

A Rule-Based System for Disambiguating French Locative Verbs and Their Translation into Arabic

Safa Boudhina

University of Gabes
safaboudhina1@gmail.com

Héla Fehri

University of Sfax
hela.fehri@gmail.com

Abstract

This paper presents a rule-based system for disambiguating French locative verbs and their translation into Arabic. The disambiguation phase is based on the use of the French Verb dictionary of Dubois and Dubois Charlier (LVF) as a linguistic resource, from which a base of disambiguation rules is extracted. The extracted rules take the form of transducers which are subsequently applied to texts. The translation phase consists in translating the disambiguated locative verbs returned by the disambiguation phase. The translation takes into account the verb tense, as well as the inflected form of that verb. This phase is based on bilingual dictionaries that contain the different French locative verbs and their translation into Arabic. The experimentation and the evaluation are done using the linguistic platform NooJ, both a language resource development environment and a tool for automatic large corpora flow (Fehri, 2012).

1 Introduction

Lexical ambiguity represents an obstacle for the automatic processing of natural language. In fact, the difficulty of automatically dealing with this phenomenon was recognized as early as the first appearance of automatic translation systems (Apidaniaki, 2008). This ambiguity can occur at the level of different grammatical categories of words, including the verb. In its conjugated form, the verb is distinguished by tense, mood, person, number and the syntactic constructions in which this verb appears.

However, the task of solving the problem of ambiguity, known as Word Sense Disambiguation (WSD), is not an easy one. In fact, it's essential to specify the elements that contribute to the selection of the appropriate meaning of the ambiguous word. Gross et al. (1997) state that "any syntactic difference corresponds to an essential semantic difference." This statement emphasizes the importance of syntactic constructions in the choice of the meaning of the verb. In other words, the meaning of the verb strongly depends on the syntactic construction in which it appears. Furthermore, to properly handle this condition, the entire sentence containing the ambiguous verb should be analyzed, and this requires a certain linguistic knowledge.

In this perspective, the dictionary of French Verbs of Jean Dubois and Françoise Dubois-Charlier (LVF) constitutes a relevant and useful resource for the WSD of verbs. Indeed, it proposes a verbal classification that relates the syntactic and semantic characteristics of the verb.

WSD may be seen as an intermediate task for some applications such as machine translation. However, to provide the suitable translation of an ambiguous word, WSD is indispensable in identifying the most appropriate meaning of the word in question. In addition, the quality of the translation does not depend solely on the preservation of the meaning of the translated word in the target language. The consideration of the form of the verb and the tense in which that verb is conjugated is also very useful and influences the quality of the translated word.

The main objective of the present work is to develop a rule-based system for disambiguating French verbs, in particular locative verbs. This system also makes it possible to translate the disambiguated verbs while taking into account the tense of the verb and its inflected form. Our system is based on the LVF dictionary using the NooJ platform.

2 Related work

According to the resources used, WSD approaches can be roughly classified into two main categories. The first one is the corpora-based approach, in which systems are trained to perform the task of word sense disambiguation (Dixit et al., 2015). This type includes supervised and unsupervised methods. The supervised methods are based on a learning corpus grouping examples of disambiguated instances of words, while unsupervised methods exploit the results of automatic meaning-acquisition methods. The second main category is the knowledge-based approach, which exploits knowledge resources to infer the senses of words in context (Dixit et al., 2015).

Supervised WSD methods require annotated corpora to be trained (Nessereddine et al., 2015). However, such corpora are only available for a few languages, which complicates supervised lexical disambiguation. In this perspective, Nassreddine et al. (2015) recently proposed a method for creating a system of lexical disambiguation for the French language. Since this language suffers from a scarcity of annotated corpora, their method consists of constructing an annotated corpus by the transfer of annotations. These annotations are obtained from the automatic translation of an English annotated corpus into French. The Bayesian classifier is then used to build the disambiguation system. The precision obtained is of the order of 52%. The limit of this method is that it strongly depends on the annotations obtained by the translation phase. In addition, the quality of the disambiguation system is restricted by the examples used during the learning phase.

In general, the problem with supervised, corpora-based disambiguation approaches is that getting large amounts of annotated text in one's way is very costly in terms of time and money, which results in a data acquisition bottleneck (Tchechmedjiev, 2012). Moreover, the quality of the disambiguation of these approaches is restricted by the examples used for training.

For unsupervised methods, the notion of meaning is directly induced by the corpus (Audibert, 2003). The method proposed by Schütze (1998), which relies on the use of the vector model can be taken as an example. In this method, a vector is associated with each word m of the corpus. This vector is derived from words that co-occur in the context of m . Subsequently, according to their degree of similarity, these vectors are grouped into clusters. Each cluster refers to a possible meaning of the word m .

