Discontinuous Arabic frozen expressions modelization and implementation

Asmaa Kourtin[1], Asmaa Amzali[1], Mohammed Mourchid[1], Samir Mbarki[2]
[1]Computer Science Research Laboratory, Faculty of Science, Ibn Tofail University, Kenitra, Morocco
[2]EDPAGS Laboratory, Faculty of Science, Ibn Tofail University, Kenitra, Morocco

Article Info	ABSTRACT
	Frozen expressions hold significant importance in the field of natural language processing, attracting considerable attention from researchers across various languages in recent years. The Arabic language, in particular, boasts a wealth of frozen expressions inherited from the pre-Islamic and early Islamic periods, with persistent usage to the present day. This linguistic richness has motivated researchers to systematically collect, classify, and elucidate these expressions. Various classifications have emerged, addressing aspects such as continuity, discontinuity, allowance for variations, and restriction from variations. Our aim is to produce lexicon-grammar tables of discontinuous Arabic frozen expressions and implement them. Our approach involves the meticulous collection and study of these expressions, followed by the transformation of their lexicon-grammar tables into dictionaries and syntactic grammars within the NooJ platform. This methodology allows us to recognize and annotate these expressions in texts and corpora, even when they exhibit discontinuity. Such recognition has the potential to address several challenges in automatic natural language processing, including the area of automatic translation.
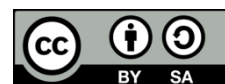
Corresponding Author:

Asmaa Kourtin
Computer Science Research Laboratory, Faculty of Science, Ibn Tofail University
B.P. 133 Kénitra, Morocco
Email: asmaa.kourtin@yahoo.fr

1. INTRODUCTION

Natural language processing, a field within computer science, centers on the interaction between computers and human language. In recent decades, there has been significant interest in applications associated with automated natural language processing. These applications encompass a range of tasks, such as automatic translation, spelling and grammar correction, corpus analysis, automated question-answering, and more. Two primary approaches are commonly employed to address these tasks: the statistical approach and the linguistic approach. While statistical techniques are widely utilized nowadays, they often fall short of delivering satisfactory results. Consequently, this has prompted the development of natural language processing platforms grounded in the linguistic approach, as exemplified by NooJ. These platforms integrate various levels of analysis, including lexical, morphological, syntactic, and semantic analysis, providing a more robust solution to language processing. The lexicon of a language contains not only simple words but also phrases and compound words. In other words, it contains both monolexical and polylexical units. The first category has undergone thorough examination in numerous linguistic studies, whereas the second has historically been overlooked, receiving only limited attention in selected works. This neglect is not due to any lack of importance since collocation is a process that ensures lexical renewal, and a language that no longer produces frozen expressions is a language doomed to extinction.

Recently, many researchers have turned their attention to frozen or idiomatic expressions, resulting in numerous studies in various languages. In particular, a large number of dictionaries have been produced on the subject. In addition, recent theoretical studies have tried to highlight either the exceptional or the non-exceptional character of these expressions. Gross [1], conducted a study of frozen expressions in French and created their lexicon-grammar tables. Baptista [2], focused on identifying the syntactic features that serve as discriminators between compound nouns and conventional noun phrases. In another study cited in [3], Baptista *et al.* [4] developed an electronic dictionary of European Portuguese frozen expressions. In addition, the integration of verbal idioms into the natural language processing system STRING [4]. In a related context, Koza and Rivas Folch carried out computational modeling of verbal idioms in Chilean Spanish based on a lexicon-grammar proposal [5]. Briskilal and Subalalitha [6], presented an ensemble prediction model for the idiomatic and literal sentences' classification, combining two basic models, BERT and RoBERTa. Adelina and Suprayogi [7], analyzed English idioms related to the human body that are equivalent to Indonesian idioms. Kocijan and Librenjak [8], used a rule-based approach and NooJ syntactic grammars to recognize any verb-based idiom in any syntactic position. Fenta and Gebeyehu [9], built a model to identify idioms for the Amharic language using a supervised machine-learning approach.

