

# Formalization of Speech Verbs with NooJ for Machine Translation: the French Verb *accuser*

Jouda Ghorbel

Language Laboratory and Automatic Processing, University of Sfax, Tunisia

joudaghorbel@yahoo.fr

## Abstract

The mediocrity of sentences generated by online translators (Jacqueline, 1998; Hutchins, 2001) prompts us to try to find a solution to have more reliable translations. This is a very difficult task due to the ambiguity of natural languages and especially the deficiencies of translation systems in terms of syntactic and semantic knowledge. How can we make automatic translation more reliable and unambiguous? Our main objective will be to generate a text where the translation of French verbs into Arabic will be without ambiguities. In this contribution, we attempt to formalize a particular class of verbs, namely the so-called verbs of speech. We shall limit ourselves to the treatment of the verb *accuser* ‘to accuse’ as presented in the Dubois & Dubois-Charlier (1997) electronic dictionary, *Les verbes français*. We shall take this verb as a prototype to show how NooJ can perform a reliable machine translation and generate a good text without ambiguities.

## 1 Introduction

The process of creating translation machines, capable of properly translating verbs, takes place in two stages: a theoretical stage and an application stage. The first phase allows analyzing and interpreting the phenomena studied which is then developed in the second phase whose role is to produce an automatic translation. In this contribution, we will try to automatically treat a class of verbs, namely the so-called verbs of speech. To do this we have adopted the NooJ platform for the syntactic description of the French verb *accuser* ‘to accuse’ as an example of speech verbs. Our two

main objectives, which are in the realm of applied linguistics, include:

- The production of a system of analysis and recognition of the syntactic patterns of the verb *accuser* according to the classification of French verbs.
- An adequate and reliable machine translation and the generation of sentences into Arabic.

Our work will therefore be divided into four parts: Derivational and inflectional formalization;

Syntactic formalization;

Implementation of the verb *accuser* in NooJ for machine translation;

Automatic translation and generation of sentences in Arabic.

## 2 Derivational and inflectional formalization

The NooJ software has its own tools for automatic verb analysis and processing (Silberztein, 2003), so we need to formalize the linguistic data, in order that the program can automatically analyze and process the verb *accuser* in all of its various shades of meaning in our corpus, and then accurately translate them.

We will therefore create the necessary paradigms to link each derivative or verb conjugated to its infinitive form.

### 2.1 Creation of derivational paradigms

We have chosen our verb example from Dubois & Dubois-Charlier’s (1997) electronic dictionary, *Les verbes français* (LVF). This work presents derivational codes which can be adjectival, such as

in *-able* and *-ant*, or nominal derivations, as in *-age*, *-ment*, *-ion*, *-eur*, and *-ure*.

The verb *accuser* in the LVF has only one nominal derivative: *accusateur* ‘accuser’, we have created in NooJ its paradigm that we called

N1 = *accuser*<B2> *ateur* / N.

The LVF dictionary clearly indicated the different derivatives for each verb, but their inflections were not mentioned in the feminine nor in the plural, so we had to create these inflectional paradigms so that NooJ could recognize the inflected derivative forms: *accusateur* = <E> / m + s | s / m + p | <B3> *rice* / f s + | <B3> *rices* / f + p.

## 2.2 Creation of inflectional paradigms

NooJ will automatically be able to recognize the conjugated forms of a verb only when describing the conjugation models indicated in LVF with the NooJ inflectional operators. For that, Max Silberstein<sup>1</sup> matched the conjugation codes of NooJ.

Example: the verb *aimer* ‘to love’ is inflected as:

<i>Aimer</i>	=	<E>/INF		<B2>ant/G	
		<B2>é/PP+m+s		<B2>ée/PP+f+s	
		<B2>és/PP+m+p		<B2>ées/PP+f+p	
		<B>/IP+s+2		<B2>ons/IP+p+1	
		<B>z/IP+p+2		<B>/PR+s+1	
		<B>s/PR+s+2		<B>/PR+s+3	
		<B2>ons/PR+p+1		<B>z/PR+p+2	
		<B>nt/PR+p+3		<B>/S+s+1	
		<B>s/S+s+2		<B>/S+s+3	
		<B2>ions/S+p+1		<B2>iez/S+p+2	
		<B>nt/S+p+3		ais/C+s+1	
		ais/C+s+2		ait/C+s+3	
		ions/C+p+1		iez/C+p+2	
		aient/C+p+3.			