The problem with this method, as well as unsupervised approaches in general, is that the senses do not correspond to any well-defined set of meanings. The distinctions of meaning can sometimes be confusing and are, moreover, often difficult to be used by other applications (Audibert, 2003).

In knowledge-based approaches, the information needed for lexical disambiguation is derived from external resources of the studied corpus. These resources are of different natures. Indeed, some studies have tested the adequacy of the definitions provided by the Collins English Dictionary (CED), Merriam-Webster New Pocket Dictionary, and Contemporary English Dictionary (LDOCE) for the automatic processing of disambiguation. Other works have exploited the information provided by Roget's International Thesaurus or by WordNet semantic lexicons explaining the meanings and relations between senses or shades of meaning (Vasilescu, 2003).

The disambiguation system established by Lesk is one of the first systems that relies on a dictionary. The principle of this method is first of all to extract from the dictionary all the possible meanings for each occurrence of the ambiguous word. Each sense corresponds to a particular definition in the dictionary. Subsequently, a score is assigned for each of these senses. This score is equal to the number of words in common between the definition of the word to be disambiguated and the definitions of the words co-occurring in its context (Vasilescu, 2003). The sense selected as being the most appropriate sense is supposed to be the one that maximizes the score. This method correctly disambiguates between 50% and 70% of the cases (Boubekeur-Amirouche, 2008).

The method proposed by Brun et al. (2001) is based on an electronic dictionary. The dictionary used in this method gives an example for each meaning. Each example is analyzed through a specific analyzer, from which rules of disambiguation are elaborated. These rules are then generalized using a semantic network that assigns a semantic class to each existing unit in the rule of origin. Subsequently, to disambiguate a word in any sentence, it is syntactically analyzed and the syntactic relationships are derived from it (Brun et al., 2001). These relationships are then compared with the disambiguation rules. The disambiguation rule whose arguments are the same is then selected and its meaning number is associated with the word in question (Brun et al., 2001).

This method was applied to 850 French sentences for the set of nouns, adjectives and verbs. The precision rate obtained for verbs is of the order of 58%. The problem with this method is that the elaborated rules of disambiguation depend on the words that exist in the examples found in the dictionary. This link has also generated a problem of coherence between the dictionary and the chosen semantic network, due to the semantic classification proposed by the thesaurus. In addition, some lexical entries haven't been disambiguated because of the lack of examples for certain meanings from which the disambiguation rules are extracted.

This method has the disadvantage of being very sensitive to the words that exist in each definition. Indeed, the choice of meanings based on a limited number of common words can be the source of errors.

3 Description of the French Verb Dictionary (LVF)

The LVF is a lexical database of French language, which contains 25,609 entries for 12,310 different verbs. The verbs are classified into 14 classes, which is based on the hypothesis of a correspondence between syntax and meaning (Sénéchal et al., 2007). The class denoted *L* presents the class of locative verbs, the topic of this work. It contains 1524 locative verbs (e.g., *baigner* “to bathe”, *apparaître* “to appear”). We note that our choice of the type of verbs is arbitrarily done.

4 The creation of electronic locative verbs dictionary with NooJ

The developed disambiguation system focuses on the use of the LVF dictionary. This dictionary has been built to be used by linguists, and is not directly exploitable by automatic text analysis programs (Silberztein, 2010). For this reason, we opt for the formalization of the LVF dictionary. More precisely, we are interested in the formalization of the locative verbs described in the LVF. This formalization consists in reformulating the information contained in LVF so that it can be exploited by automatic analysis programs, notably NooJ (Silberztein, 2010). The electronic dictionary created in NooJ, called *LocativeVerbs.dic* contains 1,524 verbal entries. Figure 1 presents an extract from this dictionary.

```

apparaître, V+VA+FLX=CONNAITRE+CONST=A31+sens=paraître+AR=ظَهَرَ
apparaître, V+VA+FLX=CONNAITRE+CONST=A11+sens=se présenter+AR=حَضَرَ
accueillir, V+VT+FLX=CUEILLIR+CONST=T11j0+sens=recevoir+AR=اِسْتَقْبَلَ
accueillir, V+VT+FLX=CUEILLIR+CONST=T3100+sens=admettre+AR=قَبِلَ
baigner, V+VP+FLX=AIMER+CONST=P10j0+sens=nager+AR=سَبَحَ
baigner, V+VN+FLX=AIMER+CONST=N1j+sens=immerger+AR=اِنْعَمَرَ
baigner, V+VT+FLX=AIMER+CONST=T13j0+sens=tremper+AR=غَمَسَ
baigner, V+VP+FLX=AIMER+CONST=P1001++sens=se tremper+AR=اِسْتَحَمَ

