Indonesian Journal of Electrical Engineering and Computer Science
Vol. 33, No. 2, February 2024, pp. 1243~1251
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v33.i2.pp1243-1251

Lexicon-grammar tables standardization and implementation

Asmaa Kourtin[1], Asmaa Amzali[1], Mohammed Mourchid[1], Abdelaziz Mouloudi[2], 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

Article history:

Received Oct 29, 2023
Revised Dec 14, 2023
Accepted Dec 16, 2023

Keywords:

C1D table
Electronic dictionary
Lexicon-grammar table
Natural language processing
NooJ platform
Syntactic and semantic analysis
Syntactic grammar

ABSTRACT

The lexicon-grammar approach is a very important linguistic approach in automatic natural language processing (NLP). It allows for the description of the lexicon of the language through readable and intuitive tables for human manual editing. However, the automatic use of the lexicon-grammar tables in the automatic NLP platforms remains difficult, given the incompatibility between the codes used to represent the properties in the lexicon-grammar tables and those used to represent the properties in the automatic NLP platforms. In this work, we present our method of standardizing the lexicon-grammar tables for the French language, since they constitute very rich lexical, syntactic, and semantic linguistic resources. First, we standardize their properties so that they can be compatible with those used in the NLP platforms. Then, to implement the standardized tables, we used a linguistic platform such as NooJ. For that, we describe the process of integrating these tables into this platform through the automatic generation of the dictionaries from these tables. Finally, to test the efficiency of the generated dictionaries, we create for some of them syntactic grammars that take into account all the grammatical, syntactic, and semantic information contained in the dictionaries.

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

Automatic natural language processing (NLP) is based on a lexical, morphological, syntactic, and semantic analysis of the language. Today, it is the basis of multiple applications such as translation applications, orthographic and grammar checker applications, and question-answering systems. This has created a critical need to represent in a readable and intuitive way the characteristics of the language's lexicon, allowing at the same time easy editing for humans and an automatic understanding and use for applications. Therefore, the lexicon-grammar approach responds perfectly to this need by offering a representation and a methodology for the study of the grammatical, syntactic, and semantic characteristics of the language's lexicon. It consists of describing the language's lexicon through readable and intuitive tables for human manual editing.

The lexicon-grammar approach was initiated by Gross [1]–[5] and his LADL team in 1975, who pioneered the construction of tables for the French lexicon, detailing the lexical classification method and the construction principles behind these tables. Leclère [6] also proposed an organization of lexicon-grammar tables for French verbs. In a parallel effort for the French language, Mathieu [7]–[10] constructed lexicon-grammar tables specifically for psychological verbs to establish a feeling system that would facilitate the recognition of these verbs in texts. Subsequently, Silberztein [11], [12] made significant contributions by

extending Gross's work, he created additional tables to complete the French lexicon and also integrated the theory of lexicon-grammar tables into the NooJ platform, allowing it to take advantage of this innovative approach. In a related context, Tolone [13]–[16] made an effort to collect existing works on lexicon-grammar tables for the French language. Her efforts focused on normalizing these works, with the ultimate goal of constructing the LGLex lexicon.

Several notable contributions to the lexicon-grammar approach have been made for different languages. In the case of Arabic, El Hannach [17] pioneered this approach by studying psychological verbs, qualitative verbs [18], and support verbs [19]. In a parallel effort, Amzali *et al.* [20] created lexicon-grammar tables for psychological verbs, studied their transformations [21], and their syntactic analysis [22]. In a related context, Kourtin *et al.* [23] developed a program to generate NooJ dictionaries from lexicon-grammar tables and created lexicon-grammar tables for modern Arabic frozen expressions [24]. For the Italian language, Vietri [25] made notable contributions by constructing an annotated corpus and developing lexicon-grammar tables specifically designed for idiomatic constructions [26], [27], locatum and location verbs [28]. For the same language, Guarasci *et al.* [29] proposed an innovative method for conducting open information extraction from natural language sentences based on lexicon-grammar. In addition, Machonis [30] studied the neutral phrasal verbs for the English language [30], [31] and constructed lexicon-grammar tables for them. For the Chilean Spanish, Koza and Folch [32] performed computational modeling of verbal idioms based on a lexicon-grammar proposal. Finally, for Portuguese, Mota *et al.* [33] worked on predicate nouns with the support verb "fazer" and integrated their lexicon-grammar table into Port4NooJ. Barreiro *et al.* [34] worked on paraphrase generation for Portuguese.

The significant number of studies conducted on this approach for different languages, including the French language, which has been the subject of several works by different researchers, has led to the creation of several inhomogeneous tables. This inhomogeneity made it difficult to automatically integrate these tables into the automatic NLP platforms, because of the incompatibility between the codes used to represent the properties in the lexicon-grammar tables and those used to represent the properties in the linguistic platforms. The objective of this work is, on the one hand, i) to standardize these lexicon-grammar tables for the French language by standardizing all their properties so that they are compatible with the format of the platforms properties, and on the other hand, and ii) to implement these tables in a linguistic platform such as NooJ.