For this inflectional paradigm of the verb *aimer* all the tenses, and moods have been described with all the personal pronouns using NooJ operators and thanks to this model we can conjugate a large number of verbs, such as *accuser* our example verb.

NooJ can now recognize all the conjugated occurrences of the verb *accuser* in our corpus, lemmatize them and link them to the list of the various uses.

## 3 Syntactic formalization

In this phase, we describe the syntactic schemas, which are written in the form of codes, replace the codes with the verb and its arguments, and assign them semantic features based on their syntactic schemas. Syntactic schemas are defined by the nature of the constituents of the sentence, their properties and their relations, and by the words of the lexicon which enter into the various types of constituents.

### a. *Accuser* direct transitive verb

- [T11b0] CONS = T11b0 + N0VN1PREPN2 + N0Hum + V + N1Hum + N2Abst + N2Vinf + PREP = « de »

Ex: *La mère accuse son enfant d’avoir oublié ses devoirs.* ‘The mother reproaches her child for having forgotten her homework’.

- [T1907] CONS = N0VN1 Nhum + VT + N1Hum N1Abst+ N1Conc

Ex: *On accuse le sort.* ‘We accuse fate’.

### b. *Accuser* pronominal verb

- [P10B0] CONS = N0seVPREPN1 + N0Hum + VP + N1Hum + N1Abst + N1Vinf + PREP = “de”

Ex: *L’enfant s’accuse d’avoir menti* ‘The child reproaches himself for having lied’.

After this phase of derivational, inflectional and syntactic formalization of the verb *accuser* with NooJ operators, we proceed to the implementation of these formal data within NooJ.

## 4 Implementation of the verb *accuser* in NooJ for automatic translation

In this phase of formalization we show how to integrate the verbal input *accuser* in NooJ for automatic translation. For this, we created a bilingual French Arabic dictionary and formal grammars for the different constructions of the verb *accuser*.

<sup>1</sup> Author of the software NooJ

#### 4.1 Creation of a bilingual Arabic dictionary

This phase of implementation of the verb in a bilingual French-Arabic dictionary aims first of all to reformulate the information of the LVF in terms of NooJ operators. This operation consists of applying the dictionary to automatically translate the text into Arabic. We added the Arabic translation to each verb and all the other words in order to generate sentences into Arabic without ambiguities.

In this phase, we reformulate the information of the LVF in connection with the verb *accuser* and apply this dictionary to the French-Arabic machine translation (Figure 1).

Accuser,  
V+COM+AUX=AVOIR+FLX=AIMER  
+CONS=T11b0+N0VN1PREPN2+N0Hum  
+V+N1Hum+PREP="de"+N2Abst+Vinf  
+CONS=B10b0+N0seVPREP1+N0Hum  
+V+PREP="de"+N1Abst+Vinf  
+Sens=reprocher à+AR="تلم"

Accuser,  
V+COM+AUX=AVOIR+FLX=AIMER  
+CONS=T1907+N0VN1+N0Hum  
+V+N1Hum+N1Abst  
+Sens= invectiver+AR="ادان"

criminel, N+FLX=S\_DE+AR="المجرم"  
juge, N+FLX=S\_0+AR="القاضي"  
mère, N+FLX=F\_S+AR="الام"  
enfant, N+FLX=S\_0+AR="ابنها"  
courir, V+FLX=COURIR+AR="الركض"  
de, PREP+AR="على"  
mentir, V+FLX=SENTIR+AR="الكذب"

Figure 1: Excerpt from the dictionary *accuser.dic*

#### 4.2 The creation of formal grammars for the different constructions of the verb *accuser*

For a reliable automatic translation of sentences containing the verb *accuser*, we tried to create formal grammars (Figure 2 for T11b0 and P10b0 constructions; Figure 3 for T1907 construct) to remove the ambiguities of the various syntactic constructions.

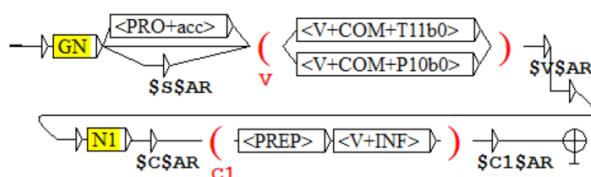


Figure 2: The formal grammars of the constructions T11b0 and P10b0.

