Revue de Traduction et Langues Volume21 Numéro1/2022
Journal of Translation Languages مجلة الترجمة واللغات
ISSN (Print): 1112-3974 EISSN (Online): 2600-6235

Une Approche Linguistique pour l'Analyse, la Reconnaissance et la Traduction Automatiques des Verbes du Domaine Administratif et de Droit

A Linguistic Approach toThe Analysis, Recognition and Automatic Translation of Administrative and Legal Verbs

Cheikhrouhou Hajer
Université de Sfax FLSHS- Tunisie
hajer.cheikhrouhou369@gmail.com
Laboratoire de Langage et Traitement Automatique
0000-0001-6481-8205
Bouattour Mohamed
Université de Sfax- FLSHS- Tunisie
mohamedbouattour@yahoo.fr
Laboratoire de Langage et Traitement Automatique
0000-0002-0489-5302

Comment citer cet article :
Cheikhrouhou, H., & Bouattour, M. (2022). Une Approche Linguistique pour l'Analyse, la Reconnaissance et la Traduction Automatiques des Verbes du Domaine Administratif et de Droit. *Revue Traduction et Langues21* (1), 28-55.

Reçu : 23/05/2022 ; **Accepté :** 21/08/2022, **Publié :** 31/08/2022

L'auteur correspondant : Cheikhrouhou Hajer

Keywords

Administration and law field; LVF; Machine Translation; NooJ; Objects classes; Verbal polysemy

Abstract

Due to globalization and with the advent of the new digital era, economic, social, and mostly administrative sectors are now required to participate in the digitization and automation of data. This will make several tasks that were previously totally performed by humans easier. Hence, in this context, the present study investigated the application of machine translation (MT) in the administrative field. Using a semantic-syntactic approach and the French Verbs (LVF) dictionary by Jean Dubois and Françoise Dubois-Charlier, we analyzed the verbs used in the administration and legal fields. Following this initial theoretical stage, and using the NooJ platform, we deployed the acquired linguistic data for the construction of a French-Arabic machine translation application. To implement this application, we will design formal grammars for the automatic analysis and recognition of the many semantic-syntactic schemes defining this class of verbs, as well as a bilingual dictionary of legal and administrative verbs. The utmost objective of this research work is the achievement of an accurate and trustworthy translation.

In the first theoretical part, we will identify the linguistic characteristics of the different verbs in the administrative and legal fields. At this first level of our approach, we will examine the syntactic constructions of these verbs and then, we will try to delimit the patterns or the distinctive semantic-syntactic schemes of the verbs used in a very specific discourse, which is the administrative discourse.

In the second stage of automatic processing, we will try to create a computer process which can:

• Automatically analyze and recognize administrative verbs using the NooJ platform.

• Properly translate the verbs studied in Arabic.

For the creation of this MT process, the interest of which mainly concerns administrative documents, for social (divorce, inheritance, etc.) and even economic reasons, we will:

a/ Integrate legal and administrative verbs into a bilingual French-Arabic electronic dictionary, describing them according to NooJ operators: inflection, conjugation, derivation and syntax operators, etc. In this step, we created:

• Paraderivational.nof derivational paradigms.

• "paraflexion.nof" bending paradigms.

• "paradigmflexionnel.nof" conjugation paradigms

• The syntactic formalization where we rewrote the syntactic schemes of the LVF.

• The translation of each verbal entry into Arabic.

Exp : accuser, V+Emploi=01+AUX=AVOIR+FLX=CHANTER+CONS=T11b0+N0VN1PREP N2+N0Hum+V+N1Hum+N2Abst+N2Vinf+PREP="de"+DRV=AV1e : ARTISTE +DRV=AV2: COUSIN+DRV=DN15: DIRECTOR+DRV=DN15a: TABLE+DOM=DRO

+CLASS=F2c+OPER="ict qn D offence"+LEXI=1+AR="اتهم"

b/ Schematize the semantic-syntactic patterns by the formal grammars of NooJ, to automatically recognize the legal verbs used in an administrative corpus.

The formal grammars used are called syntactic disambiguation grammars (Silberztein, Max 2015). In this phase, we produced disambiguation grammars that

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

allowed us to analyze only the requested syntactic constructions. The disambiguation grammar is necessary and important, not only for the automatic analysis and recognition of the correct syntactic constructions, but also for producing a correct and reliable translation.
c/ Create an automatic translation grammar that will allow us to obtain a suitable, reliable translation, which will respect the standards of the target language: Arabic.
d/ Find solutions for the disambiguation of polysemic predicates using Gaston Gross's "object class" theory which allowed us to refine the recognition grammars and subsequently obtain encouraging results, where the verbs polysemous are suitably translated into Arabic.
The originality of this study emanates from the fact that it provides, not only a linguistic description of the verbs of the administrative and legal domain, but especially of the integration of the various data in the NooJ software, in order to facilitate the treatment and the automatic translations of the two languages in question.

Mots clés

Classes d'objets ; Domaine d'administration et de droit ; LVF ; NooJ ; Traduction automatique.

Résumé

La mondialisation et la nouvelle ère numérique ont imposé à tous les secteurs économiques, sociaux et surtout administratifs d'être impliqués dans la numérisation et l'automatisation des données pour faciliter certaines tâches effectuées, auparavant, totalement par l'être humain.
Dans ce cadre, nous allons étudier le cas de la Traduction Automatique (TA) utilisée dans le domaine administratif. Dans cette étude, nous allons essayer d'analyser les verbes relevant du domaine d'administration et de droit en utilisant le dictionnaire des Verbes Français (LVF) de Jean Dubois et Françoise Dubois-Charlier et en adoptant une approche sémantico-syntaxique. Après cette première étape théorique, nous allons employer ces données linguistiques pour la création d'une application de TA français-arabe en utilisant la plate-forme NooJ. Pour la réalisation de cette application, nous allons construire un dictionnaire bilingue des verbes de droit et d'administration et nous allons créer des grammaires formelles d'analyse et de reconnaissance automatiques des différents schèmes sémantico-syntaxiques caractérisant ce type de verbes. Le but de cette recherche est d'obtenir une traduction correcte et fiable.
L'originalité de cette étude vient du fait qu'elle fournit, non seulement une description linguistique des verbes du domaine administratif et de droit, mais surtout de l'intégration des différentes données dans le logiciel NooJ, en vue de faciliter le traitement et la traduction automatiques des deux langues en question.

1. Introduction

Le discours administratif, quelle que soit sa situation d'énonciation, est une configuration de structures syntaxiques, dont la signification sémantique répond à la visée pragmatique qui la motive. En nous inspirant de la théorie des classes d'objets de Gaston

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Gross, nous postulons que le choix lexical est déterminant, non seulement au niveau syntaxique, mais surtout au niveau logico-sémantique et pragmatique. L'étude des verbes qui relèvent du domaine d'administration et de droit nous permettra de montrer que la visée pragmatique, la valeur illocutoire, opère d'abord au niveau du choix du verbe, ensuite au niveau de l'adéquation de celui-ci avec le contexte de son énonciation.

Le discours administratif ou de droit peut, à juste titre, être considéré comme langue de spécialité, en ce sens que les récurrences tant terminologiques que phrastiques pourraient constituer un paradigme spécifique (bien déterminé).

Nous allons approcher les différentes distributions des verbes qui appartiennent à ce type de discours pour mieux définir les différentes attitudes des différents intervenants dans ce domaine (président, directeur, juge, avocat, maire, conseil municipal…) par rapport aux contenus propositionnels de leurs discours, en utilisant le dictionnaire des verbes français LVF de Jean Dubois et Françoise Dubois-Charlier (abrégés en JD & FDC, 1997).

Exemple 1 : Le juge a accusé Paul de ce vol.
On adopte un enfant étranger.
Le président a agréé Mr. Denis comme ambassadeur.

La traduction automatique (TA), adoptée dans les échanges internationaux entre les sociétés et les administrations, pose un grand problème au niveau de la fiabilité des résultats et du sens exact indiqué dans le document administratif, surtout en cas de polysémie verbale. C'est ce que nous proposons de montrer en exploitant le logiciel NooJ[1], en vue d'y apporter les solutions adéquates.

2. Problématique et Méthodologie

Dans cet article, nous allons étudier les verbes de type « accréditer, administrer, allèguer, certifier, contrôler, fédéraliser, légaliser, notarier, etc. » relevant du discours administratif, qui diffère des autres discours par le lexique employé, les constructions syntaxiques, la visée énonciative et pragmatique. Notre étude sera fondée sur deux niveaux : Linguistique et Traitement Automatique des Langues (TAL).

Le premier niveau est linguistique. Nous allons y effectuer une analyse sémantico-logique des verbes administratifs, tout en utilisant le dictionnaire des verbes français LVF. Pour ce faire, nous avons filtré les verbes qui appartiennent au domaine de l'administration et du droit, marqué dans LVF par « DRO » et qui sont au nombre de neuf-cent-quatre-vingt-quatre entrées verbales. Pour la classification des verbes français, Jean Dubois et Françoise Dubois-Charlier ont «établi une classification visant à tenir compte de

[1] NooJ est un environnement de développement linguistique qui permet de construire et de gérer des dictionnaires et grammaires électroniques à large couverture, afin de formaliser divers niveaux des langues naturelles: orthographe, morphologie flexionnelle et dérivationnelle, lexique de mots simples, mots composés et expressions figées, syntaxe locale et désambiguïsation, syntaxe structurelle et transformationnelle, sémantique et ontologies.in http://www.nooj-association.org

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

l'adéquation entre la syntaxe et l'interprétation sémantique, elle-même représentée par un schéma syntactico-sémantique, donné sous la rubrique "opérateur" qui, dans son classifieur initial, renvoie au verbe de base» (Dubois, Dubois-Charlier, 1997, P. 5).

À ce premier niveau de notre démarche, nous examinerons les constructions syntaxiques de ces verbes et nous essayerons par la suite de délimiter les patrons ou les schèmes sémantico-syntaxiques distinctifs des verbes employés dans un discours bien spécifique, qui est le discours administratif.

Le deuxième niveau est formel. Nous essayerons, dans cette étape de traitement automatique, de créer un processus informatique capable de :

- Analyser et reconnaitre automatiquement les verbes administratifs en utilisant la plate-forme NooJ.
- Traduire convenablement les verbes étudiés en langue arabe.

« La description de certains phénomènes linguistiques permet de construire des logiciels du TAL. Par exemple : si l'on dispose de la liste des mots d'une langue, on peut construire un vérificateur orthographique ; si l'on dispose de la liste des règles de conjugaison, on peut construire un conjugueur automatique (…) ; si l'on dispose d'un dictionnaire et d'une grammaire bilingue, on peut construire un traducteur automatique, etc » (Silberztein, 2015, P. 20).

Selon cette citation, la règle générale pour la réalisation d'une application de traduction automatique : il faut construire par les données linguistiques :

- un dictionnaire électronique bilingue
- une grammaire formelle bilingue

Pour la création de ce processus de TA dont l'intérêt concerne surtout les documents administratifs, pour des raisons sociales (divorce, héritage, etc.) et mêmes économiques, nous allons :

- Intégrer les verbes de droit et administratifs dans un dictionnaire électronique bilingue, français-arabe, en les décrivant selon les opérateurs de NooJ : opérateurs de flexion, de conjugaison, de dérivation et de syntaxe, etc.
- Schématiser les patrons sémantico-syntaxiques par les grammaires formelles de NooJ, pour reconnaitre automatiquement les verbes de droit employés dans un corpus administratif.
- Créer une grammaire de traduction automatique qui nous permettra d'obtenir une traduction convenable, fiable, et qui respectera les normes de la langue cible : l'arabe.
- Trouver des solutions pour la désambiguïsation des prédicats polysémiques.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

3. Etude Linguistique des Verbes du Domaine Administratif et de Droit

Notre étude linguistique des verbes administratifs et de droit se fonde essentiellement sur le LVF. Pour la réalisation de cette étude, nous avons commencé par filtrer les verbes appartenant à ce domaine.