This paper is structured as follows: in the second section, we explain our method by dividing it into five sub-sections. In the first sub-section, we present the lexicon-grammar approach. In the second sub-section, we clarify how the lexicon-grammar tables are integrated into linguistic platforms such as NooJ. In the third sub-section, we cite the main motivations of our work. In the fourth sub-section, we describe our method of standardizing the properties of the French lexicon-grammar tables. In the fifth sub-section, we provide our method of standardizing the lexicon-grammar tables using the standardized properties. In the third section, we experiment this standardization for various lexicon-grammar tables on the NooJ platform by generating their dictionaries and creating their syntactic grammars to evaluate this work. In the fourth section, we provide the results of our method and discuss the benefits of this standardization. Finally, in the fifth section, we conclude this document by summarizing the key findings of this research and giving some perspectives.

2. METHOD

2.1. Overview of the lexicon-grammar approach

The lexicon-grammar approach stands as a crucial linguistic methodology in the field of NLP. It was initiated by Gross [1]–[5] and his LADL team in 1975. The lexicon of a language must be divided into several classes, which group the entries that share the same definitional construction.

A lexicon-grammar table corresponds to a class and is presented by a matrix, where the lines represent the lexical entries. The columns indicate the grammatical, syntactic, semantic, distributional, and transformational properties, and the cells contain either; i) a lexical element, ii) the symbol "+" to indicate that the corresponding lexical entry accepts the property of the corresponding column, or iii) the symbol "-" otherwise. If an element has more than one meaning, it must have multiple lexical entries, one for each sense, since the properties depend on the meaning of the lexical entry.

For example, the C1D class as shown in Figure 1 represents French frozen expressions that have the basic construction "N0 V Det1 C1", where "N0" means the subject, "V" means the verb, "Det1" means the first determinant, and "C1" means a constant object. For the 8th frozen expression "*abolir le temps*" (to abolish time), the property "<ENT> V" (the verb) is "*abolir*" (to abolish), the property "<ENT> Det1" (the first determinant) is "*le*" (the), and the property "<ENT> C1" (the frozen object with the verb) is the word "*temps*" (time). This expression accepts the properties "N0 =: N-hum", which means that the subject must be a non-human noun, and "[*passif*]", which means that the frozen expression can be transformed into the passive voice.

<ID>	N0 =: Nhum	N0 =: N-hum	Ppv =: se figé	Ppv =: en figé	Ppv =: y figé	Ppv =: Neg	<ENT>Ppv	<ENT>V	N0 V	<ENT>Det1	N0 V Det1 C1 Prép2 N2	<ENT>C1	C1 =: Npc	[passif]	Exemple
1	+	-	-	-	-	-	<E>	abandonner	-	la	-	compétition	-	+	§abandonner§la compétition
2	+	-	-	-	-	-	<E>	abandonner	-	le	-	domicile conjugal	-	+	§abandonner§le domicile conjugal
3	+	-	-	-	-	-	<E>	abandonner	-	les	-	lieux	-	+	§abandonner§les lieux
4	+	-	-	-	-	-	<E>	abandonner	+	la	-	partie	-	+	§abandonner§la partie
5	+	-	-	-	-	-	<E>	abandonner	-	la	+	pose	-	+	§abandonner§la pose
6	+	-	-	-	-	-	<E>	abandonner	-	le	-	terrain	-	+	§abandonner§le terrain
7	+	-	-	-	-	-	<E>	abjurer	-	la	-	foi catholique	-	+	§abjurer§la foi catholique
8	-	+	-	-	-	-	<E>	abolir	-	le	-	temps	-	+	§abolir§le temps
9	+	+	-	-	-	-	<E>	accélérer	-	le	-	mouvement	-	+	§accélérer§le mouvement
10	+	-	-	-	-	-	<E>	accompagner	-	le	-	pas	-	+	§accompagner§le pas

Figure 1. Excerpt from the lexicon-grammar table C1D for French frozen expressions

2.2. Integration of lexicon-grammar tables in NooJ

The lexicon-grammar tables allow researchers to edit the lexical, syntactic, semantic, and transformational entries' characteristics in an intuitive, readable, and easy way without requiring computer skills. This has created a capital need to integrate these lexicon-grammar tables into the automatic NLP platforms. NooJ is a linguistic development environment that allows the modeling of linguistic knowledge through electronic dictionaries and grammars to make it exploitable by the machine. To integrate the lexicon-grammar tables in this platform, Silberztein [12] proposed to transform them into dictionaries and syntactic grammars, as illustrated in Figure 2. His idea is to build for each table a dictionary and syntactic grammar, both having the same name, preferably the name of the table to which they refer. This dictionary and this syntactic grammar must be created in the same folder on the NooJ platform, which is "Lexical Analysis". Each dictionary's entry corresponds to an entry in the lexicon-grammar table.