For Arabic, there are no dictionaries devoted to idiomatic sentences in the lexical heritage. A specific dictionary of this type of structure was not published until the twentieth century when Al-Din [10] wrote a dictionary of idiomatic expressions, which he called " التعبير الأصطلاحي: دراسة في تأصيل المصطلح ومفهومه ". Siny [11] and several researchers also wrote a dictionary, which they called "ومجالاته الدلالية وأنماطه التركيبية ". In the twenty-first century, several dictionaries appeared, including " معجم التعبيرات الإصطلاحية في القرآن الكريم: دراسة في التراكيب " by Dawood [12] and " التعبير الاصطلاحي في العربية المعاصرة الدلالة " by Ghorab [13]. In addition, numerous researchers have dedicated their efforts to examining frozen expressions and their translation, particularly focusing on their French translation. Alqahtni [14], analyzed idiomatic expressions extracted from the Saudi newspaper "Al Riyadh." Boulaalam [15], delved into the verbal sentence category in Arabic idioms and scrutinized their translation into French. Addressing the issue of emergence and interlanguage comparison through a psycholinguistic and experimental investigation, Yaiche contributed to this area [16]. The research involved both native and non-native French-speaking adults, specifically Tunisian Arabic speakers. Nseme [17], conducted a theoretical study on Arabic frozen expressions, encompassing the classification of their morphosyntactic structures and the analysis of constituent elements within these expressions. Ali [18], focused on translating idiomatic expressions extracted from the Qur'an into the French language. Abdelmaksoud [19], studied the French translation of idiomatic expressions extracted from the Qur'an. Dziri [20], focused on the frozen structure translation following the lexicographic approach. Kourtin *et al.* [21], studied fixed and continuous frozen expressions in modern Arabic and created their lexicon-grammar tables.

Our primary goal is to develop linguistic resources for modern Arabic frozen expressions. To achieve this, we will construct lexicon-grammar tables for these expressions, enabling us to articulate their complete grammatical, syntactic, and semantic characteristics. This work is a continuation of our previous work on constructing lexicon-grammar tables for continuous frozen expressions that do not allow variation [21].

In this paper, we aim to study the discontinuous Arabic frozen expressions, such as "لقي مصرعه" (*laqiya masra'ahu*; he was killed). To this end, we will first collect and study these expressions. We will then create their lexicon-grammar tables, which we will implement within a linguistic platform such as NooJ by transforming them into dictionaries [22] and building their corresponding syntactic grammars [23], [24]. This framework will enable the identification and analysis of these expressions in texts and corpora and can be used in natural language processing applications, including automatic translation.

This paper is organized as follows: in the second section, we explain our method by dividing it into three sub-sections. In the first sub-section, we present the lexicon-grammar approach. In the second sub-section, we study the Arabic frozen expressions. In the third sub-section, we provide our proposed lexicon-grammar table for discontinuous Arabic frozen expressions. In the third section, we implement this lexicon-grammar table in the NooJ platform by creating its dictionary and syntactic grammar to detect these expressions in texts and corpora. In the fourth section, we present the results of our method and discuss their advantages. Finally, in the fifth section, we conclude this document by summarizing the main results of this research and giving some perspectives.

2. METHOD
2.1. The lexicon-grammar approach
The lexicon-grammar approach is a central pillar of automatic natural language processing (ANLP), which helps machines understand and process human language. This methodology was first introduced by Maurice Gross at the LADL laboratory in 1975 [1], [25]-[28]. The lexicon-grammar approach revolutionized the field by providing a structured framework that allowed machines to analyze and manipulate linguistic

elements and contributed significantly to developing ANLP systems and applications. The lexicon-grammar approach continues to influence speech technology development and natural language understanding. The language's lexicon is systematically organized into different classes, each containing a set of entries sharing a particular definitional structure.

The lexicon-grammar tables use a coding system, presented in a matrix format, to represent these classes. This matrix configuration has rows corresponding to individual lexical entries and columns containing various properties such as syntactic, semantic, distributional, morphological, and transformational aspects. Within each matrix cell, we can find either a lexical element or the symbols "+" or "-," indicating the presence or absence of the corresponding property in the respective column. In cases where a lexical entry has multiple meanings, it is necessary to create different lexical entries, each dedicated to articulating a different meaning or interpretation.

2.2. The Arabic frozen expressions

A frozen expression is a sentence whose overall meaning cannot be understood by simply understanding the significance of its components and combining them. It is a group of words that together form a sense different from their lexical meaning, both singly and in combination, and this signification comes from the agreement of a group of linguists on a concept that they give to this verbal grouping. The composition of an idiomatic expression is based on a series of words limited by semantic and structural factors that make them a new semantic unit.

The importance of idiomatic expressions in modern linguistic research cannot be overestimated, as they constitute a significant part of the lexicon of any language. These expressions concentrate meaning, articulating it with precision and clarity, thus ensuring linguistic communication concisely and eliminating ambiguity in the message conveyed to the recipient. Furthermore, frozen expressions serve as a linguistic treasure trove, enriching the resources available to both parties in the communication process with a wealth of possibilities for conveying diverse meanings, such as:

الشعب كاهل الحرب أثقلت (*Athqalati al harbu kahila al-chaâbi*)