Exemple 2: accuser (04)[2], affecter (02), congédier(01), déposer (04), fouiller (05), nommer (06), etc.

B542 DRO

	A	B	C	D	E	F	G	H	I	J	K
1	M	DOM	CLASSE	OPER	SENS	PHRASE	C	CONST	DER	N	L
2	abandonner 01	DRO	D2a	dat qc A qn	laisser,léguer	On a~ses biens à ses enfants,à une fondation.	1aZ	T13a0	---- -D --RB --	3*	1
3	abdiquer 01	DRO	S2e	dgrp qc trône	renoncer au trône	Le roi a~le trône.Le roi a~en 1580.	1aZ	T1300 A10	---- -- 2A-- --	-	2
4	abolir 01	DRO	F4b	ict total loi	annuler,supprimer	Le Parlement a~la peine de mort par une loi.	2aZ	T1308 P3000	---- -- RE-- --	-	2
5	abroger	DRO	F4b	ict total loi	annuler,abolir	On a~la loi,la peine de mort.Cette loi ne peut pas s'a~.	1fZ	T1308 P3000	22- -- RB-- --	-	3
6	abstenir 02(s)	DRO	U2c	dli soi D+inf	se dispenser de	On s'a~de voter.On s'a~lors du vote,de la discussion.	3gZ	P10b6	---- -- RF-- --	-	5
7	abuser 04	DRO	S2c	grp D qn p viol	violer	On a~de cette femme au cours de cette orgie.	1aZ	N1b	---- -- ---- --	-	5
8	accabler 01	DRO	F2b	ict+qt qn D abs	surcharger de	On a~les gens d'impôts.Les impôts a~les contribuables.	1aZ	T11b0	---- -- ---- --	-	2
9	accabler 04	DRO	C1i	loq.mvs qn	confondre,écraser sous	On a~l'accusé par ces témoignages.Les témoins a~P.	1aZ	T1107	--1 -- ---- --	-	5
10	accéder 05	DRO	E2c	ire A qc	devenir propriétaire	On a~à la propriété par les crédits.	1gZ	N1a	--1 -D 2C-- --	4S	5
11	accenser 01	DRO	D2a	dat A qn c cens	louer	Le seigneur a~un domaine à un fermier.	1aZ	T13a6	---- -1 ---- --	-I	5
12	accenser 02	DRO	N2a	mun qn D cens	donner bail à qn	Le seigneur a~un fermier d'une terre.Un fermier peut s'a~.	1aZ	T11b0 P10b0	---- -1 ---- --	-I	5
13	accepter 05	DRO	D3c	dat.val à loi	ratifier,approuver	On a~une loi.Ce projet ne peut s'a~dans cet état.	1bZ	T1300	2-- -- RB-- --	-	5
14	acclamer 01	DRO	S2f	grp qn tel p cris	nommer par cris	Les soldats a~P empereur.	1aZ	T1106	---- -- RB-- --	-	2
15	accomplir 01	DRO	R4b	fac durée d qc	réaliser,achever	On a~son temps de service,son mandat.	2aZ	T1300 P3000	-2- -1 ---- --	-	2
16	accorder 03	DRO	D2a	dat qc A qn	octroyer	On a~un congé à P.On s'a~un jour de repos.Le congé est a~.	1aZ	T13a0 P30a0	21- -- ---- --	2*	5
17	accréditer 01	DRO	S2f	grp qn c diplomate	donner lettres accrédit	Le gouvernement a~P comme ambassadeur auprès de cet Etat.	1aZ	T1106	-1- -- RBRA --	-	3
18	accuser 01	DRO	F2c	ict qn D délit	inculper	Le juge a~P d'un délit.P est a~,est un a~.	1aZ	T11b0	21- -- RBRB --	-	1
19	accuser 04	DRO	F2c	ict qn D crime	inculper qn de	On a~P de ce vol.	1aZ	T11b0	---- -- ---- --	-	5
20	acenser 01	DRO	D2a	dat A qn c cens	accenser	Le seigneur a~une terre à un fermier.	1aZ	T13a6	---- -1 ---- --	-I	7
21	acenser 02	DRO	D2c	abda qc A qn loué	louer	Le fermier a~une terre au seigneur.Cette terre peut s'a~.	1aZ	T13a6 P3006	---- -1 ---- --	-	5
22	acquérir 01	DRO	D2e	abda qc D p achat	acheter à	On a~cette maison d'un riche propriétaire.	3jZ	T13b0	---- -- 2NRA --	-	2
23	acquêter	DROt	D2e	abda qc D p achat	acquérir de	L'époux a~les biens du mort par l'héritage.	1bZ	T13b0	---- -D ---- --	2*	7

Figure 1. Extrait des verbes du domaine administratif et de droit « DRO[3] ».

[2] Dans LVF, les verbes qui ont plusieurs emplois sont numérotés (01, 02, 03…) comme le verbe « abattre » qui présente 10 emplois.

[3] Pour bien préciser le domaine, Dubois ajoutent des lettres minuscules qui indiquent les niveaux de Langues et régionalismes. Exemple : DROt la lettre « t » marque le niveau de langue littéraire.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

En analysant les verbes administratifs et de droit, nous relevons plusieurs caractéristiques.

3.1 Au Niveau Sémantique

Les prédicats relevant du domaine « DRO » appartiennent à maintes classes sémantiques, parmi lesquelles nous citons les classes suivantes.

3.1.1 Classe des verbes de communication C (111 entrées verbales)

Selon la classification de Dubois, les verbes de communication expriment l'échange de messages, d'informations, d'idées, de sons, avec ou sans le recours à la parole. Les classes génériques sont repérées par des lettres correspondant à une classe générale. Exemple 3 : C symbolise la classe de la communication. Ils sont divisés en quatre sous-classes sémantico- syntaxiques C1/C2/C3/C4 (Dubois, Dubois-Charlier, 1997, P. 3)

- C1 : « s'exprimer par un son, une parole »
- C2 : « dire/demander qc »
- C3 : « montrer qc »
- C4 : « dire ou montrer qc »

Exemple 4 : déclarer (03), juger (01), avouer (01), etc.

3.1.2 Classe des verbes de Don D (129 entrées verbales)

Cette classe regroupe les verbes qui peuvent être utilisés pour décrire des événements de transfert, de don ou de privation.

Exemple 5 : déléguer (01), donner (14), accorder (03), etc.

3.1.3 Classe des verbes de mouvement E (entrée/sortie) (113 entrées)

Les prédicats de cette classe indiquent le mouvement d'entrée et de sortie et ils sont caractérisés par un argument *locatif [Nloc]*.

Exemple 6 : verser (06), libérer (02), démissionner (01), etc.

3.1.4 Classe des verbes « frapper ou toucher qn » F (80 entrées)

Ces prédicats expriment frapper ou toucher qn avec ou sans instrument ou outil.

Exemple 7 : accabler (01), attaquer (03), électrocuter (02), etc.

3.1.5 Classe des verbes « être ou mettre qn dans tel état » H

Exemple 8 : obliger (01), régir (01), etc.

Nous signalons qu'il y a d'autres classes sémantiques pour les verbes du domaine administratif et de droit comme les classes de : U (union, réunion), T (transformation, changement), R (réalisation, mise en état), L (locatif), S (saisir, serrer, posséder), etc.

3.2 Au Niveau Syntactico-Sémantique

En étudiant les verbes du domaine administratif et de droit, nous constatons qu'ils peuvent avoir différentes constructions syntaxiques : transitive directe et indirecte,

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

intransitive et pronominale. Dans LVF, les constructions syntaxiques sont codifiées sous la forme alphanumérique. Les codes numériques symbolisent les traits sémantiques du sujet et des arguments : +Hum, +Conc, +Abst, etc.
Les lettres indiquent la construction syntaxique et la préposition (A : intransitif, N : transitif indirect, T : transitif direct et P : pronominal). (Dubois, Dubois-Charlier, 1997, P. 7)

3.2.1 Verbes intransitifs : A10, A17, A1q, etc.
Exemple 9 : voter (01) : A16 : N0Hum+V+Cpt Modalité.
On vote ce dimanche.
3.2.2 Verbes transitifs indirects : N1a, N1k, N3j, etc.
Exemple 10 : repostuler : N1a : N0Hum+V+Prep : à+N1Objet.
On repostule à ce poste.
3.2.3 Verbes transitifs directs : T11b0, T1307, T1900, T3900,
Exemple 11 : valider : T1308 : N0Hum+V+N1Objet.
On valide un bulletin de vote.
3.2.4 Verbes pronominaux : P3000, P10a0, P10b0, etc.
Exemple 12 : saisir (s) 16 : P10b0 : N0Hum+V+Prep :de+N1Objet.
Le conseil se saisit de l'affaire.

3.3 Au niveau des Opérateurs Sémantico- Syntaxiques

Des mots (latins, français ou anglais) abrégés symbolisent les propriétés sémantiques et syntaxiques du verbe et analysent sémantiquement les schémas syntaxiques.
Exemple 13 : dic ordre à qn (dire ordre à qn), dic mots juridiques (dire mots juridiques), fac crime (faire crime), etc.
Agréer : dat val à qc : donner valeur à qc.
Exemple 14: Entrée verbale: approuver 02



Figure 2. L'entrée verbale "approuver 02".

Le verbe approuver (02), par exemple, appartient à la sous-classe sémantique D3c "Donner validité à qc ou valeur à qn" dont la construction syntaxique est [T1300] qui signifie que le verbe est transitif direct avec un sujet [+Hum] et que le complément direct est [N1Conc].

L'opérateur sémantico-syntaxique est "dat.val à loi" (donner valeur à la loi). Ce verbe se dérive en "approbation" selon l'indication "2D" et se conjugue comme le verbe "aimer" et avec l'auxiliaire "avoir" (1aZ). Exemple15 : Le président approuve cette loi.

Notre objectif, dans cette étude, est la création d'une application et plus précisemment la création d'un processus de traduction automatique des verbes du domaine

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

administratif et de droit. Après avoir réalisé la première étape théorique, qui consiste à analyser et interpréter les phénomènes linguistiques à traduire, nous passerons à l'étape applicationnelle où nous allons exploiter les résultats de la première phase, pour la réalisation d'un système de traduction automatique.

4. Le Processus de Traduction Automatique Français-Arabe

Notre programme de traduction automatique se fonde principalement sur :

- L'implémentation des verbes dans un dictionnaire bilingue, français-arabe[4].
- La création des grammaires formelles de désambiguïsation syntaxique (Silberztein, M 2015).
- La traduction automatique de ces prédicats verbaux en arabe, en utilisant le logiciel NooJ.