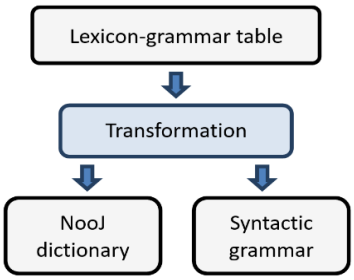

Figure 2. The lexicon-grammar tables integration in NooJ

2.3. Motivations

In the present era, automation has garnered significant importance owing to its inherent benefits, particularly in optimizing time and reducing human effort. Our earlier contribution focused on the automatic generation of NooJ dictionaries from lexicon-grammar tables [23]. Nevertheless, this automated process faced challenges resulting from the lack of standardization in the lexicon-grammar tables, necessitating manual preprocessing for each table. Indeed, the representation format of the syntactic-semantic properties in these tables doesn't conform to the representation format of the platforms properties. In addition, we can have the same property represented by different codes or several different properties represented by the same code.

As illustrated in Figure 3, for the grammatical category "verb", some researchers use the code "V", others use the code "<ENT> V", or even the code "verbe". For the property denoting a human or animated noun, some use the code "N0 =: Nhum", others use the code "N0hum", or even the code "N0 = +hum". Also, each one uses a different assignment symbol, such as "=:", "=", or "=:"" [23].

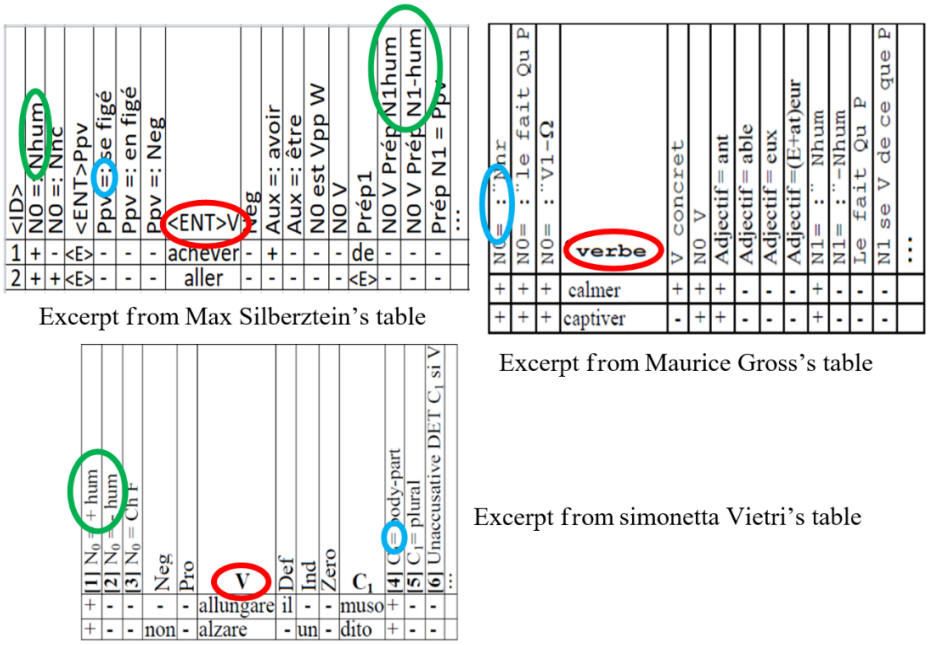

Figure 3. The non-standardization of the lexicon-grammar tables' properties

This non-standardization represents an obstacle to the well-functioning of the automatic generation process of the electronic dictionaries from these lexicon-grammar tables. In addition, the non-existence of a standard helping and assisting linguists to define the properties' names in a unified way makes the automatic generation task difficult. This leads to the necessity of the lexicon-grammar tables' properties standardization, which is the aim of this work.

2.4. Strandardization of the lexicon-grammar tables' properties

To standardize the French lexicon-grammar tables' properties, we started by listing the properties used in these tables and giving their meanings. Their number of occurrences, and the tables where they are used. This phase allowed us to identify 515 properties for verbs and 225 properties for frozen expressions.

Then, we proposed for each property a meaningful name to be used in the future, which would conform to the format of the platform's properties. For example, the property labeled "Nhum" or "N =: Nhum", denoting a human or animated noun, appears 251 times in the tables "V_1", "V_3", "V_4", "V_5". This property will be replaced by "H" in general, "H0" if it represents a subject, or "H1" if it represents a first object as shown in Table 1.