↓ ↓ ↓ ↓ ↓

N2+Det C1 N0+Det V
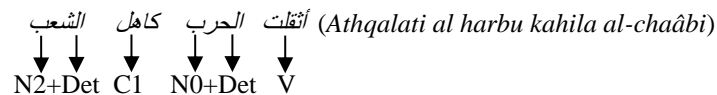
in this sentence, the frozen expression "كاهل أثقلت" is present, where "أثقلت" (*athqalati*) that means "to make heavy" serves as the verb denoted by V, "الحرب" (*al harbu*) that refers to "the war" represents the specified subject in the form of N0+Det, "كاهل" (*kahila*) representing "the shoulders" stands as the constant object denoted by C1, and "الشعب" (*al-chaâbi*) translated as "the people" functions as the second object in the form of N2+Det. However, the global meaning of this sentence is "the war burdened the people". As demonstrated above, this meaning is unique and results from a collective interpretation by linguists, which cannot be derived by simply combining the individual significance of each component.
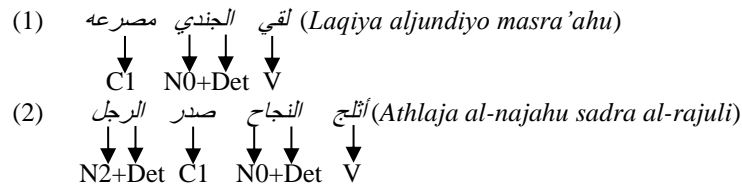
The Arabic language is very rich in frozen expressions. Recognizing the importance of these expressions, several researchers have taken the initiative to collect, categorize, and elucidate them. Several classification schemes have emerged according to the needs of each researcher. Some classifications are based on the degree of frozenness, while others are organized according to structural features, and categories [21]. In Table 1 these classifications provide a structured framework for linguistic analysis and research to understand and categorize the variety of frozen expressions in a language.

Table 1. Frozen expressions' classification

Degree of frozenness	Structure	Category
Fixed	C0 C1	Noun phrases (جمل اسمية)
Semi-flexible	V C0	Verbal phrases (جمل فعلية)
Flexible	V C0 C1	Genitive (الإضافة)
	V C0 Prép C1	Adjective (نعت)
	V N0 C1	Preposition phrases (الجار و المجرور)
	V N0 C1 N2	Single word (كلمة واحدة)
	…..	.....

In this paper, we aim to study the discontinuous Arabic frozen expressions. For that, we will create their lexicon-grammar tables, generate their associated dictionaries, and create syntactic grammars for them to analyze texts and corpora. Afterward, our research included an extensive exploration of Arabic dictionaries dedicated to frozen expressions, including "معجم التعبير الإصطلاحي في العربية المعاصرة" [12]. This effort resulted in a collection of about 2,400 frozen expressions. We then specifically extracted the discontinuous frozen

expressions and studied them by describing their linguistic properties. We identified numerous structural patterns within this category, including "VN0C1," "VN0Prép1N1," etc. In this particular study, we focused on discontinuous verbal frozen expressions, especially those in the forms "VN0C1" and "VN0C1N2", and amassed a collection of about 400 such expressions, e.g.:

(1) لقي الجندي مصرعه *(Laqiya aljundiyo masra'ahu)*

$$C1 \quad N0{+}Det \quad V$$

(2) أثلج النجاح صدر الرجل *(Athlaja al-najahu sadra al-rajuli)*

$$N2{+}Det \quad C1 \quad N0{+}Det \quad V$$

the sentence (1), which means "the soldier was killed," is a verbal phrase made up of "verb+subject+object," and the sentence (2), expressing "success warmed the man's heart," is composed of "verb+subject+first object +second object".

2.3. Lexicon-grammar tables for discontinuous Arabic frozen expressions

In this paper, we have worked on the discontinuous verbal frozen expressions for the Arabic language with the basic constructions "VN0C1" and "VN0C1N2". For this purpose, we have created the EFDA1 class, which groups these expressions. Figure 1 displays an excerpt from our lexicon-grammar table labeled "EFDA1." The entries in this table belong to the grammatical category of verbs, denoted by "V." The columns within this table encompass syntactic-semantic features, with "N0Hum" indicating a human subject, "C" serving as a noun for frozen expressions, and the features "VN0C1" and "VN0C1N2" representing the structures of frozen expressions included in the table. For example, the frozen expression "أثلج صدر" (*athlaja sadra*) is employed to convey "to give him peace of mind and make him feel satisfied." This expression constitutes a verbal phrase with the structure "VC1," where "V" corresponds to "أثلج" (*athlaja*; to snow) and "C1" corresponds to "صدر" (*sadra*; chest).