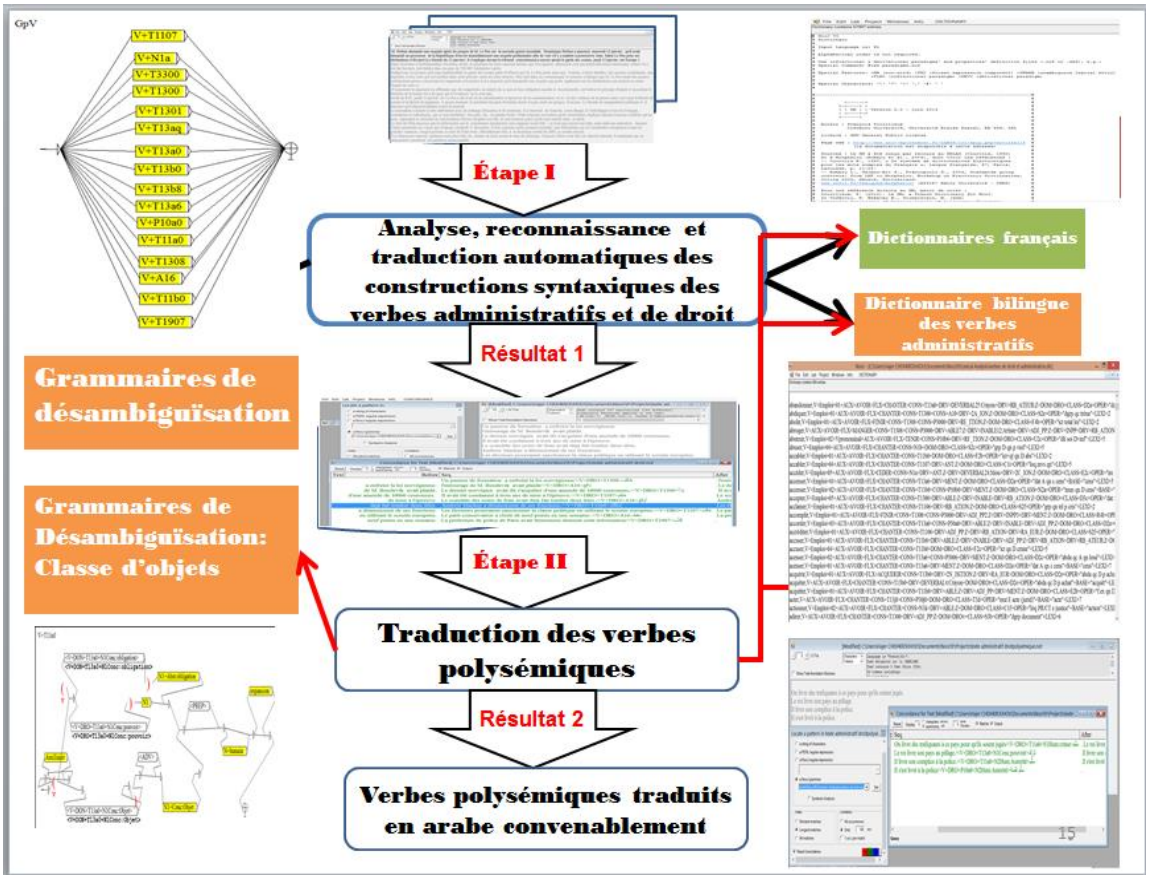


Figure 3. *La démarche adoptée pour la TA des verbes français en arabe.*

[4] NooJ est une plate-forme à laquelle contribuent des centaines de chercheurs de langues différentes. Ce sont ces chercheurs qui mettent en exergue les équivanlences lexicales et sémantiques entre les différentes langues en vue de créer des dictionnaires bilingues.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

4.1 La Création d'un Dictionnaire Electronique Bilingue : « verbes de droit et administration.dic »

Pour l'intégration des données théoriques des verbes du domaine administratif et de droit, relevés dans LVF, nous devons les reformuler par les opérateurs et le langage de NooJ. Cette première ressource linguistique est indispensable à la phase de reconnaissance et d'analyse linguistiques mais aussi à la phase de TA.

Dans cette étape, nous avons créé :

- Des paradigmes dérivationnels « *paraderivational.nof* ».
- Des paradigmes de flexion « *paraflexion.nof* ».
- Des paradigmes de conjugaison *« paradigmeflexionnel.nof »*
- La formalisation syntaxique où nous avons réécrit les schèmes syntaxiques du LVF.
- La traduction de chaque entrée verbale en arabe[5].

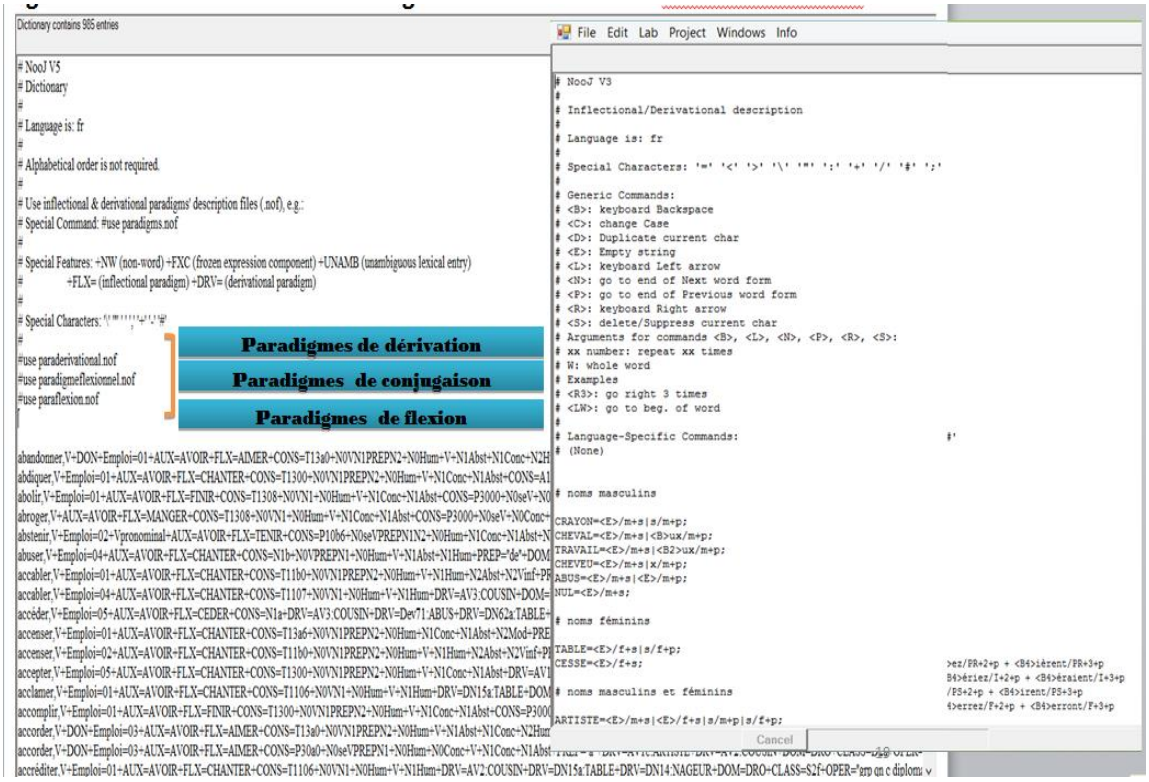


Figure 4. *Extrait du dictionnaire bilingue « verbes de droit et administration.dic ».*

[5] La traduction est le produit de la mise en œuvre des pradigmes dérivationnels, des pradigmes de flexion, des paradigmes de conjugaison et de la construction sémantico-syntaxique. Cette traduction n'est pas un choix mais c'est le produit d'un traitement automatique qui s'appuie sur les quatre étapes précédentes. L'évaluation et la pertinence de la traduction se fera par un expert.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Pour la reconnaissance automatique des différentes formes dérivées ou fléchies de chaque verbe, nous devons indiquer les « fichiers.nof » dans le dictionnaire bilingue, en vue de réaliser les opérations de dérivation verbale, de flexion nominale et adjectivale et de conjugaison[6].

Table 1. Exemples de paradigmes dérivationnels et de flexion.

Exemples de paradigmes dérivationnels	Exemples de paradigmes de flexion
1/Dérivation adjectivale # défendre=défendable=indéfendable AV1e=<B2>able/A|<B2>able<LW>in/A;	**1/ Noms masculins** CRAYON=<E>/m+s|s/m+p;
2 / Dérivation nominale # diviser=diviseur : DN14=<B2>eur/N ; # diviser=division : DN14a=<B2>ion/N;	**2/ Noms féminins** TABLE=<E>/f+s|s/f+p;
3 / Mot de base ou déverbal # fatiguer=fatigue :Dev1=<B>/N; # présignaliser=signal : Dev133=<B4><LW><R3><B3>;	**3/Noms masculins et féminins** DIRECTEUR = <E>/m+s | <B3>rice/f+s | s/m+p | <B3>rices/f+s;
	4/Noms épicènes ARTISTE=<E>/m+s|<E>/f+s|s/m+p|s/f+p ;

La production d'un dictionnaire bilingue nous montre qu'il y a une grande différence entre l'entrée verbale du dictionnaire LVF et celle du dictionnaire de NooJ.

Table 2. *L'entrée verbale "accuser 01".*

Verbe « accuser 01 » du domaine administratif										
Entrée verbale dans LVF										
accuser 01	DRO	F2c	ict qn D délit	inculper	Le juge a~P d'un délit.P est a~.est un a~.	1aZ	T11b0	21---RBRB --	-	1
Entrée verbale dans NooJ										
accuser,V+Emploi=01+AUX=AVOIR+FLX=CHANTER+CONS=T11b0+N0VN1PREPN2+N0Hum+V+N1Hum+N2Abst+N2Vinf+PREP="de"+DRV=AV1e:ARTISTE+DRV=AV2:COUSIN+DRV=DN15:DIRECTEUR+DRV=DN15a:TABLE+DOM=DRO+CLASS=F2c+OPER="ict qn D délit"+LEXI=1+AR="إتهم"										

[6] Les paradigmes de conjugaison nommés « paradigmeflexionnel.nof » sont créés par le professeur Max Silberztein.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Accuser,V+Emploi=01+AUX=AVOIR+FLX=CHANTER+CONS=T11b0+N0VN1PREPN2+N0Hum+V+N1Hum+N2Abst+N2Vinf+PREP="de".

Cette entrée verbale représente l'emploi 01 du verbe « **accuser** ». Le verbe se conjugue selon le paradigme *CHANTER* (FLX=CHANTER) avec l'auxiliaire avoir (AUX=AVOIR). Il suit la construction syntaxique (T : transitive directe).
+DRV=AV1eARTISTE+DRV=AV2 : COUSIN+DRV=DN15 : DIRECTEUR + DRV =DN15a : TABLE

Il admet la dérivation adjectivale « accusable » (AV1e) qui se fléchit selon le modèle « ARTISTE » et la dérivation « accusé » (AV2) dont le paradigme de flexion est « COUSIN ». Pour la dérivation nominale, nous trouvons la forme « accusateur » (DN15) qui se fléchit comme le mot « DIRECTEUR » et le nom « accusation » (DN15a) qui suit le paradigme « TABLE ».
+DOM=DRO+CLASS=F2c+OPER="ict qn D délit"+LEXI=1+AR="اِتَّهَمَ"

L'entrée verbale « accuser » appartient au domaine de « droit et administration (DRO) » qui relève de la classe sémantique « F2c », verbes de type « frapper ou toucher ». Le sens du verbe est « inculpé ». Ce verbe est extrait du « *dictionnaire fondamental de 1500 mots* » codé par « LEXI=1 ». Enfin, on ajoute la traduction arabe « +AR="اِتَّهَمَ"».