For the lexicon-grammar tables' properties that are not taken into account, we put some conditions to standardize them:

- The property must have a meaningful name that is not previously used as a property, so as not to cause ambiguities with the existing properties.
- The property's name must not contain unrecognized symbols in the electronic dictionaries such as "=:", or special characters such as "+" or "#" for the NooJ platform.
- In preference, the property's name must not contain a space, and there must be no space before and after the symbol "=" if it exists.
- Two different columns must not contain the same information.

2.5. Standardization of French lexicon-grammar tables

To standardize the lexicon-grammar tables of the French language, we replaced the old non-standardized properties in these tables with their standardized equivalents. To do this, we have listed 67 tables for the verbs and 69 tables for the frozen expressions. For example, to standardize the table C1D that represents the French frozen expressions having the basic construction "N0 V Det1 C1" (see Figure 1), we made the following changes as shown in Figure 4:

- The property "N0 =: Nhum", denoting a human or animated subject, became "H0".
- The property "N0 =: N-hum", denoting a non-human subject, became "NH0".

- The property “Ppv =: se figé”, denoting that the clitic pronoun “*se*” is frozen with the verb “V”, became “Ppv=SeFigé”.
- The property “Ppv =: en figé”, denoting that the clitic pronoun “*en*” is frozen with the verb “V”, became “Ppv=EnFigé”.
- The property “Ppv =: y figé”, denoting that the clitic pronoun “*y*” is frozen with the verb “V”, became “Ppv=YFigé”.
- The property “C1 =: Npc”, denoting that the object is a body or a body part of a person, became “C1pc”.

Table 1. Excerpt from the standardization of the French lexicon-grammar tables’ properties

Property name before standardization	Property name after standardization	Meaning	Number of occurrences	Tables
NHum | N =: Nhum | N0 =: Nhum | …	H, H0, H1, H2, …	Human or animated noun	251	V_1, V_3-V_16, V_18, V_31R, V_32H, V_32R3, V_33, V_34L0, V_35L, V_35LD, …
N-hum | N =: N-hum | N0 =: N-hum |…	NH, NH0, NH1, NH2, …	Non-human noun	195	V_1, V_3, V_5-V_16, V_18, V_32A, V_32CL, V_32CV, V_32D, V_32NM, …
NAbstrait | N =: Nabs | N0 =: Nabs | …	A, A0, A1, A2, …	Noun denoting an abstract entity	10	V_4, V_31H, V_32C, V_32D, V_32H, V_35L, V_35LD, V_35LR, …
Nnr | N =: Nnr | N0 =: Nnr | …	Nr, Nr0, Nr1, Nr2, …	Noun denoting a person, a concrete object, an abstract entity, or take the form of a sentential subject or an infinitive clause	35	V_3, V_6-V_16, V_18, V_31R, V_32H, V_32NM, V_32R2, V_33, V_34L0, V_35L, V_35LD, V_35LR, V_35LS, V_35R, V_35S, …
Npc | N =: Npc | N0 =: Npc | …	Pc, Pc0, Pc1, Pc2, …	Noun denoting the body or a body part of the person	17	V_36SL, V_38LD, V_38LR, V_38LS, V_32NM, V_32C, V_36S, V_37M1, V_37M2, …
…..	…..	…..	…..	…..

<ID>	H0	NH0	Ppv=SeFigé	Ppv=EnFigé	Ppv=YFigé	Ppv=Neg	Ppv	V	N0V	Det1	N0VDet1C1Prép2N2	C1	C1pc	Passif	<OPT>Exemple
1	+	-	-	-	-	-	<E>	abandonner	-	la	-	compétition	-	+	§abandonner§la compétition
2	+	-	-	-	-	-	<E>	abandonner	-	le	-	domicile conjugal	-	+	§abandonner§le domicile conjugal
3	+	-	-	-	-	-	<E>	abandonner	-	les	-	lieux	-	+	§abandonner§les lieux
4	+	-	-	-	-	-	<E>	abandonner	+	la	-	partie	-	+	§abandonner§la partie
5	+	-	-	-	-	-	<E>	abandonner	-	la	+	pose	-	+	§abandonner§la pose
6	+	-	-	-	-	-	<E>	abandonner	-	le	-	terrain	-	+	§abandonner§le terrain
7	+	-	-	-	-	-	<E>	abjurer	-	la	-	foi catholique	-	+	§abjurer§la foi catholique
8	-	+	-	-	-	-	<E>	abolir	-	le	-	temps	-	+	§abolir§le temps
9	+	+	-	-	-	-	<E>	accélérer	-	le	-	mouvement	-	+	§accélérer§le mouvement
10	+	-	-	-	-	-	<E>	accompagner	-	le	-	pas	-	+	§accompagner§le pas

Figure 4. Excerpt from the standardized lexicon-grammar table C1D for French frozen expressions

3. IMPLEMENTATION

To implement the standardized lexicon-grammar table C1D for French frozen expressions, we used the NooJ platform as a natural language platform. For that, we create the dictionary “C_c1d.dic” by running our program that automatically generates the NooJ dictionaries from lexicon-grammar tables [23]. The generated dictionary contains 1,661 entries with all their grammatical, syntactic, and semantic characteristics coded in the lexicon-grammar table. For each entry, we have added i) the code “c1d”, which represents the name of the class “C1D” to reference the lexicon-grammar table from which the entry is taken, and ii) the specific property “FXC” for frozen expressions, which means that the lexical entry should not be treated as an ambiguous linguistic unit [11] as shown in Figure 5.