<Opt>Exemple	<OPT>ExpFig	FLX	FR	ENG	VN0C1N2	VN0C1	C1=Adv1N1	C1=N1ConjV1	C1=N1Prép2N2	C1=ConjV1	C1=N1Prép2	C1=N1	N2Hum	N2Hum	C1pc	C1NHum	C1Hum	N0NHum	N0Hum	C1	V
أبلى الجندي المصري بلاء حسنا في حرب أكتوبر 1973م.	أبلى بلاء حسنا	<E>	<E>	<E>	-	+	-	-	-	-	+	-	-	+	-	-	+	بلاء حسنا	أَبْلَى
قانون الصحافة أثار ضجة كبيرة.	أثار ضجة	V_atara	<E>	<E>	-	+	-	-	-	-	+	-	-	+	-	+	+	ضجة	أَثَارَ
الشعب الأفغاني أقلت كاهله الحروب و الفتن.	أقل كاهله	V_atkala	<E>	<E>	+	-	-	-	-	+	-	+	+	-	-	+	+	كاهل	أَقْلَ
قال الأب لابنه : نجاحك أثلج صدري.	أثلج صدر	V_atkala	<E>	<E>	+	-	-	-	-	+	-	+	+	-	-	+	+	صدر	أَثْلَج
التنمية الصناعية أثمرت ثمارها في اليابان و الصين.	أثمر ثماره	V_atkala	<E>	<E>	-	+	-	-	-	+	-	-	-	+	+	-	أثمر <Pos+ثمار+N>	أَثْمَرَ
نجيب محفوظ أحرز قصب السبق في الرواية العربية.	أحرز قصب السبق	V_atkala	<E>	<E>	-	+	-	-	-	+	-	-	+	-	-	+	قصب السبق	أَحْرَزَ
أحكمت إسرائيل قبضتها على أرض فلسطين و شعبها.	أحكم قبضته على ...	V_atkala	<E>	<E>	+	-	-	-	+	-	+	+	+	-	+	+	على <Pos+قبضة+N>	أَحْكَمَ
أحيت وزارة الأوقاف ليلة القدر في حفل أقيم بهذه المناسبة.	أحيى ليلة	<E>	<E>	<E>	+	-	-	-	-	+	+	-	+	-	-	+	ليلة	أَحْيَى
أخذ الإرهاب حيزا كبيرا في إعلام أمريكا.	أخذ حيزا في ...	V_kataba	<E>	<E>	+	-	-	-	+	-	+	-	+	-	+	-	حيزا في	أَخَذَ
تقاعد المدير فأخذ نائبه مكانه.	أخذ مكان...	V_kataba	<E>	<E>	+	-	-	-	-	+	-	+	-	+	-	+	مكان	أَخَذَ
استطاعت إسرائيل أن تغتصب فلسطين ثم أخرجت لسانها للعرب.	أخرج لسانه ل...	V_atkala	<E>	<E>	-	+	-	-	+	+	-	-	+	+	أخرج <N+لسان><Pos+N+PrépLi>	أَخْرَجَ
حكمت المحكمة ببراءة المتهم و أخلت سبيله.	أخلى سبيل...	<E>	<E>	<E>	+	-	-	-	-	+	-	+	-	+	-	+	سبيل	أَخْلَى
أدار المذيع الحديث بمهارة ولباقة.	أدار الحديث	V_atara	<E>	<E>	-	+	-	-	-	+	-	-	+	-	+	الحديث	أَدَارَ
أدارت الفتاة رؤوس زملائها بجمالها الأخاذ.	أدار رأس...	V_atara	<E>	<E>	+	-	-	-	-	+	+	+	-	+	+	أدار <N+رأس>	أَدَارَ
أدار الشباب ظهورهم للقيم الأصلية.	أدار ظهره ل...	V_atara	<E>	<E>	-	+	-	-	+	+	+	-	-	+	أدار <N+ظهر><Pos+N+PrépLi>	أَدَارَ
أزاح الوزير الستار عن المشروع الجديد.	أزاح الستار عن ...	V_atara	<E>	<E>	+	-	-	-	+	-	+	-	+	+	+	الستار عن	أَزَاحَ