Pour vérifier le fonctionnement de notre dictionnaire, nous avons appliqué le dictionnaire « *verbes de droit et administration.dic* » au corpus administratif [7] *« texte administratif droit.not »*. Pour l'extraction des verbes du domaine administratif et de droit, nous avons utilisé la requête <V+DRO>.

[7] Le corpus utilisé rassemble des articles de journaux traitant des sujets administratifs et juridiques comme : La Nouvelle République du Centre Ouest, Agence France Presse, Le Monde.fr, L'Humanité, Le Télégramme, etc.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

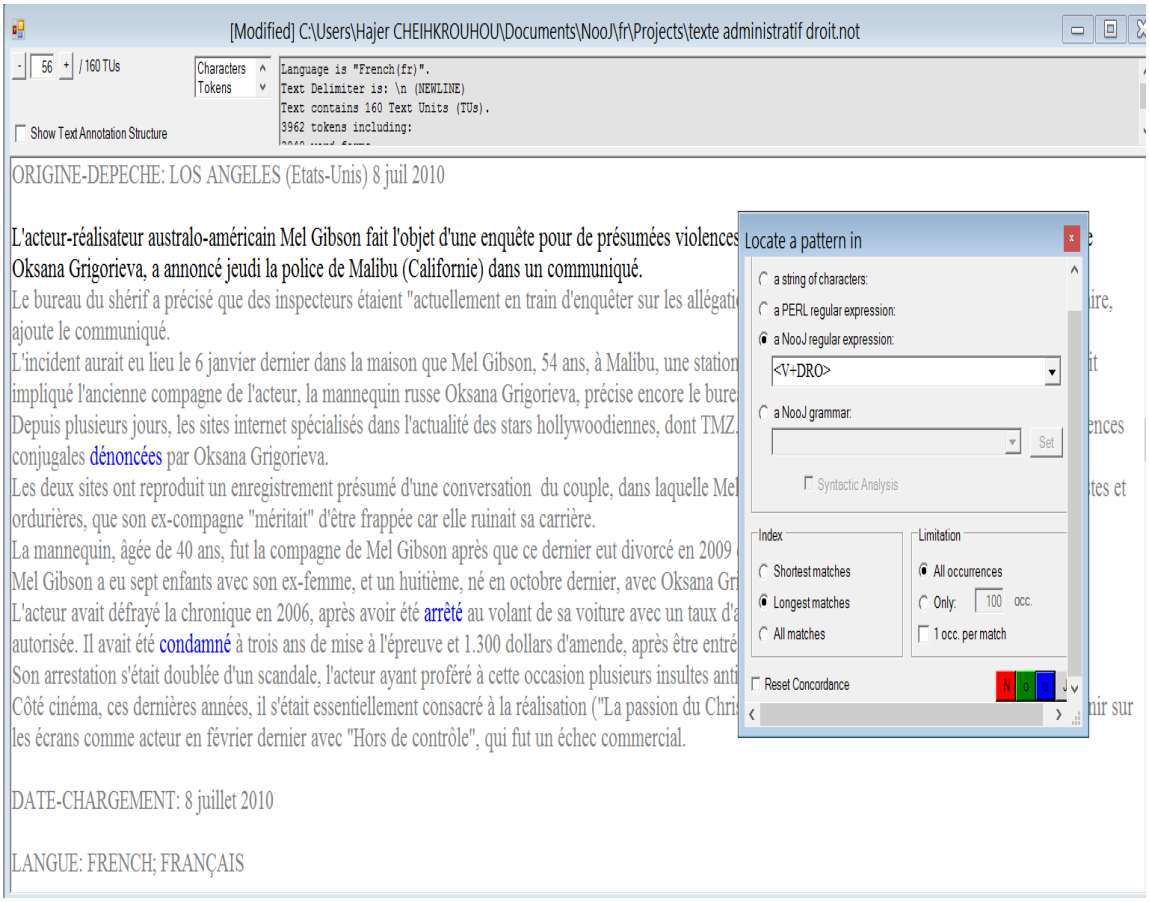


Figure 5. *La requête <V+DRO>.*

Cette opération a affiché les résultats suivants :

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

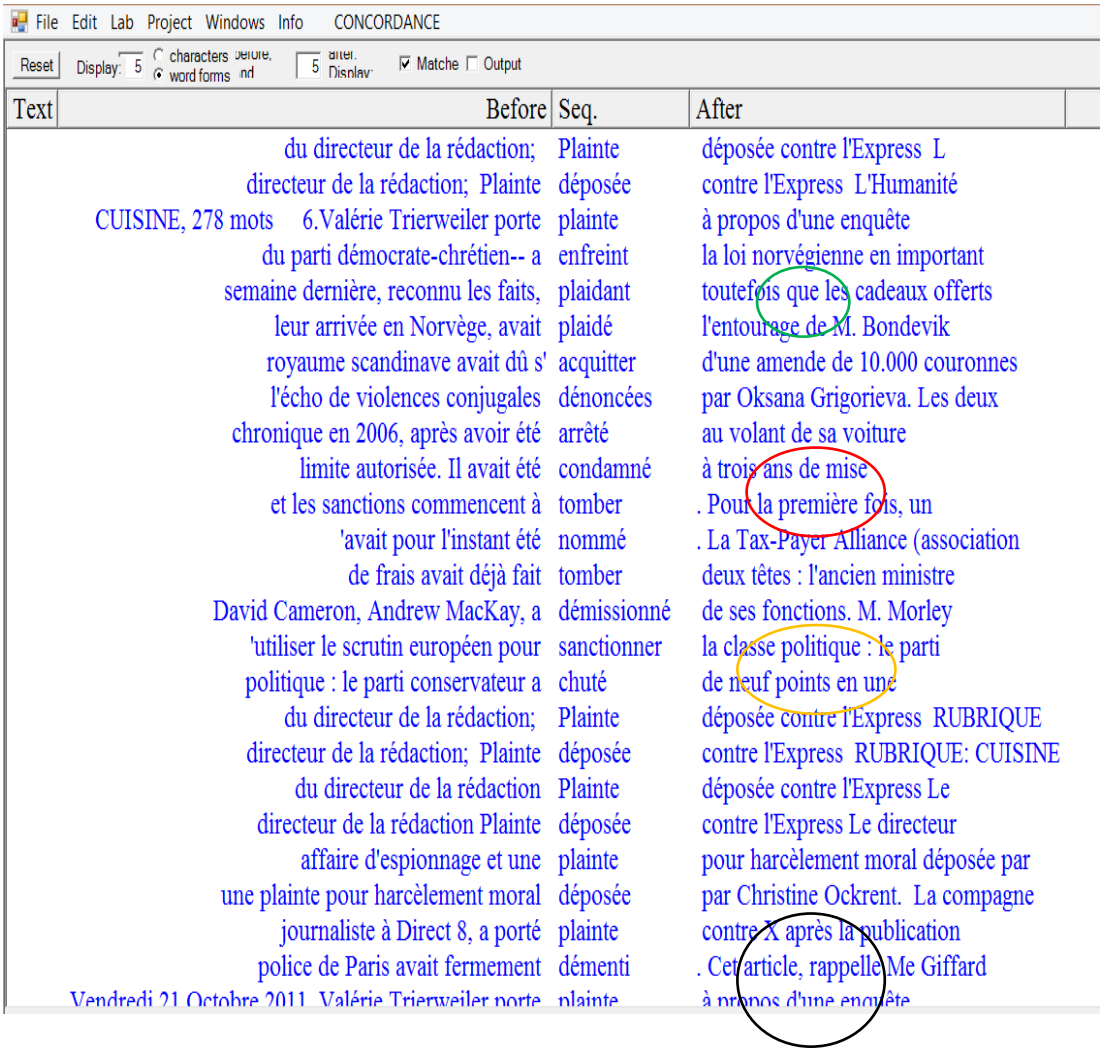


Figure 6. *L'extraction automatique des verbes.*

Nous constatons que le système informatique a pu analyser le corpus et relever les verbes appartenant au domaine administratif et de droit sous les différentes formes fléchies et dérivées, comme *« a enfreint, avoir été arrêté, avait été condamné, a démissionné, sanctionner, plainte, etc. »*.

Après la validation de la première étape, nous passerons à l'étape suivante qui consiste à créer des grammaires formelles pour la reconnaissance automatique des différents patrons sémantico-syntaxiques des verbes du domaine administratif et de droit.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

4.2 La Création des Grammaires d'Analyse et de Reconnaissance des Patrons Syntaxiques

Dans cette phase, nous allons essayer de créer des grammaires formelles qui doivent analyser les phrases pour en reconnaitre les différents schèmes argumentaux, trouvés dans un corpus ou un texte, et en ressortir les constructions syntaxiques convenables, sans aucune ambiguïté.

Les grammaires construites sont appelées **des grammaires de désambiguïsation syntaxique** (Silberztein, Max 2015). Dans cette phase, nous avons produit des grammaires de désambiguïsation qui nous permettent d'analyser seulement les constructions syntaxiques demandées.

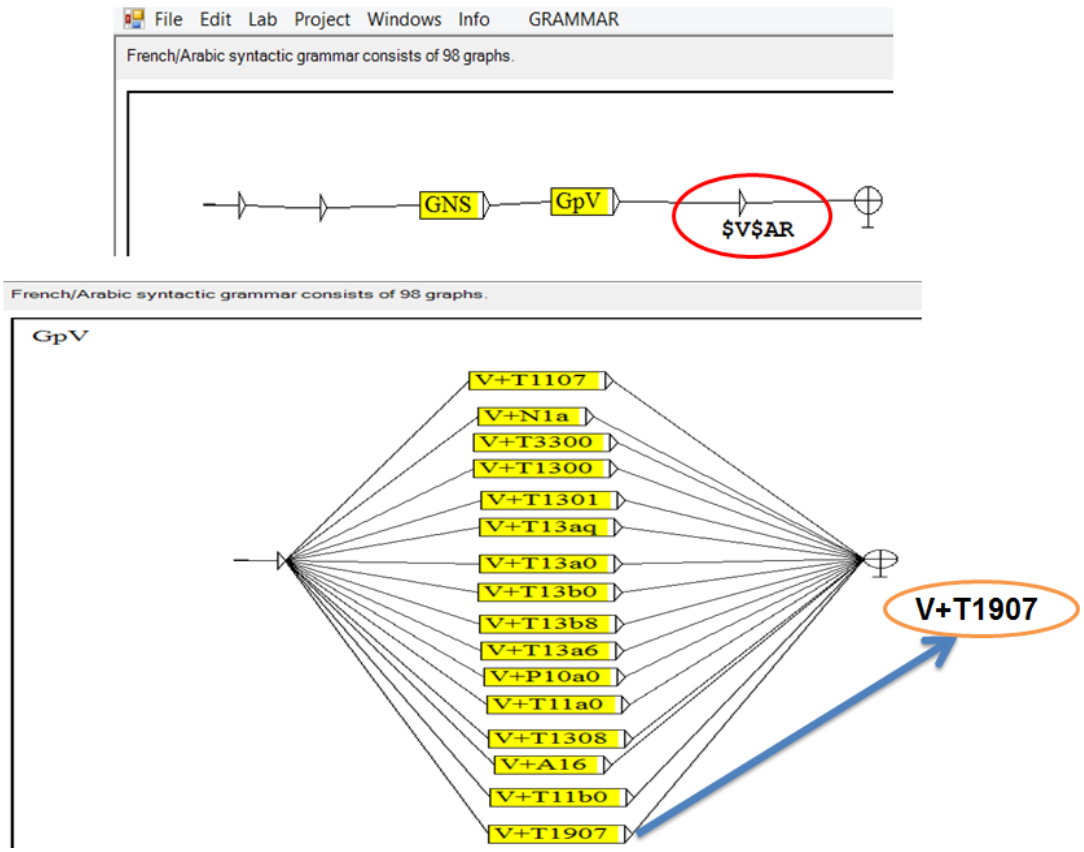


Figure 7. *Les grammaires de désambiguïsation.*

Par exemple, la construction syntaxique [V+T1907] nécessite un prédicat transitif direct dont le sujet est [+Hum] et le complément direct peut être [+Hum] ou [+Conc]. Nous trouvons cette construction syntaxique avec l'entrée verbale « démentir 01 ».