```

Figure 1. Extract of electronic dictionary *LocativeVerbs.dic*

As can be seen in Figure 1, each locative verbal entry is associated with a set of features: a lemma that corresponds to the verb in the infinitive form; a label indicating the grammatical category (*V*); the *FLX* property which gives the inflectional model related to the verb; the *CONST*, which shows the syntactic-semantic properties of the verb; the *sens*, which indicates the meaning of the verb; and finally the translation into the target language (*AR*). It's worth-noting that a verb can be associated with more than 10 senses. The verb *baigner* “to bathe” is used as an example. Since this verb has four different meanings described in the LVF dictionary (*baigner* 01=*nager*, *baigner* 02=*immerger*, *baigner* 03=*tremper*, *baigner* 04=*se tremper*), it appears four times in the *LocativeVerbs* dictionary, as shown in Figure 1.

The NooJ implementation of our system involves a two-phase process: (1) disambiguation of French locative verbs and (2) translation into Arabic. Each phase requires the construction of proper transducers.

5 Phase of disambiguation

Each verbal entry described in the LVF dictionary is specified by the different possible meanings that correspond to it. These meanings are illustrated by different uses, semantic indicators, syntactic characteristics, etc. The information related to the given meaning of the ambiguous verb, thus paves the way to extract the necessary disambiguation rules. The extracted rules then make it possible to identify the syntactic patterns that will be transformed into transducers to achieve the task of disambiguating verbs.

5.1 Identification of disambiguation rules

The French verbs described in the LVF are classified into semantic classes defined by syntactic arguments, which make it possible to assign the appropriate meaning to each occurrence of the verb. That is, each use of the verb is described by a particular syntactic construction, and then the meaning of the verb is determined by the construction in which the verb appears. In fact, the verb can have four syntactic constructions: intransitive (A), an indirect transitive (N), direct transitive (T), and pronominal (P). Each construction is composed of a set of components (subject, object, complement), which are also associated with semantic features, such as human, inanimate, animal, etc. For example, the word *apparaître* “to appear” has two different constructions:

A11: Inanimate subject, Human object

A31: Human subject, Complement

To solve these semantic and syntactic ambiguities, we created formal grammars that take into consideration the differences between syntactic constructions and the various types of components (human, inanimate, animal) in each construction. The elaboration of such a grammar is not an easy task because it is necessary to find a representation that considers recursion, as well as the length of the syntactic construction. Since the syntactic construction is formed by other components, its length is not known in advance. The proposed representation is illustrated by syntactic patterns.

5.2 Identification of syntactic patterns

In order to facilitate the identification of the transducers necessary for the disambiguation of verbs, we transformed the definition of each syntactic construction into a syntactic pattern. A major advantage of this method is that it makes it possible to arrange the various constituents of the construction in a linear manner that can be easily transformed into graphs (Fehri, 2012). According to the classification proposed by the LVF, we can construct 56 syntactic patterns that describe the locative verbs. A sample pattern is shown below:

Pattern N1j: = <human subject><verb><Prepositional complement, PREP = « en, dans »>

This pattern describes the different components of the *N1j* construction. According to this construction, we can distinguish the meaning of the verb when the latter appears in a sentence which has a human subject and a prepositional complement introduced by the preposition (PREP) *en* “in, to” or *dans* “in, inside”. We notice that the identified patterns have been refined.

5.3 Creation of transducers for locative verbs

For a reliable disambiguation system of locative verbs, we transformed each syntactic pattern into a transducer, which is manually built in NooJ. The meaning of the verb is related to the type of the construction and its constituents. The created transducer is represented in Figure 2.

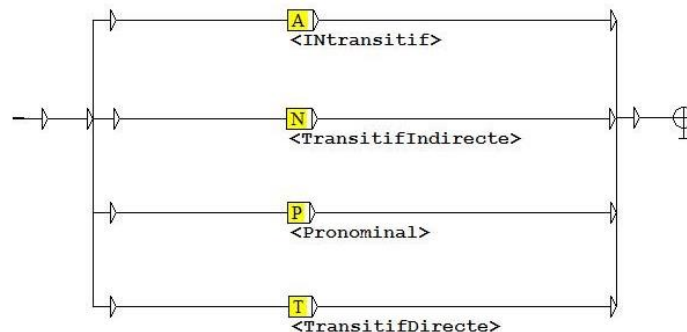


Figure 2. Main transducer of verb disambiguation

The transducer illustrated in Figure 2 contains four sub-graphs. Each sub-graph represents a category of syntactic construction -- intransitive (A), an indirect transitive (N), a pronominal (P), and a direct transitive (T),

6 Phase of Translation

Understanding a word sense is an inherent part of correct translation of words whose meanings depend on the context (Turdakov, 2010). For this reason, the proposed method of translation takes as input the disambiguated verbs returned by the phase of disambiguation. In addition, the quality of the translation of the verb does not depend only on the preservation of the meaning of the translated verb in the target language but also on the consideration of the tense in which the verb is conjugated and the inflected form of that verb. It also influences the quality of the translated verb. In this context, we built a syntactic grammar allowing the translation of each disambiguated locative verb by taking into account the tense of the verb and its inflected form.