Figure 1. Excerpt of the lexicon-grammar table EFDA1

3. IMPLEMENTATION

The integration of lexicon-grammar tables into the NooJ platform involves two primary steps. Initially, each table must undergo conversion into a NooJ dictionary. Subsequently, a syntactic grammar for each table needs to be constructed, utilizing the linguistic knowledge embedded in the table to identify sentences [23], [24]. It is important to note that both the dictionary and the syntactic grammar should share the same name and be located in the "Lexical Analysis" folder within the NooJ platform. Figure 2 provides an excerpt from the automatically generated dictionary "EFDA1.dic," derived from the lexicon-grammar table "EFDA1" using the program designed to generate NooJ dictionaries from lexicon-grammar tables [22]. This generated dictionary encompasses all the information encoded in the respective lexicon-grammar table.

```
# NooJ V5
# Dictionary
#
# Language is: ar
#
# Alphabetical order is not required.
#
# Use inflectional & derivational paradigms' description files (.nof), e.g.:
# Special Command: #use paradigms.nof
#
# Special Features: +NW (non-word) +FXC (frozen expression component) +UNAMB (unambiguous lexical entry)
#                   +FLX= (inflectional paradigm) +DRV= (derivational paradigm)
#
# Special Characters: '\' '"' ' ' ',' '+' '-' '#'
#
#use Flex.nof
أُبْـلَى,V+EFDA1+FXC+C1=حسناً بلاء+N0Hum+C1NHum+C1=N1+VN0C1
أَثَّـرَ,V+EFDA1+FXC+C1=صحة+N0Hum+N0NHum+C1NHum+C1=N1+VN0C1+FLX=V_atara
أَثْـقَلَ,V+EFDA1+FXC+C1=هل كـ+N0Hum+N0NHum+C1pc+N2Hum+C1=N1+VN0C1N2+FLX=V_atkala
أَثْـلَجَ,V+EFDA1+FXC+C1=صدر+N0Hum+N0NHum+C1pc+N2Hum+C1=N1+VN0C1N2+FLX=V_atkala
أَثْـمَرَ,V+EFDA1+FXC+C1=<N+ثمار+Pos>+N0NHum+C1NHum+C1=N1+VN0C1+FLX=V_atkala
أَحْـرَزَ,V+EFDA1+FXC+C1=قصب السبق+N0Hum+C1NHum+C1=N1+VN0C1+FLX=V_atkala
أَحْـكَمَ,V+EFDA1+FXC+C1=<N+قبضة+Pos> على+N0Hum+N0NHum+C1pc+N2Hum+N2NHum+C1=N1Prép2+VN0C1N2+FLX=V_atkala
أَحْـيَى,V+EFDA1+FXC+C1=ليلة+N0Hum+C1NHum+N2NHum+C1=N1+VN0C1N2
أَخَـذَ,V+EFDA1+FXC+C1=في ا حيز+N0NHum+C1NHum+N2NHum+C1=N1Prép2+VN0C1N2+FLX=V_kataba
أَخَـذَ,V+EFDA1+FXC+C1=مكان+N0Hum+C1NHum+N2Hum+C1=N1+VN0C1N2+FLX=V_kataba
أَخْـرَجَ,V+EFDA1+FXC+C1=<N+لسان+Pos><N+PrépLi>+N0Hum+N0NHum+C1pc+N2Hum+N2NHum+C1=N1Prép2+VN0C1+FLX=V_atkala
أَخْـلَى,V+EFDA1+FXC+C1=سبيل+N0Hum+N0NHum+C1NHum+N2Hum+C1=N1+VN0C1N2
أَذَاعَ,V+EFDA1+FXC+C1=الحديث+N0Hum+C1NHum+C1=N1+VN0C1+FLX=V_atara
```

Figure 2. Excerpt of the generated dictionary EFDA1.dic

To recognize all the sentences in the previous dictionary, we must develop a morphological grammar designed to identify nouns with suffixes such as "لسانه" (*lisanahu*; his tongue), where "لسان" represents the noun and "ه" serves as the suffix. Figure 3 shows an excerpt of the morphological grammar "Morpho.nom," created for our table EFDA1, which allows recognizing all the forms constituted by a name or adjective followed by a suffix such as "ه", "ك", "ي".

The creation of NooJ dictionaries is not sufficient in the process of integrating lexicon-grammar tables into the NooJ platform. Additionally, the development of syntactic grammar is essential to this process. For that, we have created a syntactic grammar "EFDA1.nog" for the class "EFDA1", allowing us to recognize all the discontinuous verbal frozen expressions contained in this table. As shown in Figure 4, the created grammar takes into account all the grammatical, syntactic, and semantic characteristics contained in his associated dictionary, "EFDA1.dic," such as conditions on the subject, the second complement, and the sentence's structure.

Figure 3. Excerpt of the morphological grammar "Morpho.nom" for EFDA1 table

Figure 4. Syntactic grammar for EFDA1 table: (EFDA1.nog)

4.   RESULTS AND DISCUSSION

To evaluate our generated dictionary "EFDA1.dic" as shown in Figure 2 and the syntactic grammar "EFDA1.nog" as shown in Figure 4, we analyzed a few sentences containing discontinuous frozen expressions of the lexicon-grammar table EFDA1. Figure 5 shows the annotations generated by the constructed grammar "EFDA1.nog" of the Arabic sentences " أبلى الجندي بلاء حسنا" (the soldier did well) with the structure "VN0C1" in Figure 5(a), " أثقلت الحرب كاهل الشعب" (the war burdened the people) with the structure "VN0C1N1" in Figure 5(b), and " لقي أحمد نحبه" (Ahmed is dead) with the structure "VN0C1" in Figure 5(c). We can see in this figure that the grammar was able to recognize and annotate the discontinuous Arabic frozen expressions " أبلى بلاء حسنا", " أثقل كاهل" and " لقي نحبه" A notable observation is the grammar's ability to recognize intricate structures, as evidenced by the identification of the object in the last sentence, " نحبه" The object is composed of the noun " نحب" followed by a possessive adjective " ه," which the morphological grammar "Morpho.nom" as shown in Figure 3 aptly recognizes.