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

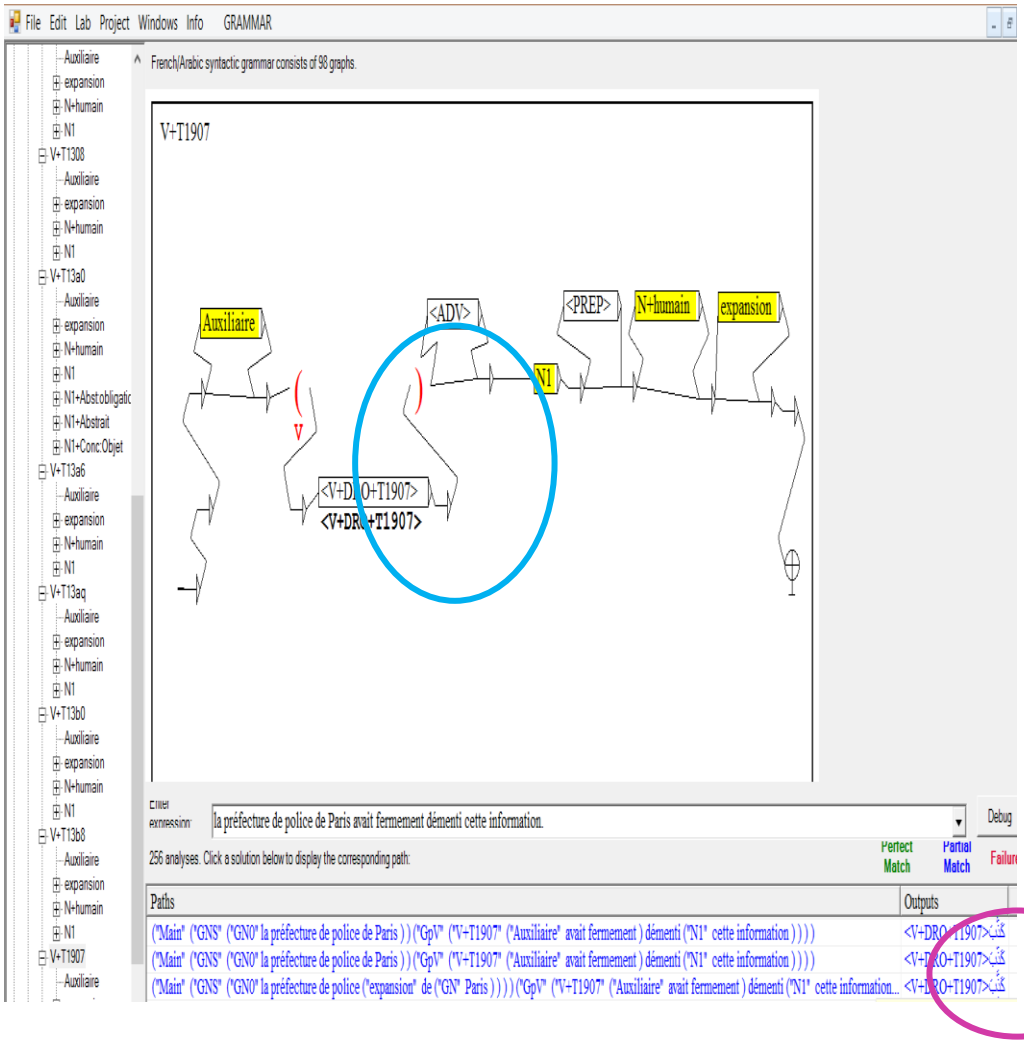


Figure 8. *La grammaire de la construction syntaxique [V+T1907].*

L'entrée verbale « démentir » :

démentir,
V+Emploi=01+AUX=AVOIR+FLX=DORMIR+CONS=T1907+N0VN1+N0Hum+V+N1Hum+N1Abst+N1Conc+DOM=DRO+CLASS=C1i+OPER="loq. mvsqn, qc"+LEXI=2+AR="كَذَّبَ"

Exemple 16 : La préfecture de police de Paris a démenti cette information.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Dans la création de notre grammaire pour cette construction syntaxique et même pour les autres constructions, nous avons précisé que le transducteur ne doit reconnaitre que le type syntaxique mis en filtre <V+DRO+T1907>.

La grammaire de désambiguïsation est nécessaire et importante, non seulement pour l'analyse et la reconnaissance automatiques des constructions syntaxiques adéquates, mais aussi pour produire une traduction correcte et fiable.

Nous notons que le système a pu analyser convenablement la phrase selon le filtre donné et il a aussi fourni la bonne traduction trouvée dans le dictionnaire bilingue, laquelle est « كَذَّبَ ».

Après avoir testé la validité de notre grammaire de désambiguïsation, nous allons appliquer cette dernière aux différentes phrases que nous avons trouvées dans le corpus.

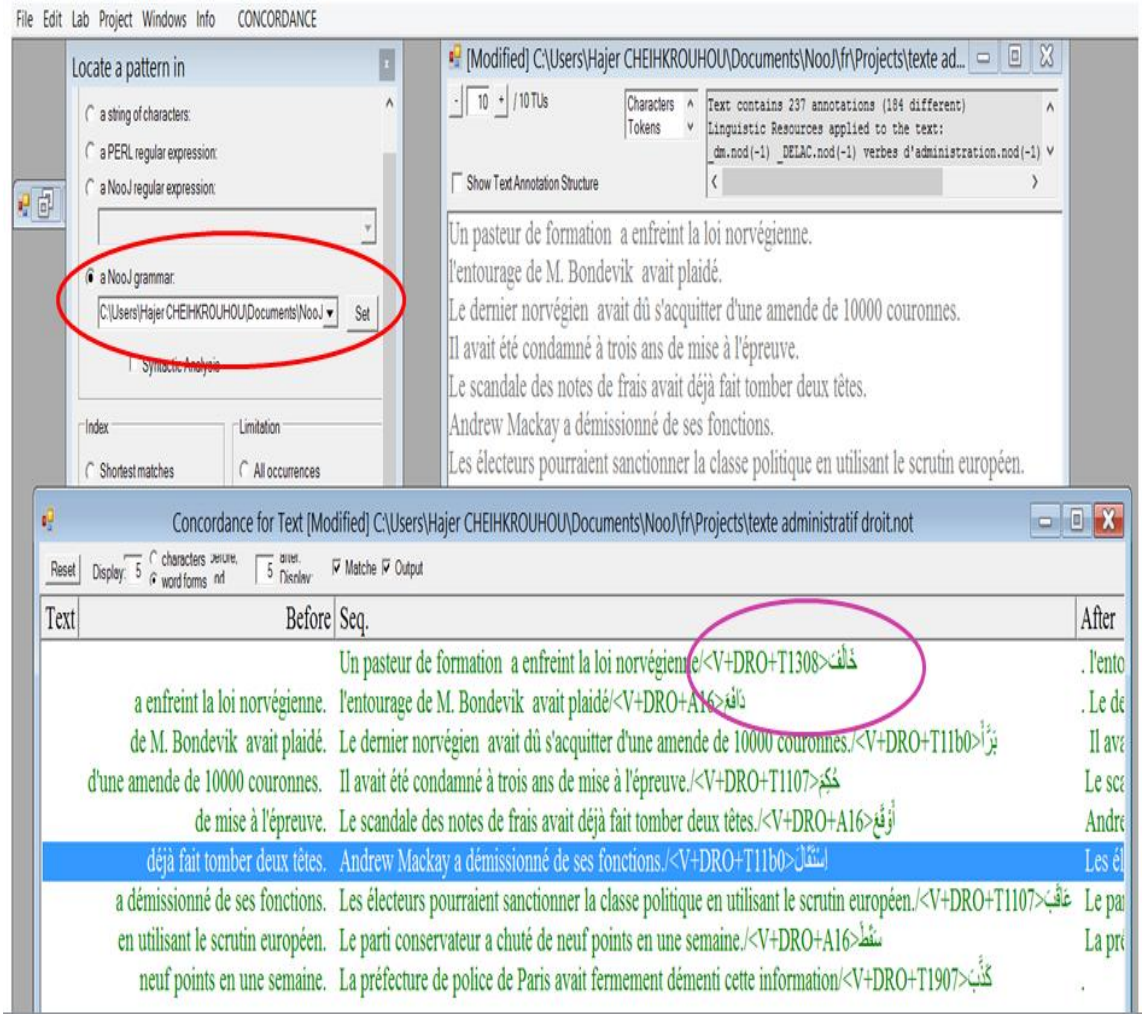


Figure 9. *L'analyse et la reconnaissance automatiques des différentes constructions sémantico-syntaxiques.*

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

Les résultats obtenus montrent bien l'efficacité de cette grammaire de désambiguïsation à l'étape de reconnaissance automatique et par la suite à l'étape de traduction automatique.
En fait, plus les règles de reconnaissance et d'analyse sont minutieuses et non-ambiguës, plus la traduction automatique sera fiable et adéquate.
Comme résultat, nous trouvons la phrase :
Andrew Mackay a démissionné de ses fonctions.
Cette phrase a été annotée convenablement par le type syntaxique <V+DRO+T11b0> et elle a été traduite par le verbe arabe « اِسْتَقَالَ ».

En créant le processus de TA, nous avons rencontré quelques ambiguïtés, surtout au niveau des emplois verbaux polysémiques. En effet, ce type d'ambiguïté nécessite un traitement spécifique et plus détaillé, surtout au niveau de la description des traits sémantiques des arguments de chaque emploi verbal.

5. La traduction des verbes polysémiques

La polysémie verbale pose un grand problème pour le TAL. Ce phénomène, omniprésent dans les langues naturelles, implique que le sens d'un verbe dépend du contexte dans lequel il est employé ; ce qui nous ramène à dire que la construction verbale est déterminante pour donner un sens au verbe ; ce qui justifie la remarque de Wittgenstein : « un mot n'a pas de sens en soi, il acquiert un sens seulement par son contexte ».

