging to test corpus. The obtained results are ambitious and show that several agglutinated words can be detected. As perspectives, we will increase the coverage of our designed dictionaries. Furthermore, we will improve our morphological grammars by adding morphological rules recognizing other linguistic phenomena.

References

1. Boudchiche, M., Mazroui, A., Bebah, M. O. A. O., Lakhouaja, A., & Boudlal, A.: AlKhalil Morpho Sys 2: A robust Arabic morpho-syntactic analyzer. Journal of King Saud University-Computer and Information Sciences 29(2), 141–146 (2017).

2. Buckwalter, T.: Buckwalter Arabic Morphological Analyzer Version 2.0. Linguistic Data Consortium, LDC catalog number LDC2004L02, ISBN 1-58563-324-0 (2004).
3. Eskander, R., Habash, N., & Rambow, O.: Automatic extraction of morphological lexicons from morphologically annotated corpora. In Proceedings of the 2013 conference on empirical methods in natural language processing, Seattle, Washington, USA, Association for Computational Linguistics, pp. 1032-1043 (2013).
4. Ghezaiel Hammouda, N., Torjmen, R., Haddar, K.: Transducer Cascade to Parse Arabic Corpora. In: Silberztein, M., Atigui, F., Kornyshova, E., Métais, E., Meziane, F. (eds.) NLDB 2018. LNCS, vol. 10859, pp. 230–237. Springer, Cham (2018).
5. Graff, D., Maamouri, M., Bouziri, B., Krouna, S., Kulick, S., & Buckwalter, T.: Standard Arabic morphological analyzer (SAMA) version 3.1. Linguistic Data Consortium, LDC catalog number LDC2009E73. (2009).
6. Hamdi, A., Boujelbane, R., Habash, N., & Nasr, A. : Un système de traduction de verbes entre arabe standard et arabe dialectal par analyse morphologique profonde. In: Traitement Automatique des Langues Naturelles, pp. 396-406 (2013).
7. Harrat, S., Meftouh, K., Abbas, M., Smaili, K.: Building resources for Algerian arabic dialects. In: Fifteenth Annual Conference of the International Speech Communication Association, Singapore, 14–18 September 2014, pp. 2123–2127 (2014).
8. Karmani, N. B., Soussou, H., & Alimi, A. M.: Intelligent Tunisian Arabic morphological analyzer. In: 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), Agadir, Morocco, pp. 1-8 (2016).
9. Mesfar, S.: Analyse morpho-syntaxique et reconnaissance des entités nommées en arabe standard. Doctoral dissertation, Thèse, Université de franche-comté, France (2008).
10. Pasha, A., Al-Badrashiny, M., Diab, M. T., El Kholy, A., Eskander, R., Habash, N., & Roth, R.: MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014), Reykjavik, Iceland, May 2014, pp. 1094–1101 (2014).
11. Salloum, W., & Habash, N.: Adam: Analyzer for dialectal arabic morphology. Journal of King Saud University-Computer and Information Sciences, 26(4), 372–378 (2014).
12. Salloum, W., & Habash, N.: Dialectal arabic to english machine translation: Pivoting through modern standard arabic. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (NAACL-HLT), Atlanta, GA, pp. 348–358 (2013).
13. Silberztein, M.: NooJs dictionaries. In: Proceedings of LTC, Poland, 21–23 April 2005, vol. 5, pp. 291–295 (2005)
14. Tachicart, R., Bouzoubaa, K., Jaafar, H.: Building a Moroccan dialect electronic dictionary (MDED). In: 5th International Conference on Arabic Language Processing, pp. 216–221 (2014).
15. Torjmen R., Haddar K.: Construction of Morphological Grammars for the Tunisian Dialect. In: Mauro Mirto I., Monteleone M., Silberztein M. (eds) Formalizing Natural Languages with NooJ 2018 and Its Natural Language Processing Applications. NooJ 2018. Communications in Computer and Information Science, vol 987, pp. 62-74. Springer, Cham (2019).
16. Torjmen, R., Haddar, K.: Morphological analyzer for the Tunisian dialect. In: Sojka, P.,Horák, A., Kopeček, I., Pala, K. (eds.) International Workshop on Temporal, Spatial, and Spatio-Temporal Data Mining, vol. 11107, pp. 180–187. Springer, Cham (2018).