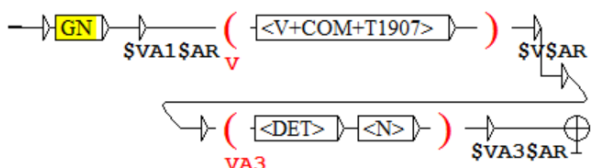


Figure 3: The formal grammars of the construction T1907

Boons et al. (1976) claim that the meaning of the verb is related to the type of subject and the complement (human, concrete, non-animated, etc.); we created formal grammars that take into consideration the type of the object and the type of the complement. The verb *accuser* has three different syntactic constructs: T11b0 / T1907 / P10b0, all have a human subject with a different complement:

- **[T11b0]** N0VN1PREPN2 + N0Hum + VT + N1Hum + N2Abst + N2Vinf + PREP = “de”

Ex: *La mère accuse son enfant de courir.* ‘The mother accuses her child of running’.

- **[P10B0]** = N0seVPREP1 + N0Hum + VP + N1Hum + N1Abst + N1Vinf + PREP = “de”

Ex: *On s'accuse d'avoir menti.* ‘One reproaches oneself for having lied’.

- **[T1907]** = N0VN1 NHum + VT +N1Hum N1Abst + N1Conc

Ex: *Le juge accuse le criminel.* ‘The judge accuses the criminal’.

For the development of a formal grammar capable of correctly recognizing and translating the sentences which contain the verb *accuser*, we have to add other electronic dictionary resources and dictionaries for the detection of lexical units. We used the electronic dictionaries already integrated into the NooJ platform, such as:

Before	Seq.	After
s'est-il exclamé. "On président de l'Assemblée nationale, "fils de Le Pen" et de la fondation. En somme, le 24 mars 1999 qui fit 39 morts. le nombre de meurtres qu' M. Bartolone ne décolère pas. à la radicalisation du mouvement, le 28 mars contre Michel Rocher. qui que ce soit. Quand Avec une telle argumentation, on veut tuer son chien, on jeunes Sénégalais de 20 à 25 ans premier secrétaire du PS a-son éthique personnelle et qu' positionnement de M. Henno qu' le rappeur Monsieur R. qu' influence, je dis non. Si des faits de maltraitance dont William Bourdon. Huit citoyens birmanes, de l'état d'urgence, Politique - Société TITRE : Les témoignages refuge pour les survivants tutsis. TITRE : Au procès Matra-Thomson, la nation entre 1997 et 2002 " et d'entrée des migrants malades.	nous accuse de parler/ادان qui accuse le président/ادان l'avoit accusé d'antisémitisme/ادان la France accuse un certain/ادان Il accuse sa hiérarchie d'avoir pris/لام elle l'accuse d'avoir commis/لام Il accuse la majorité/ادان des proviseurs accusent des adultes/ادان On l'accuse d'avoir favorisé un laboratoire/لام son mari l'accuse d'avoir mis le doigt/لام vous accuse de racisme/ادان l'accuse de la rage/ادان qui accusent le prêtre/ادان t-il accusé le gouvernement/ادان on l'accuse d'avoir agi/لام il accuse d'avoir/ادان ils accusent d'incitation/ادان vous m'accusez d'avoir monté un service/لام ils accusaient leurs parents/ادان qui accusaient l'entreprise/ادان il accusait la gauche/ادان qui accusent l'armée/ادان Les témoins accusent les militaires d'avoir facilité les enlèvements/لام la défense accuse le plaignant d'avoir usurpé son titre/لام l'accuse d'avoir creusé la dette/لام L'association accuse le ministre/ادان	librement et même parfois un de l'UMP de vouloir et d'homophobie. Monté sur retard sur le marché international à la légère l'incendie . f.<DATE="050304"> § <PAGE="article_400155"> RUBRIQUE : France du PS d'essayer de - parents d'élèves, enseignants - et vidéo, VDM, à Courbevoie, où dans les fesses de ses et de nationalisme. Comment réagissez ", fait remarquer, avant toute chose sont d'anciens prostitués, prêts , vendredi 12 août, de " laisser les contre l'intérêt des actionnaires " fait une campagne d'extrême à la haine pour son parallèle qui n'était pas . Les six accusés du procès de travail forcé, ont retiré d'être " à l'unisson française au Rwanda Cinq hommes de Tutsis par les milices de victime LE PROCÈS ouvert de " 300 milliards d'euros " alors de l'intérieur d'avoir

Figure 4: Extraction from the corpus Le Monde and translation of the verb accuser into Arabic

- The delaf.nod which contains all the inflected forms of the single words corresponding to the lexical entries of the DELAS (Silberztein 1993);
- *Le DM* dictionary of French words (Trouilleux, 2011). This linguistic resource is integrated into the NooJ platform and it contains 67,997 entries composed of determiners, pronouns, prepositions, conjunctions, numerals, adverbs, nouns, interjections, adjectives and verbs.