En étudiant les verbes du domaine administratif et de droit, nous avons trouvé plusieurs verbes polysémiques, qui ont nécessité d'autres stratégies d'analyse pour enlever toutes les ambigüités sémantico-syntaxiques et pour que le système informatique NooJ puisse traduire convenablement ces entrées verbales. Parmi ces verbes nous trouvons : attaquer (03, 10, 11), casser (05, 08, 11,13), commettre (01, 02), établir (03, 04, 07, 09), etc.
Exemple 17 : livrer possède 3 emplois polysémiques : livrer 01, 02, 03.

livrer 01	DRO	D1b	dat qn A étranger	extrader	On l'~des trafiquants à ce pays pour qu'ils soient jugés.	1aZ	T11a0	-1-
livrer 02	DRO	D2a	dat qc A qn ennemi	abandonner au pouvoir	On l'~son pays à l'étranger.	1aZ	T13a0	---
livrer 03	DRO	D1b	dat qn A autorité	dénoncer	On l'~son complice à la police.On se l'~à la police.	1aZ	T11a0 P10a0	---

Figure 10. *Les emplois polysémiques du verbe "livrer".*

- **Livrer 01 :** "dat qn A étranger". La construction synatxique est T11a0 dont le sujet et le complément d'objet sont [+ Hum] et +AR=سَلَّمَ

Exemple 18 : On livre des trafiquants à ce pays pour qu'ils soient jugés.
→ [T11a0] + AR = « سَلَّمَ»

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

- **Livrer 02 :** “dat qc A qn ennemi". La construction synatxique est T13a0 dont le sujet +Hum et le complément d’objet +Conc et +AR=«تَرَكَ»

Exemple 19 : Le roi livre son pays au pillage.
→ [T13a0] + AR = « تَرَكَ»

- **Livrer 03 :** “dat qn A Autorité". Deux constructions synatxiques :

Transitif direct T11a0 dont le sujet et le complément d’objet sont [+ Hum] et +AR=سَلَّمَ

Exemple 20 : Il livre son complice à la police.
→ [T11a0] + AR = « سَلَّمَ »

Pronominal P10a0 dont le sujet et le complément d’objet indirect [+ Hum] et +AR= سَلَّمَ نَفْسَهُ

Exemple 21 : Il s’est livré à la police. → [P10a0] + AR = « سَلَّمَ نَفْسَهُ»

➡ Nous notons que les constructions syntaxiques du verbe “**livrer**” sont très proches et leur distinction est très difficile, surtout dans le cas des constructions transitives directes.

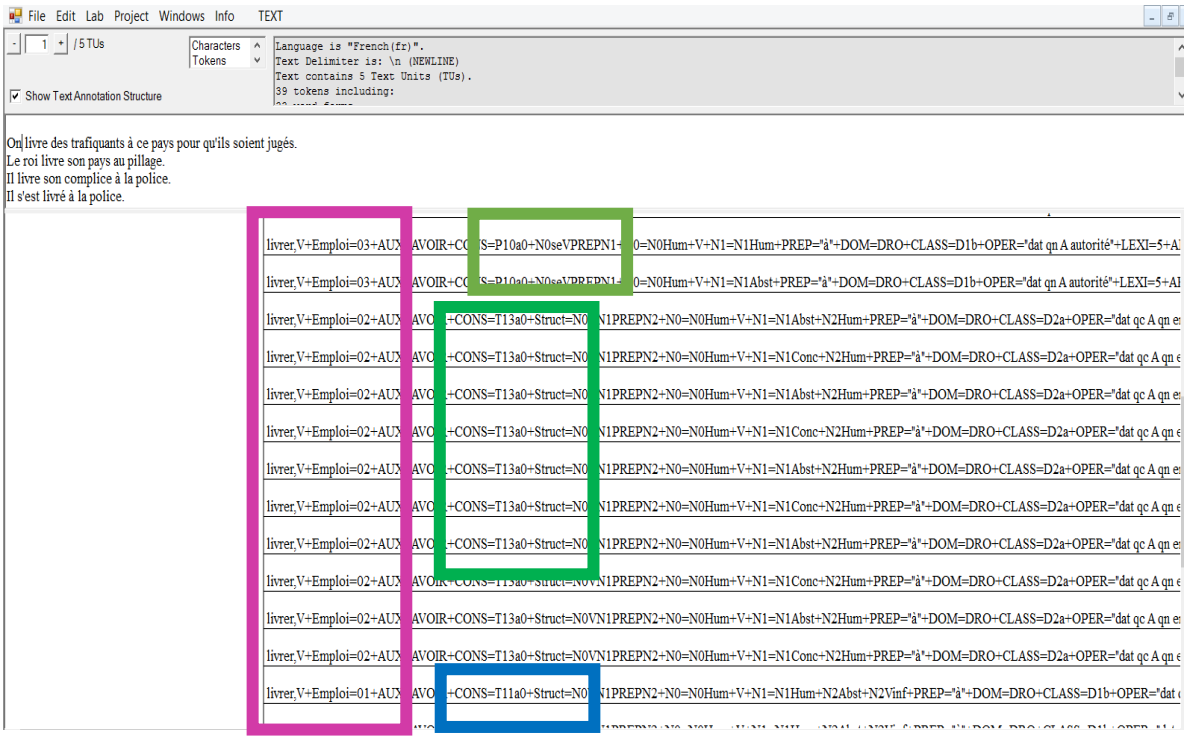


Figure 11. *Les premières analyses syntactico-sémantiques du verbe polysémique “livrer”.*

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

En suivant la même démarche d'analyse et de reconnaissance automatiques, nous avons trouvé que le système a pu analyser les différentes constructions sémantico-syntaxiques, mais qu'il a affiché des annotations aberrantes.

Exemple : On livre des trafiquants à ce pays pour qu'ils soient jugés.

Cette phrase a été interprétée par les différentes constructions et a aussi été annotée par tous les emplois trouvés dans le dictionnaire bilingue. Cette ambigüité au niveau de la première phase d'analyse et de reconnaissance va entrainer, en conséquence, une traduction fausse et aberrante.

Comme dans le cas de cet exemple, le verbe « livrer » est traduit par deux verbes arabes « سَلَّمَ» et « تَرَكَ », or la traduction correcte est « سَلَّمَ» propre à l'emploi 01.

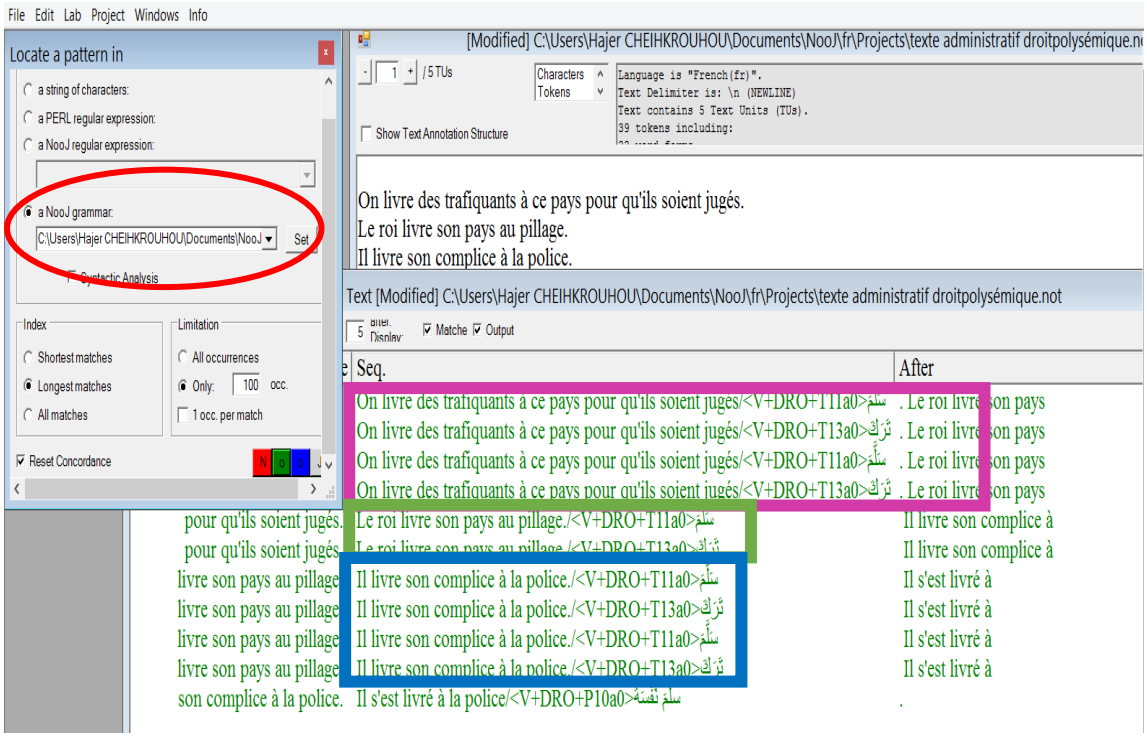


Figure 12. *Des annotations et des traductions aberrantes et erronées du verbe polysémique "livrer".*

Pour résoudre cette ambiguïté, Gaston Gross a trouvé la solution en ajoutant des informations supplémentaires concernant les arguments. Il sous-catégorise le sens selon des sous-classes sémantiques appelées **classes d'objets** (Gross. G :1994).

Le but ultime de cette théorie est de discriminer le sens du prédicat avec la précision exigée pour le traitement automatique. Une bonne analyse du verbe ne doit pas se limiter

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

à la détection des arguments, elle doit aussi s'intéresser à la clarification de leur nature et de leur classe sémantique, pour produire une traduction correcte et adéquate.

➡ Pour obtenir des résultats compatibles avec le sens du verbe, qui est conditionné par la sous-classe sémantique de son argument, nous devons réécrire les schémas argumentatifs de ces verbes dans les dictionnaires et dans les grammaires, tout en précisant la nature de la sous-classe sémantique des arguments, pour supprimer toutes les annotations non- appropriées.

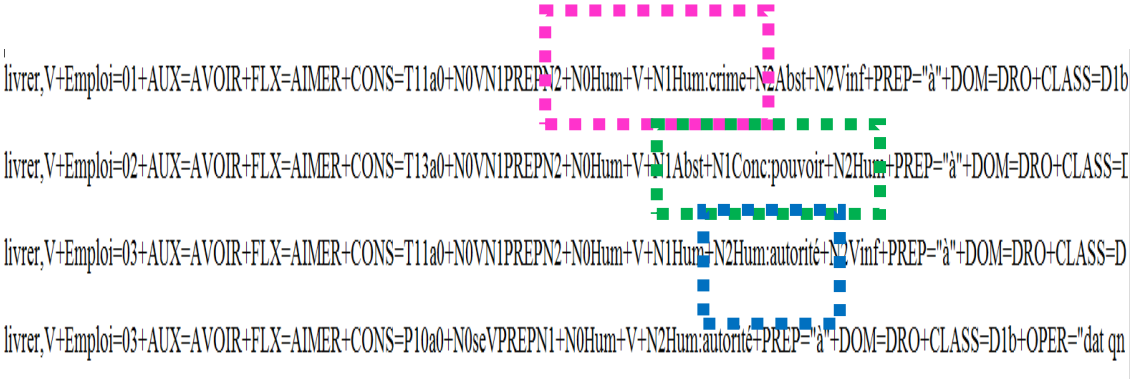
Livrer 01+T11a0+N1Hum : crime + AR = سَلَّمَ

Livrer 02+T13a0+N1Conc : pouvoir + AR = تَرَكَ

Livrer 03+T11a0+N2Hum : autorité + AR = سَلَّمَ

Livrer 03+P10a0+N2Hum : autorité + AR = سَلَّمَ نَفْسَهُ

Ces informations sont ajoutées à la description des verbes dans les dictionnaires et dans les grammaires de désambiguïsation.