By undertaking this study, we anticipate gaining a nuanced understanding of the syntactic structures and semantic nuances embedded within discontinuous frozen expressions in the Arabic language. The process of creating lexicon-grammar tables facilitates the systematic organization of linguistic elements, while the generation of dictionaries provides a practical tool for reference and analysis. Additionally, the creation of syntactic grammars allows for the implementation of structured rules to recognize and interpret these expressions in various linguistic contexts. Ultimately, this research not only contributes to advancing our understanding of Arabic linguistic structures but also provides valuable resources and tools for linguistic analysis, facilitating more accurate and nuanced interpretations of texts and corpora containing discontinuous frozen expressions.
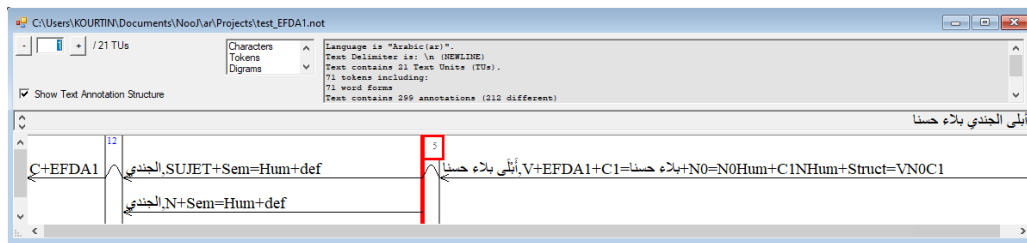
(a)
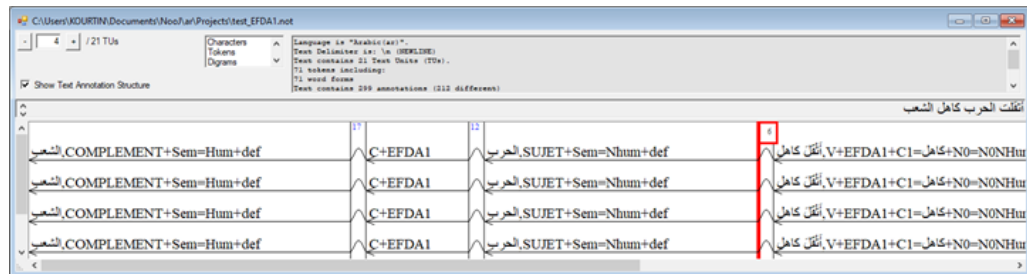
(b)
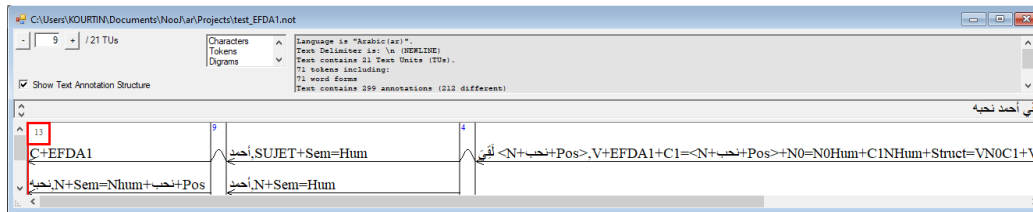
(c)

Figure 5. The annotations of the sentences; (a) " أبلى الجندي بلاء حسنا", (b) " أثقلت الحرب كاهل الشعب", and (c) " لقي أحمد نحبه"

5.   CONCLUSION AND PERSPECTIVES

The Arabic language faces challenges due to the limited availability of linguistic resources, especially in the frozen expressions area. Research on contemporary Arabic frozen expressions and the development of specialized lexicon-grammar tables are being undertaken to address this issue. This

undertaking has the potential to significantly enrich linguistic resources and facilitate the use and updating of linguistic data. By delving into the study of Arabic frozen expressions and creating corresponding lexicon-grammar tables, the language community will gain a valuable reference tool. This comprehensive resource would provide a detailed insight into the structure, the meaning, and the usage of these expressions. Such a reference would not only simplify the management of linguistic data but would also streamline subsequent modifications. This will contribute to more efficient and accurate analysis and translation of Arabic texts. This paper focuses on discontinuous Arabic frozen expressions, with the primary objective of creating their lexicon-grammar tables. Additionally, we have developed dictionaries and syntactic grammars for these expressions to seamlessly integrate them into natural language processing platforms, notably the NooJ platform. The integration of these resources into a natural language processing platform enables users to efficiently analyze, process, and translate Arabic texts, considering the specificities of frozen expressions. Therefore, this research makes a significant contribution to the advancement of the study of Arabic frozen expressions. It also provides valuable tools for linguistic analysis and translation tasks, especially in the specific context of the NooJ platform.

As perspectives, we plan to extend this research by developing lexicon-grammar tables for additional categories of frozen expressions in Arabic. We aim to incorporate translation features for these expanded categories, thus extending our study to a wide range of frozen expressions, including idioms, proverbs, and other commonly used linguistic expressions. In addition, we aim to transform these comprehensive lexicon-grammar tables into electronic dictionaries, facilitating their seamless integration into natural language processing platforms. This integration will be complemented by creating specific syntactic grammars tailored to these frozen expressions. This strategic enhancement aims to provide the platform with advanced capabilities for detecting and analyzing these expressions in different texts and corpora.