A syntactic grammar was made by Max Silberztein in the French linguistic resources of the NooJ platform for the lexicon-grammar table C1D, but it doesn’t recognize all the properties appearing on this table. To that end, we have created a syntactic grammar that adds to the grammar previously constructed by

Max Silberztein all the grammatical, syntactic, and semantic characteristics contained in the dictionary, such as conditions on the subject, the clitic pronoun, and the sentence's structure as shown in Figure 6. The syntactic grammar created has the name "C_c1d.nog", which must be the same as the name of the dictionary, "C_c1d.dic", to which the syntactic grammar refers.

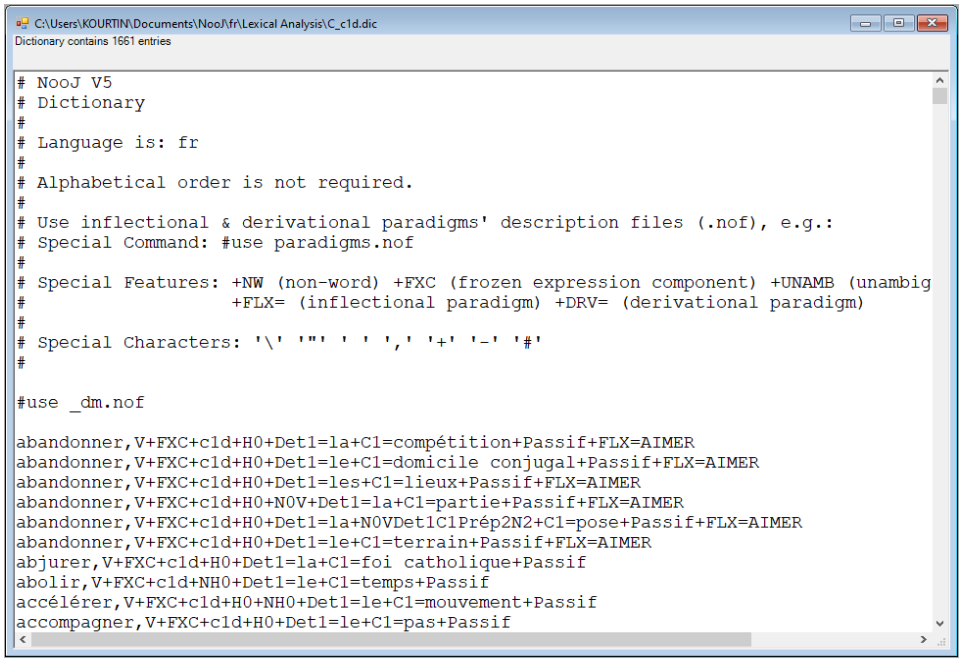

Figure 5. Excerpt from the dictionary "C_c1d.dic" generated from the standardized lexicon-grammar table C1D

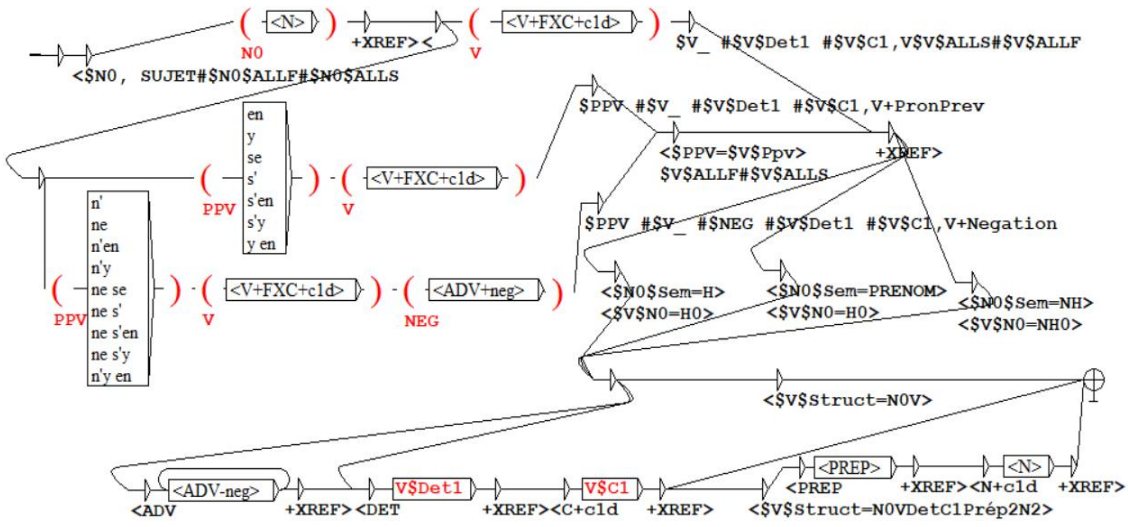

Figure 6. The syntactic grammar "C_c1d.nog" for the standardized lexicon-grammar table C1D