The grammar we have constructed must know all possible constructions without any ambiguity. For this reason, we chose to create for each syntactic construct its path independently of the other constructions to remove all the ambiguities using the precise semantic traits (concrete, abstract, human etc.) of each argument.

## 5 Automatic translation and generation of verbal predicates into Arabic

The automatic translation process is applied to the communication predicates obtained after the analysis and recognition phases.

We have already tried to implement our process of analysis and recognition of syntactic patterns on a newspaper corpus of *Le Monde*. We obtained results where the patterns and verbs found are disambiguated and annotated simply by the appropriate syntactic constructions (Figure 4).

We can see that the formal grammars allowed us not only to automatically recognize the different

constructions, but also to translate the verb automatically into Arabic. However, our system not only allows us to translate a single verb but also the entire sentence.

Therefore, we added the grammars of translation (\$v\$AR) to each formal grammar. And we obtained the translation into Arabic of the different sentences that contain the verb *accuser* in its various constructions:

- Translation of *accuser*+T11b0+AR= لام (Lāma) 'To reproach'

Ex: *La mère accuse son enfant de courir.*

→ الأم تلوم ابنها على الركض

- Translation of *accuser*+P10b0+AR = لام (Lāma)

Ex: *On s'accuse de mentir.*

→ تلوم أنفسنا على الكذب

- Translation of *accuser*+T1907+AR= أدان (adāna) 'to accuse'

Ex: *Le juge accuse le criminel.*

→ أدان المجرم القاضي

Our system has succeeded in automatically translating the sentences into Arabic taking into account the meaning of the verb which varies according to the construction.

The software has thus made it possible to generate sentences naturally into Arabic without syntactic and semantic ambiguities.

The naturalness of the generated text is due to an automatic semantic syntactic analysis of open corpus, based on a broad description of the vocabulary. Our example adds to the examples of Silberztein (2015, 2016) to confirm that NooJ brings a significant qualitative leap for text generation.

## 6 Conclusion

We tried in this contribution to formalize the verb *accuser* as an example of verbs of speech, and to integrate it into the NooJ software. Thanks to the linguistic richness of the LVF verbal entries which helps the computer tools make a syntactic, semantic and morphological analysis of the verbs, we succeeded in automatically recognizing, extracting, and processing cases of *accuser* in a corpus of considerable size, the newspaper *Le Monde*.

This formalization was needed in order to obtain a reliable automatic translation. Thus, we created formal grammars to remove the ambiguity of the syntactic construct, since the meaning of the verb depends on the type of subject and the complement (human, concrete, abstract, etc.).

These grammars have led to the automatic recognition of syntactic constructions, which in turn removes ambiguities and generates sentences into

Arabic taking into account the meaning of the verb in its original French context.

## References

- Jean-Paul Boons, Alain Guillet and Christian Leclère. 1976. *La structure des phrases simples en français. Constructions intransitives*. Genève : Droz.
- Jean Dubois. 1979. *Dictionnaire de français langue étrangère*. Niveau II. Paris: Larousse.
- Jean Dubois, Françoise Dubois-Charlier. 1997. *Les Verbes Français*. Paris: Larousse-Bordas (diffuseur exclusif).
- John W. Hutchins.2001. Machine Translation over fifty years. *Histoire épistémologique du langage*. Volume 23 Numéro 1 pp 7-31
- Jacqueline Léon. 1998. Les débuts de la traduction automatique en France (1959- 1968): à contretemps?, *Modèles Linguistiques*, tome XIX, fascicule 2, pp 55-86.
- Max Silberztein. 2003. *NooJ Manual*. Available at: [www.nooj4nlp.net](http://www.nooj4nlp.net)
- Max Silberztein. 2015. *La formalisation des langues: l'approche de NooJ*. ISTE Editions, London,UK.
- Max Silberztein. 2016. *Formalizing Natural Languages: The NooJ Approach*. Wiley ISTE, London,UK.
- François Trouilleux. 2011. *Le DM*, A French dictionary for NooJ. In *Automatic Processing of Various Levels of Linguistic Phenomena: Selected Papers from the NooJ 2011 International Conference* (eds. K. Vučković, B. Bekavac and M. Silberztein). Cambridge Scholars Publishing, pp 16-28.