Cette étape est conditionnée par une autre, celle de définir les noms, enregistrés dans le dictionnaire "dm.dic", par les caractéristiques sémantiques appropriées.

Exemple 22 :

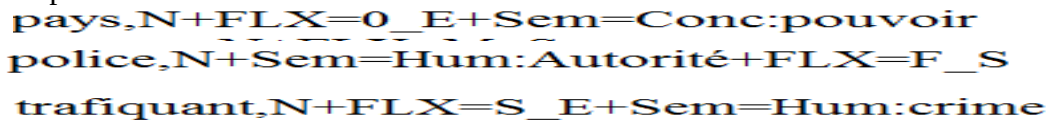
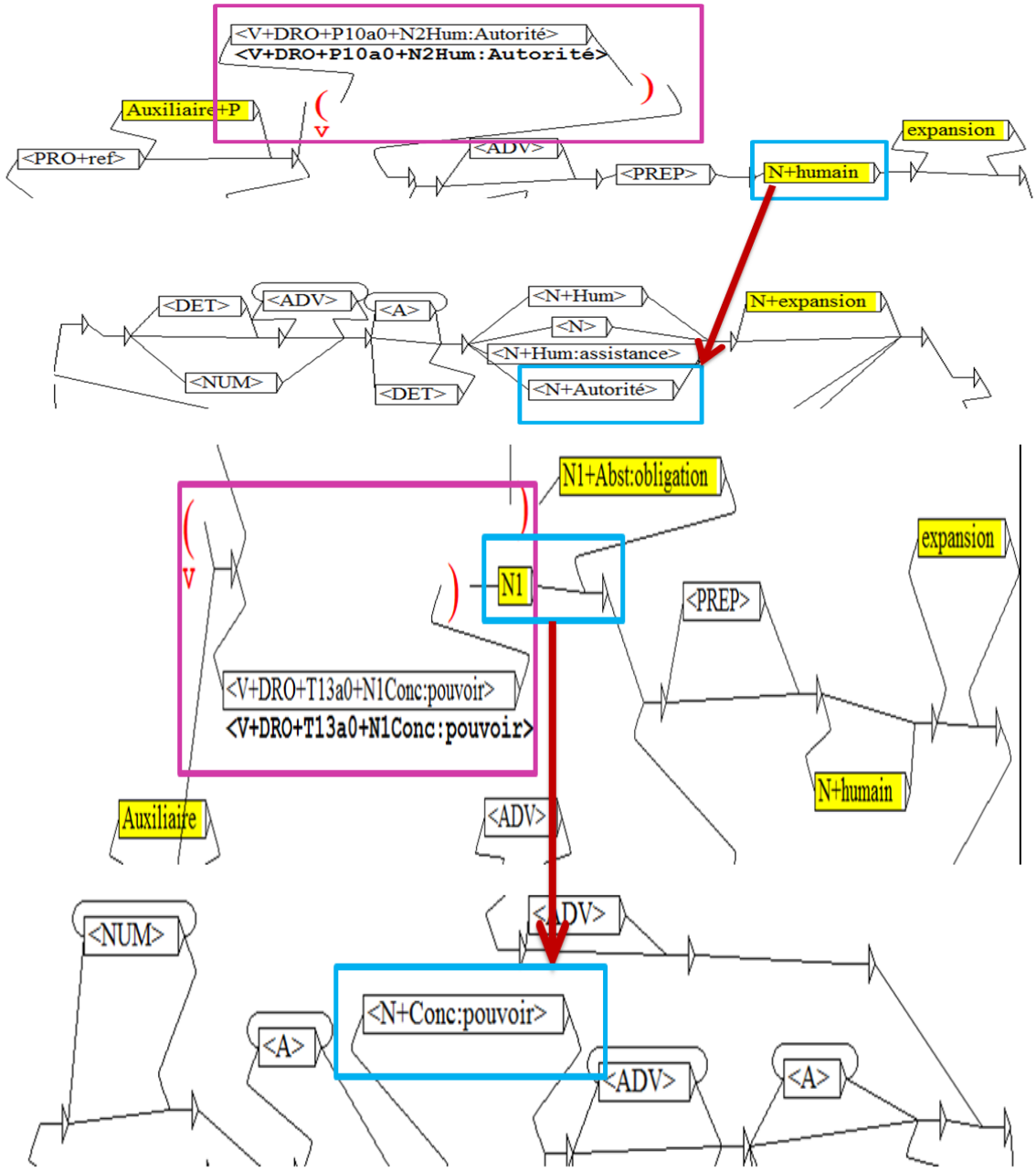


Figure 14. *Les sous-classes sémantiques des noms.*

Nous devons modifier les grammaires de désambiguïsation, en spécifiant les filtres par ces nouvelles sous-classes sémantiques.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International



Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

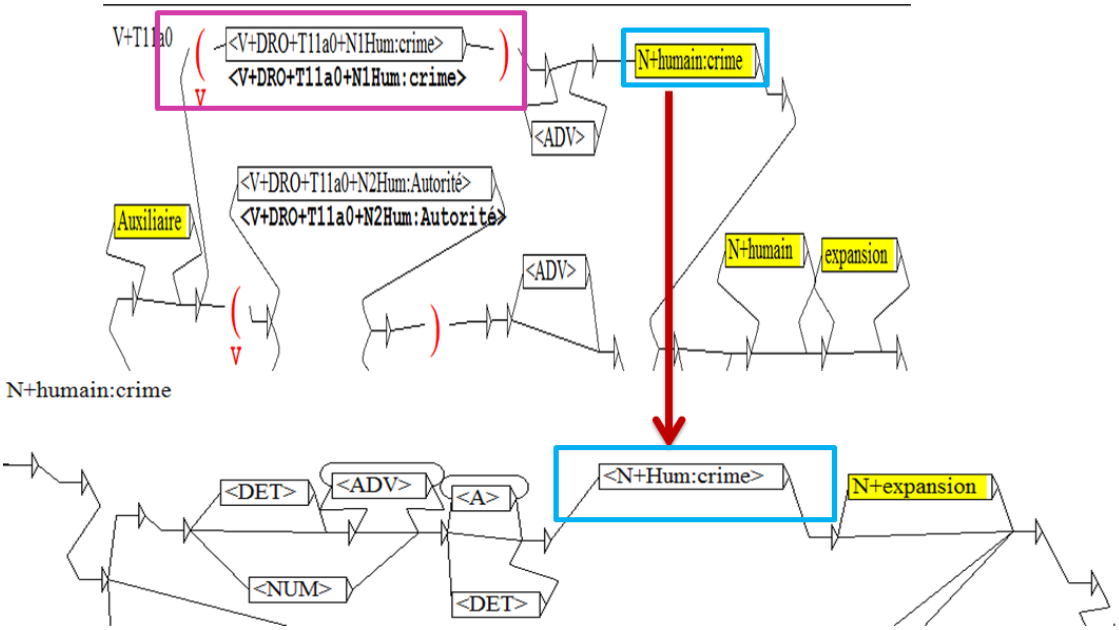


Figure 15. *La modification des filtres en ajoutant les sous-classes sémantiques convenables pour chaque grammaire de désambiguïsation <V+P10a0>, <V+T13a0> et <V+T11a0>.*

Après avoir inséré les modifications nécessaires aux entrées verbales dans les dictionnaires et les transducteurs syntaxiques de notre grammaire, nous avons réappliqué les grammaires à notre texte et nous avons trouvé les résultats suivants :

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

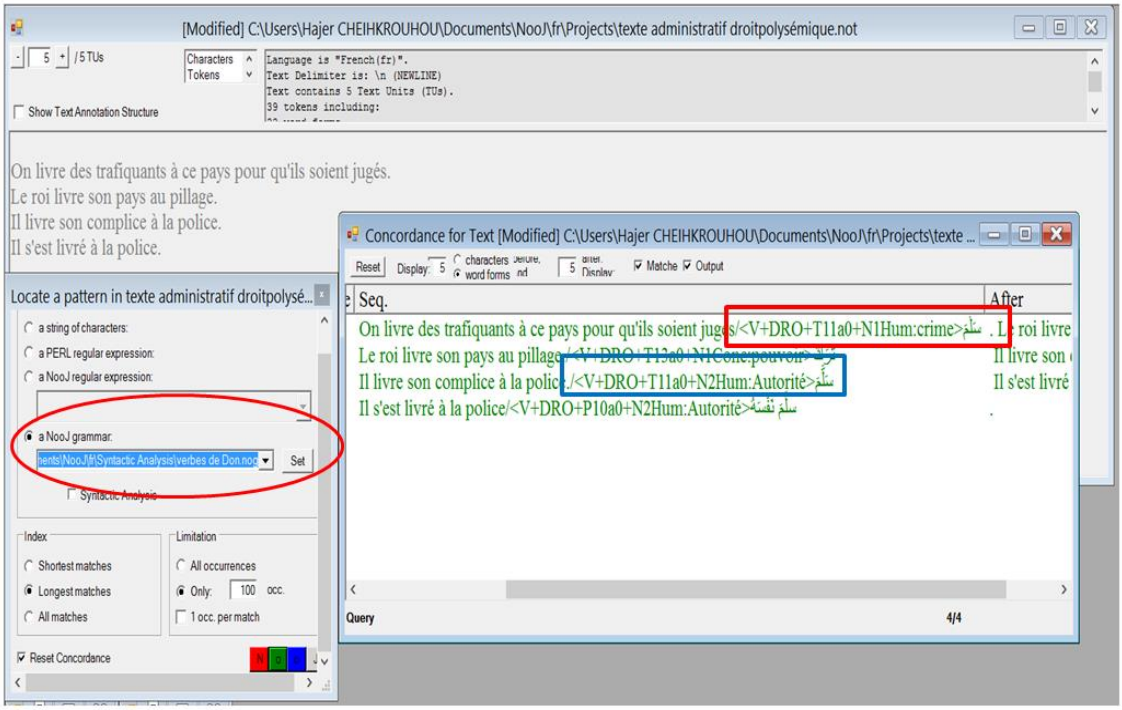


Figure 16. *Les résultats obtenus après l'application de la nouvelle grammaire.*

Nous remarquons que les phrases sont bien analysées et qu'elles sont marquées seulement par les annotations correctes et adéquates, de sorte que les prédicats sont traduits convenablement.

Par exemple 23 :

Il livre son complice à la police.