4. RESULTS AND DISCUSSION

To evaluate our linguistic resources constructed in the previous section, consisting of the dictionary "C_c1d.dic" (see Figure 5) and the syntactic grammar "C_c1d.nog" (see Figure 6), we analyzed a few sentences containing continuous or discontinuous frozen expressions of the lexicon-grammar table C1D. In particular, the study of sentences, exemplified by the analysis of "*Luc abandonne fréquemment la partie*" (Luc frequently gives up), illustrates the effectiveness of the constructed grammar. Figure 7 visually encapsulates the annotations derived from this evaluation, showing the remarkable ability of the grammar to

recognize and annotate the discontinuous frozen expression "*abandonner la partie*" (to give up). This result not only underscores the model's ability to navigate linguistic complexity but also highlights its versatility in capturing nuanced language structures. The successful handling of discontinuity represents a valuable contribution to the robustness of the lexicon-grammar approach, which holds promise for applications in NLP and linguistic analysis tasks.

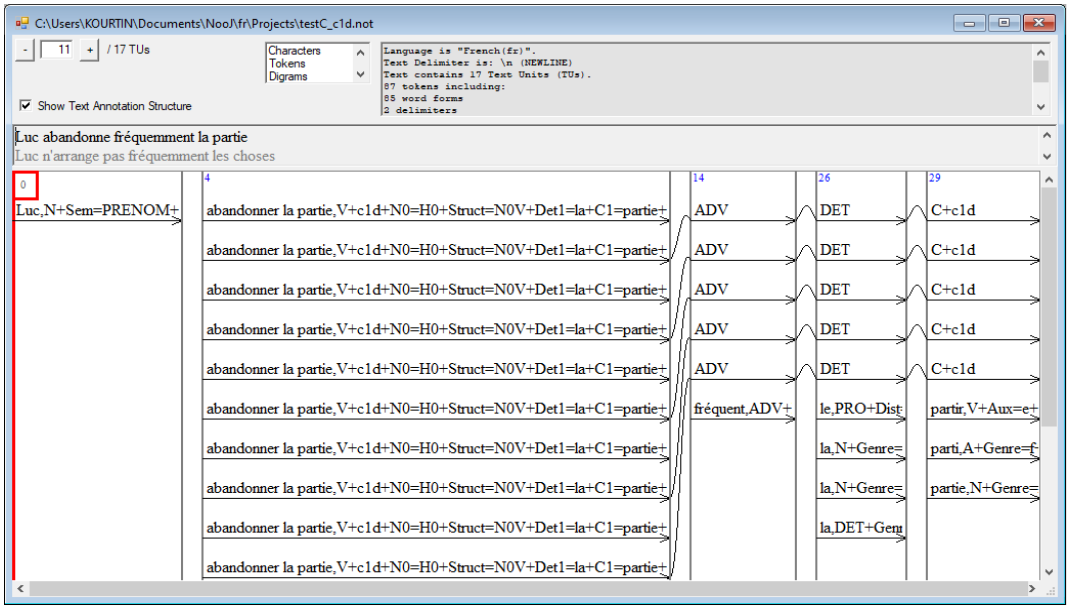

Figure 7. The annotations of the sentence "*Luc abandonne fréquemment la partie*"

The results of this study show a successful standardization process applied to lexicon-grammar tables for the French language, aligning their properties with compatibility standards used in linguistic platforms. The standardized lexicon-grammar tables were seamlessly integrated into the NooJ platform, demonstrating the practical application of this standardization. The automated generation of dictionaries from the lexicon-grammar tables, facilitated by a dedicated program, streamlined the integration process into NooJ. This automation not only increases efficiency but also ensures a consistent and coherent representation of the language's lexicon within the linguistic environment. Subsequent evaluations involved the creation of syntactic grammars that took into account the full range of grammatical, syntactic, and semantic properties from the associated dictionaries. The results of testing these generated dictionaries underscore the effectiveness of the standardized lexicon-grammar tables in capturing the intricate nuances of the language. Overall, the standardization effort not only automated the integration process but also leveraged the benefits of intuitiveness and simplicity in encoding grammatical, syntactic, and semantic properties within the linguistic platform. This represents a significant step toward promoting accessibility and usability in linguistic analysis, paving the way for enhanced applications in NLP and linguistic research.