REFERENCES

[1] M. Gross, "French fixed phrases (in French: *les phrases figées en français*)," *L'information grammaticale*, vol. 59, no. 1, pp. 36–41, 1993, doi: 10.3406/igram.1993.3139.
[2] J. Baptista, "Compositional vs Frozen Sequences," *Journal of Applied Linguistics, Special Issue on Lexicon-Grammar*, pp. 81–92, 2004.
[3] J. Baptista, A. Correia, and G. Fernandes, "Frozen sentences of portuguese: formal descriptions for NLP," in *ACL Workshop on Multiword Expressions: Integrating Processing*, T. Tanaka, A. Villavicencio, F. Bond, and A. Korhonen, Eds., Barcelona, Spain, pp. 72–79, 2004, doi: 10.3115/1613186.1613196.
[4] J. Baptista, N. Mamede, and I. Markov, "Integrating verbal idioms into an NLP system," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8775, pp. 250–255, 2014, doi: 10.1007/978-3-319-09761-9.
[5] W. Koza and N. Rivas Folch, "Computational modelization of verbal idioms of chilean spanish based-on lexicon-grammar proposal," *MonTI*, pp. 94–120, Jan. 2021, doi: 10.6035/MonTI.2020.ne6.3.
[6] J. Briskilal and C. N. Subalalitha, "An ensemble model for classifying idioms and literal texts using BERT and RoBERTa," *Information Processing and Management*, vol. 59, no. 1, p. 102756, Jan. 2022, doi: 10.1016/j.ipm.2021.102756.
[7] C. Adelina and S. Suprayogi, "Contrastive analysis of English and Indonesian idioms of human body," *Linguistics and Literature Journal*, vol. 1, no. 1, Art. no. 1, Jun. 2020, doi: 10.33365/llj.v1i1.235.
[8] K. Kocijan and S. Librenjak, "Recognizing verb-based Croatian idiomatic MWUs," *Communications in Computer and Information Science*, vol. 607, pp. 96–106, 2016, doi: 10.1007/978-3-319-42471-2_9.
[9] A. A. Fenta and S. Gebeyehu, "Automatic idiom identification model for Amharic language," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 22, no. 8, 2023, doi: 10.1145/3606864.
[10] K. Z. H. al-Din, *Idiomatic expression: a study of the term's rooting, its concept, its semantic fields, and its syntactic patterns* (in Arabic: التعبير الأصطلاحي: دراسة في تأصيل المصطلح ومفهومه ومجالاته الدلالية وأنماطه التركيبية), Anglo-Egyptian Library, Egypte,1985.
[11] M. I. Sini, *Contextual dictionary of idiomatic expressions Arabic-Arabic* (in Arabic: المعجم السياقي للتعبيرات الاصطلاحية عربي - عربي), Librairie du Liban Publishers, Beirut, 1996.
[12] M. M. Dawood, *A dictionary of idiomatic expression in modern Arabic* (in Arabic: معجم التعبير الإصطلاحى فى العربية المعاصرة), Dar Gharib for printing, publishing and distribution, Egypte, 2003.
[13] A. H. Ghorab, *Idiomatic expressions in the Holy Qur'an: A study in structure and meaning* (in Arabic: التعبيرات الاصطلاحية في القرآن الكريم: دراسة في التراكيب والدلالة), Nancy Library, 2005.
[14] H. M. Alqahtni, "The structure and context of idiomatic expressions in the Saudi press," Ph.D. dissertation, University of Leeds, 2014.
[15] A. Boulaalam, "Idiomatic expressions and translation: a lexico-semantic study (in Arabic: التعابير المسكوكة والترجمة : دراسة معجمية دلالية)," Ph.D. dissertation, Sidi Mohamed Ben Abdellah University, Fez, Morocco, 2017.
[16] S. Yaiche, "Figment and preaching in Arabic and French: linguistic and psycholinguistic studies (in French: *figement et prédication en arabe et en français : études linguistiques et psycholinguistiques*)," Ph.D. dissertation, Sfax and Paris 8 universities, 2014.
[17] S. A. Minko-Mi-Nseme, "Modeling Arabic idioms to build a lexical database (in French: *modélisation des expressions figées en arabe en vue de la constitution d'une base de données lexicale*)," Ph.D. dissertation, Lyon 2, 2002.
[18] M. S. Ali, "Translating fixed expressions: language and culture (in French: *la traduction des expressions figées : langue et culture*)," *Traduire. Revue française de la traduction*, no. 235, pp. 103-123, Dec. 2016, doi: 10.4000/traduire.865.
[19] M. Abdelmaksoud, "Idiomatic expressions in the Qur'an and their French translation: A contrastive analytical study of Arabic to French in three French interpretations of the meaning of the Qur'an (in French: *les expressions idiomatiques dans le Coran et leur traduction française : Étude analytique contrastive de l'arabe vers le français dans trois interprétations françaises du sens du Coran*)," Ph.D. dissertation, Mansoura University, Egypte, 2018.