→ Cette phrase est annotée par l'annotation adéquate <V+DRO+T11a0+N2Hum : Autorité> et par la traduction juste سَلَّمَ. Cette annotation est spécifique à l'emploi : livrer 03. Pour supprimer les annotations fausses et inappropriées, nous utilisons la commande de NooJ <CONCORDANCE → Annotate Text (add/remove annotations)> qui nous permet de ne conserver que les annotations convenables. Cette commande a supprimé 70 annotations aberrantes.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

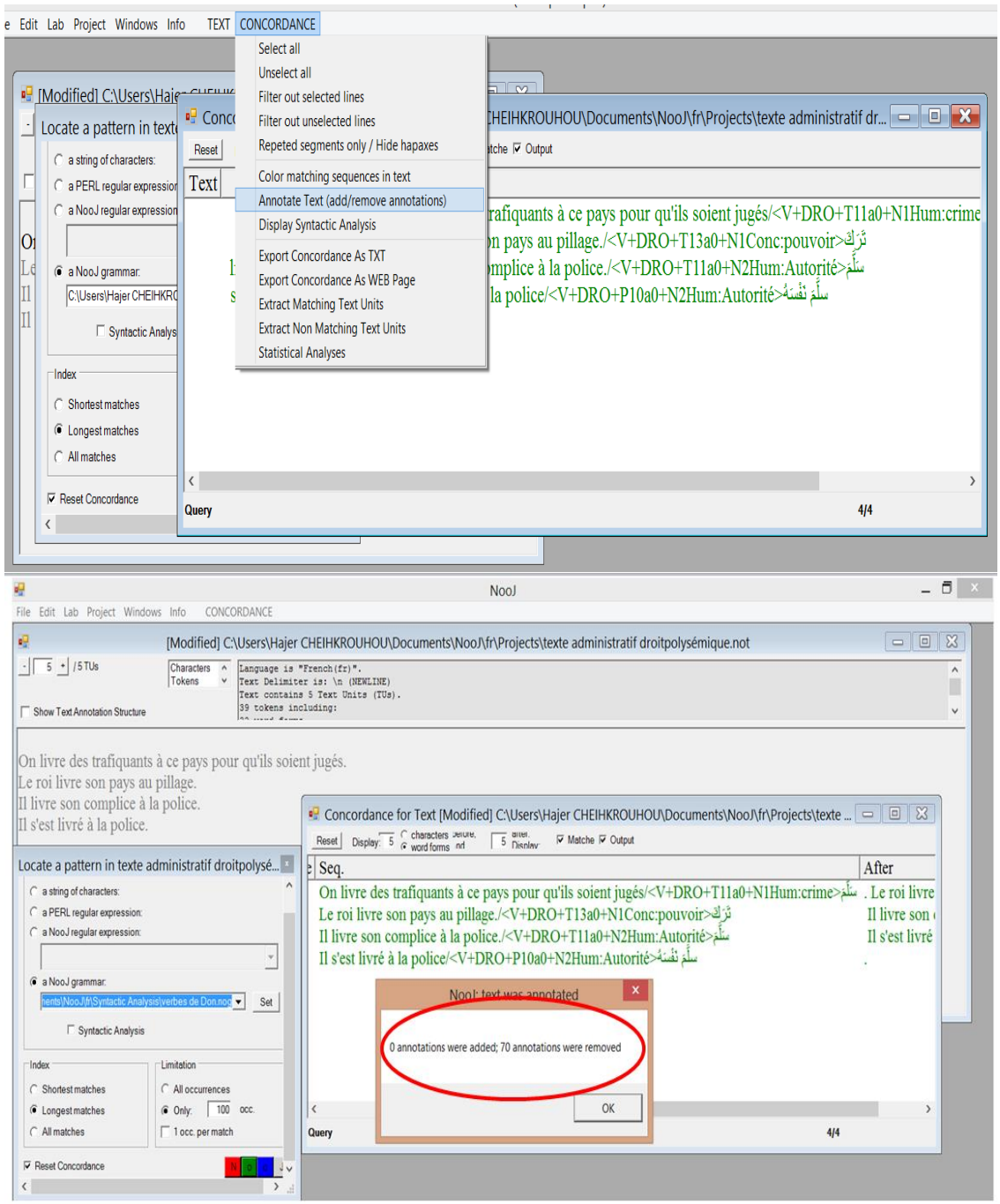


Figure 17. *La commande <CONCORDANCE → Annotate Text (add/remove annotations)>.*

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Comme résultat final, nous trouvons que la phrase <Il s'est livré à la police.> est analysée seulement par l'entrée verbale « livrer 03 » qui suit la construction syntaxique pronominale P10a0, dont le complément indirect est N2Hum : Autorité.

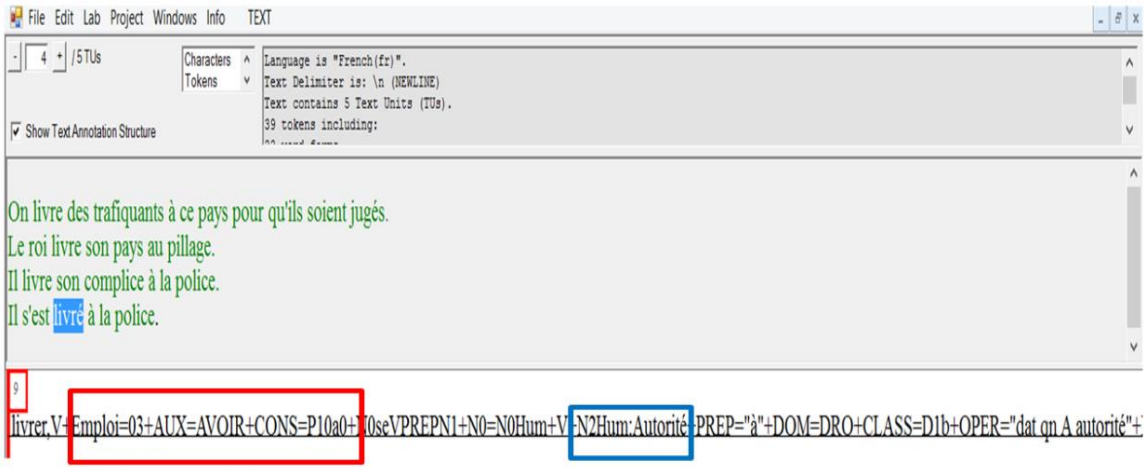


Figure 18. *La suppression des annotations inappropriées.*

Nous déduisons que les détails sémantiques que nous avons ajoutés ont réussi à supprimer toutes les ambiguïtés et les annotations incompatibles avec le filtre ; ce qui a conduit à la résolution de toutes les ambiguïtés aux niveaux de la reconnaissance syntaxique et de la traduction automatique.
En fait, la théorie des classes d'objets garantit la désambiguïsation des emplois polysémiques et assure une traduction adéquate et fiable.

6. Conclusion et Perspectives

Dans ce travail, nous avons essayé de créer un système de traduction automatique des verbes relevant du domaine administratif et de droit. Ce système s'est réalisé sur deux phases:

- La phase de reconnaissance et de désambigüisation des différentes constructions syntaxiques.
- La phase de traduction automatique.

L'application de la théorie des classes d'objets de Gaston Gross nous a permis d'affiner les grammaires de reconnaissance et d'obtenir par la suite des résultats encourageants, où les verbes polysémiques sont convenablement traduits en arabe.

Il nous reste évidemment beaucoup à faire. Dans le prolongement de ce travail, nous envisageons d'étendre cette étude à tous les verbes de droit et d'administration en créant un dictionnaire bilingue comparable à ceux que nous avons réalisés pour les verbes de communication (2013) et de mouvement (2014).

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Références

[1] Cheikhrouhou, H. (2013). Recognition of Communiction Verbs with NooJ Platform. *Formalising Natural Languages with NooJ 2013*, Cambridge Scholars Publishing, British (2014), pp. 155–169.

[2] Cheikhrouhou, H. (2015). The Formalisation of Movement Verbs for Automatic Translation using NooJ Platform. *Formalising Natural Languages NooJ 2014*, Cambridge Scholars Publishing, British, pp. 14–21.
https://www.cambridgescholars.com/resources/pdfs/978-1-4438-7558-5-sample.pdf

[3] Cheikhrouhou, H. (2019): Automatic recognition and translation of polysemous verbs using the platform NooJ. *Formalising Languages with NooJ 2019 and its Natural Processing Applications 13th International Conference*, Springer, Heidelberg (2019), pp. 39–51.
https://doi.org/10.1007/978-3-030-38833-1_4

[4] Cheikhrouhou H. (2021). « The Recognition and the Automatic Translation of Dative Verbs ». *Formalizing Natural Languages: Applications to Natural Language Processing and Digital Humanities. NooJ 2021*, Communications in Computer and Information Science, Vol 1520, Springer, Cham, pp. 174- 186.
https://doi.org/10.1007/978-3-030-92861-2_15

[5] Dubois, J., Dubois-Charlier, F. (1997). *Les Verbes Français*. Larousse, Paris.

[6] Gross, G. (2008). « Les classes d'objets ». *Lalies*, Presses de l'ENS, Editions rue d'Ulm, pp.111-165.
https://halshs.archives-ouvertes.fr/halshs-00410784

[7] Le Pesant, D., J, François. et D, Leeman. (2007). Présentation de la classification des Verbes Français de Jean Dubois et Françoise Dubois- Charlier. *Langue Française*, 153, Larousse, Armand Colin.
https://www.cairn.info/revue-langue-francaise-2007-1-page-3.htm

[8] Silberztein, M. (2015) : *La Formalisation des Langues l'Approche de NooJ*. Collection Science Cognitive et Management des Connaissances. ISTE Editions.

Remerciements

Nous tenons à remercier tous ceux qui ont contribué à la conception et à l'organisation de ce colloque ! Nous exprimons d'abord notre fierté et notre honneur au réseau LTT, dont les membres ont eu la lourde tâche d'évaluer les actes du colloque !
Les sponsors du colloque LTT2021 n'ont pas lésiné sur les moyens, en prodiguant tout ce qui est nécessaire à l'organisation en présentiel et à distance d'une manifestation dont les participants ont certainement profité aussi bien scientifiquement que sur le plan relationnel.
Nos remerciements vont aussi bien à tous les memebre de l'équipe du CERIST (Centre de Recherche sur l'Information Scientifique et Technique), hébergeur de la revue TRANSLANG, pour tous les efforts consentis. Qu'ils trouvent tous ici l'expression de

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

notre estime et surtout de notre reconnaissance de nous avoir donné cette chance de participer et de publier nos travaux dans les actes de ce grand colloque !

Notice biographique des auteurs

Dr. Hajer Cheikhrouhou, Docteure en langue française spécialité Linguistique et Sciences du langage. Elle a soutenu sa thèse, intitulée
« Le verbe et la problématique de la traduction-automatique français-arabe (utilisation de la plate-forme NooJ) », le 19 juin 2017 à l'INACO-Paris sous la direction des professeurs Mohamed Bouattour et Max Silberztein dans le cadre d'une cotutelle entre l'Université de Sfax-Tunisie et l'Université de Franche-Comté-France. Membre du Laboratoire de Langage et Traitement Automatique (Université de Sfax-Tunisie). Elle a participé à plusieurs conférences comme la conférence Internationale NooJ. La majorité de ses publications ont essentiellement traité la reconnaissance et la traduction automatiques des verbes de communication (2013), de mouvement (2014), les auxiliaires français (2016), les verbes polysémiques français (2019), les prédicats nominaux arabes (2022), etc.

Prof. Mohamed Bouattour, Docteur d'Etat en langue et littérature françaises, directeur du laboratoire Langage et Traitement Automatique de l'Université de Sfax.
Mohamed Bouattour a encadré plusieurs thèses portant sur les questions de polysémie et d'ambiguïté syntaxique et des difficultés qu'elles posent au niveau de la traduction, surtout automatique. En plus des articles et des ouvrages collectifs, Pr. Mohamed Bouattour a publié trois ouvrages, le premier, en 2007, s'intitule Interaction et argumentation dans les fables de La Fontaine, le second publié en 2021, s'intitule Les stratégies définitoires dans les Provinciales de Blaise Pascal, le dernier en date, publié en 2022, s'intitulé La stratégie entre construction géométrique et construction rhétorique dans les Provinciales de B. Pascal.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155