5. CONCLUSION AND PERSPECTIVES

The lexicon-grammar tables are used to describe the language lexicon through readable and intuitive tables for human manual editing. In this work, we have standardized these lexicon-grammar tables for the French language by standardizing their properties to be compatible with those used in the linguistic platforms. Then, we implement the standardized lexicon-grammar tables into the NooJ platform as an example. For that, we automatically generated their dictionaries by using the program that automatically generates the NooJ dictionaries from lexicon-grammar tables. Finally, to test the efficiency of some generated dictionaries, we have created some syntactic grammars that take into account all the grammatical, syntactic, and semantic entries' characteristics contained in the associated dictionary.

This standardization made it possible to automate the process of integrating these lexicon-grammar tables into the linguistic environment, thus profiting from their advantages such as intuitiveness and simplicity while entering the grammatical, syntactic, and semantic properties of the language's lexicon. The standardization, therefore, serves as a key enabler for realizing the full potential of lexicon-grammar tables,

making them more accessible and usable in a wide range of linguistic applications. In the future, we will extend this standardization work to other languages that already have non-standardized lexicon-grammar tables. Also, we will create lexicon-grammar tables for the Arabic language, as there is a lack of research in this area.

REFERENCES

[1] M. Gross, *Methods in syntax: regime of completion constructions,(in French: Méthodes en syntaxe: régime des constructions complétives)*, Universite de Paris-Vincennes, 1975.

[2] 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.

[3] 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*, 1984, pp. 275–282.

[4] M. Gross, "Fixed sentences in French (in French: les phrases figées en français)," *L'information grammaticale*, vol. 59, no. 1, pp. 36–41, 1993, doi: 10.3406/igram.1993.3139.

[5] M. Gross, "A local grammar of the expression of feelings (in French: une grammaire locale de l'expression des sentiments)," *Langue Francaise*, vol. 105, no. 1, pp. 70–87, 1995, doi: 10.3406/lfr.1995.5294.

[6] C. Leclère, " Organization of the lexicon-grammar of French verbs (in French: organisation du lexique-grammaire des verbes français)," *Langue Francaise*, vol. 87, no. 1, pp. 112–122, 1990, doi: 10.3406/lfr.1990.6333.

[7] Y. Y. Mathieu, "Semantic predicate interpretation of argument structures. Feeling: an application to psychological verbs (in France: Interpretation par predicats semantiques de structures d'arguments. Feeling : une application aux verbes psychologiques)," Ph.D. disseration, Fundamental computing, Paris 7 Univ., Paris, France, 1994.

[8] Y. Y. Mathieu, " A lexical and semantic knowledge base of psychological verbs (in France: une base de connaissances lexicale et sémantique des verbes psychologiques)," *LINX*, vol. 34, no. 1, pp. 307–313, 1996, doi: 10.3406/linx.1996.1438.

[9] Y. Y. Mathieu, "A semantic lexicon for emotions and feelings," in *AAAI Spring Symposium - Technical Report*, pp. 89–93, 2005.

[10] Y. Y. Mathieu, "From linguistic descriptions to their representations (in French: des descriptions linguistiques à leurs représentations),", Ph.D. disseration, Dept. Linguistique, Paris-Diderot - Paris VII Univ., Paris, France, 2009.

[11] M. Silberztein, "Complex annotations with NooJ," in *Proceedings of the 2007 International NooJ Conference*, Cambridge Scholars Publishing, pp. 214-227, 2008.

[12] M. Silberztein, *Formalizing Natural Languages: The NooJ Approach*, ISTE Editions, pp. 1-333, London, UK, 2016, doi: 10.1002/9781119264125.

[13] E. Tolone, " Lexique-grammaire tables in TAL format (in France: les tables du lexique-grammaire au format TAL)," in *Actes de MajecSTIC 2009*, Avignon, France, p. 8, Nov. 2009.

[14] E. Tolone and B. Sagot, "Using lexicon-grammar tables for French verbs in a large-coverage parser," in *Human Language Technology. Challenges for Computer Science and Linguistics*, Z. Vetulani, Ed., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2011, pp. 183–191. doi: 10.1007/978-3-642-20095-3_17.

[15] E. Tolone, "Maintenance of the lexicon-grammar: definitional formulas and classification tree," *TAL Traitement Automatique des Langues*, vol. 52, no. 3, pp. 153–190, 2011.

[16] E. Tolone, "Syntactic analysis using French lexicon-grammar tables (in French: analyse syntaxique à l'aide des tables du lexique-grammaire du français)," *Lingvisticæ Investigationes*, vol. 35, no. 1, pp. 147–151, Jan. 2012, doi: 10.1075/li.35.1.07tol.

[17] M. E. Hannach, "Syntax of psychological verbs in Arabic (in French: syntaxe des verbes psychologiques en Arabe)," Ph.D. disseration, Dept. Linguistique, Paris 7 Univ., Paris, France, 1988.