[20] H. Dziri, "The bilingual dictionary as an assistant tool to the translator: the case of frozen structures," *Synergies Algerie*, vol. 23, pp. 87–98, 2016.
[21] A. Kourtin, A. Amzali, M. Mourchid, A. Mouloudi, and S. Mbarki, "Lexicon-grammar tables for modern arabic frozen expressions," *Communications in Computer and Information Science*, vol. 1520 CCIS, pp. 28–38, 2022, doi: 10.1007/978-3-030-92861-2_3.
[22] A. Kourtin, A. Amzali, M. Mourchid, A. Mouloudi, and S. Mbarki, "The automatic generation of NooJ dictionaries from lexicon-grammar tables," *Communications in Computer and Information Science*, vol. 1153 CCIS, pp. 65–76, 2020, doi: 10.1007/978-3-030-38833-1_6.
[23] M. Silberztein, "Complex annotations with NooJ," in *Proceedings of the 2007 International NooJ Conference*, Cambridge Scholars Publishing, pp. 214-227, 2008.
[24] M. Silberztein, *Formalizing Natural Languages: The NooJ Approach*. in Formalizing Natural Languages: The NooJ Approach, ISTE Editions, pp. 1-333, London, UK, 2016, doi: 10.1002/9781119264125.
[25] M. Gross, Syntactic methods: completive construction regime (in French: *Méthodes en syntaxe: régime des constructions complétives)*. Universite de Paris-Vincennes., 1975.
[26] M. Gross, "Building electronic dictionaries (in French: *la construction de dictionnaires électroniques*)," *Annales des Télécommunications.*, vol. 44, no. 1, pp. 4–19, Jan. 1989, doi: 10.1007/BF02999875.
[27] M. Gross, "Lexicon-grammar and the syntactic analysis of French," in *The 10th International Conference on Computational Linguistics, COLING 1984 and 22nd Annual Meeting of the Association for Computational Linguistics, ACL 1984*, 1984, pp. 275–282.
[28] M. Gross, "A local grammar for expressing feelings (in French: *une grammaire locale de l'expression des sentiments*)," *Langue française*, vol. 105, no. 1, pp. 70–87, 1995, doi: 10.3406/lfr.1995.5294.

BIOGRAPHIES OF AUTHORS
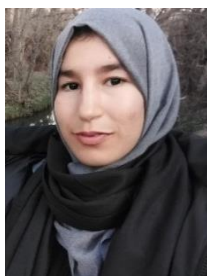
Asmaa Kourtin received the high school baccalaureate in Mathematics Sciences A, the B.Sc. degree in Mathematics and Computer Science from Ibn Tofail University, Kenitra, Morocco, and the M.Sc. degree in computer science from Ibn Tofail University, Kenitra, Morocco. She is currently in the final stages of getting her Ph.D. in the computer science field of computational linguistics with a project on Arabic Natural Language Processing. Her research interests include software engineering, artificial intelligence, computational linguistics, and automatic natural language processing. She can be contacted at email: asmaa.kourtin@yahoo.fr.

Asmaa Amzali received the High School Baccalaureate in physics, the B.Sc. degree in Mathematics and Computer Science from Ibn Tofail University, Kenitra, Morocco, and the M.Sc. degree in computer science from Ibn Tofail University, Kenitra, Morocco. She is currently in the final stages of getting her Ph.D. in the computer science field of computational linguistics with a project on Arabic Natural Language Processing. Her research interests include software engineering, artificial intelligence, computational linguistics, and automatic natural language processing. She can be contacted at email: asmamzali@hotmail.com.

Mohammed Mourchid Doctorate Degree in Computer Science in 1999, Associate Professor at the Computer Science Department at the Faculty of Sciences at Ibn Tofail University in Kenitra, Morocco. He is a supervisor of computational linguistic projects. He has been working on an Arabic dictionary since 2017 in the same department. His research interest score is 108.9, the number of citations is 94, and his h-index is 5. His research interests include natural language processing, web semantics, and information systems. He can be contacted at email: mourchidm@hotmail.com.

Samir Mbarki Professor at the Department of Computer Science at the Faculty of Sciences, Ibn Tofail University. He does research in Model Driven Engineering, Software Engineering, Artificial Intelligence, and Natural Language processing. Their current project is resource allocation in wireless networks. His research interest score is 540.6, the number of citations is 616, and his h-index is 12. His research interests include programming languages, software engineering, artificial intelligence, MDA, and natural language processing. He can be contacted at email: mbarkisamir@hotmail.com.