[18] M. El Hannach, "Arabic lexicon-grammar: class of qualitative verbs (in French: lexique-grammaire de l'Arabe : classe des verbes qualitatifs)," *Linguistica Communicatio*, Vol. I, Fasic. 1, pp. 9-18, Fasic. 2, pp.31-41, Fès, 1989.

[19] M. El Hannach, "Linguistics tools to develop an Arabic syntax analyzer," in *The 2013 9th International Conference on Innovations in Information Technology, IIT 2013*, 2013, pp. 248–252. doi: 10.1109/Innovations.2013.6544427.

[20] A. Amzali, A. Kourtin, M. Mourchid, A. Mouloudi, and S. Mbarki, "Lexicon-grammar tables development for arabic psychological verbs," *Communications in Computer and Information Science*, vol. 1153, pp. 15–26, 2020, doi: 10.1007/978-3-030-38833-1_2.

[21] A. Amzali, M. Mourchid, A. Mouloudi, and S. Mbarki, "Arabic psychological verb recognition through nooj transformational grammars," *Communications in Computer and Information Science*, vol. 1389, pp. 74–84, 2021, doi: 10.1007/978-3-030-70629-6_7.

[22] A. Amzali, A. Kourtin, M. Mourchid, A. Mouloudi, and S. Mbarki, "Syntactic analysis of sentences containing arabic psychological verbs," *Communications in Computer and Information Science*, vol. 1520, pp. 51–61, 2022, doi: 10.1007/978-3-030-92861-2_5.

[23] 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, pp. 65–76, 2020, doi: 10.1007/978-3-030-38833-1_6.

[24] 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, pp. 28–38, 2022, doi: 10.1007/978-3-030-92861-2_3.

[25] S. Vietri, "The construction of an annotated corpus for the analysis of Italian transfer predicates," *Lingvisticæ Investigationes*, vol. 37, no. 1, pp. 69–105, Jan. 2014, doi: 10.1075/li.37.1.03vie.

[26] S. Vietri, "The lexicon-grammar of Italian idioms," in *Proceedings of Workshop on Lexical and Grammatical Resources for Language Processing*, Dublin, Ireland: Association for Computational Linguistics and Dublin City University, pp. 137–146, 2014, doi: 10.3115/v1/W14-5817.

[27] S. Vietri, *Idiomatic Constructions in Italian: a lexicon-grammar approach*. John Benjamins Publishing Company, Lingvisticae Investigationes Supplementa, vol. 31, 2014.

[28] S. Vietri, "Locatum and location verbs in Italian: a lexicon-grammar analysis," *Lingvisticæ Investigationes*, vol. 44, no. 2, pp. 153–203, Dec. 2021, doi: 10.1075/li.00061.vie.

[29] R. Guarasci, E. Damiano, A. Minutolo, M. Esposito, and G. De Pietro, "Lexicon-Grammar based open information extraction from natural language sentences in Italian," *Expert Systems with Applications*, vol. 143, no. 112954, Apr. 2020, doi: 10.1016/j.eswa.2019.112954.

[30] P. A. Machonis, "Compositional phrasal verbs with up: direction, aspect, intensity," *Lingvisticae Investigationes*, vol. 32, no. 2, pp. 253–264, 2009, doi: 10.1075/li.32.2.08mac.

[31] P. Machonis, "Neutral phrasal verbs in english," *Cahiers du CENTAL*, Presses universitaires de Louvain, pp. 229-237, 2010.

[32] W. Koza and N. R. 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.

[33] C. Mota, L. Chacoto, and A. Barreiro, "Integrating the lexicon-grammar of predicate nouns with support verb fazer into Port4NooJ," *Communications in Computer and Information Science*, vol. 811, pp. 29–39, 2018, doi: 10.1007/978-3-319-73420-0_3.
[34] A. Barreiro, C. Mota, J. Baptista, L. Chacoto, and P. Carvalho, "Linguistic resources for paraphrase generation in portuguese: a lexicon-grammar approach," *Language Resources and Evaluation*, vol. 56, no. 1, 2022, doi: 10.1007/s10579-021-09561-5.

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 NLP. Her research interests include software engineering, artificial intelligence, computational linguistics, and automatic NLP. 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 NLP. Her research interests include software engineering, artificial intelligence, computational linguistics, and automatic NLP. 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 NLP, web semantics, and information systems. He can be contacted at email: mourchidm@hotmail.com.

Abdelaziz Mouloudi Professor at the Department of Computer Science at the Faculty of Sciences, Ibn Tofail University. He does research in artificial neural networks, artificial intelligence, computer security, and reliability. His research interest score is 323.0, the number of citations is 495, and his h-index is 13. His research interests include evolutionary algorithms encryption cryptography, information security, and data security. He can be contacted at email: mouloudi_aziz@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 NLP. 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 NLP. He can be contacted at email: mbarkisamir@hotmail.com.