However, the creation of such a grammar is not an easy task because this study deals with two very different languages: French, an Indo-European language, and Arabic, a Semitic language. We now examine some cases of translation of French locative verbs into Arabic. For example, unlike the French verb, the Arabic verb is not based on tense but on the aspect. In fact, the Arabic verbal system, is essentially represented by three aspects: accomplished, unaccomplished and imperative. Consider the following dictionary entry and French sentence:

accueillir 02: meaning = recevoir (to welcome) + AR= **اِسْتَقْبَلَ** istaqbala “welcomed”

La fille accueillait une amie dans sa maison.

The tense of the verb *accueillir* is in the imperfect. In Arabic, the imperfect is translated by **كَانَ يَفْعَلُ** Kāna yafaalu “has done” and the verb *accueillait* will be translated by **كَانَتْ تَسْتَقْبِلُ** Kānat tastaqbilu “have welcomed.”

Another difference between the two language can be seen in the case of gender. In French, the verb is conjugated with the same form for both the third person singular subject masculine pronoun *il* “he” and the feminine pronoun *elle* “she” – *il écrit* “he writes”, *elle écrit* “she writes”. In Arabic, however, the conjugation of the verb differs for the two subject pronouns. Indeed, *il écrit* is translated into Arabic as **كَتَبَ** kataba “he writes”, while *elle écrit* is translated by **كَتَبَتْ** katabat “she writes”.

Finally, the Arabic language specifically notes when there are two of something, called **الْمُتَنَّى** al-muṭanā “dual”. This form does not exist in French, which has only singular and plural. In Arabic, the dual number has its own suffixes, which totally differ from those of the plural.

After, this level of analysis, we move to the practice of automatic translation using the NooJ platform. In this step, we will integrate locative verbs into a bilingual dictionary. We also create a formal grammar, seen in Figure 3.

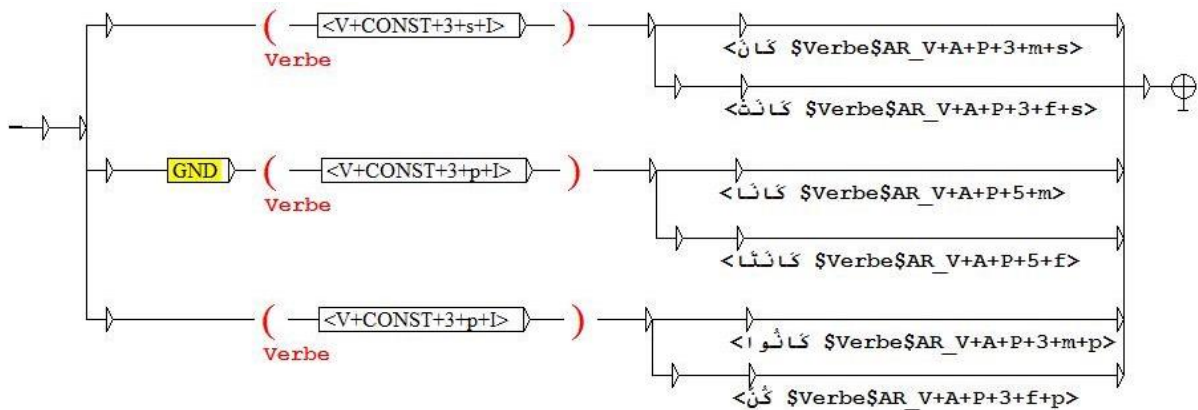


Figure 3. Transducer for the translation of conjugated verbs in imperfect

Figure 3 represents an extract of the transducer relating the translation of a verb when it is conjugated in the imperfect. The node that exists between brackets called *Verbe*, represents the form of the French verb to translate. For example, <V+CONST+3+s+I> indicates that the verb is conjugated in the imperfect (I) with the third person singular (3+s).

7 Experimentation and evaluations

WE tested the developed system is done with NooJ. As mentioned above, this platform uses already built syntactic local grammars (Fehri et al., 2011). Table 1 gives an idea about the dictionaries which we added to the resources in NooJ.

Dictionaries	Number of inputs	Annotation in the dictionary
<i>Human nouns</i>	250	N+Hum
<i>Inanimate nouns</i>	305	N+Chose
<i>Animal nouns</i>	480	N+Anl

Table 1. Added dictionaries

In addition to the dictionaries mentioned in Table 1, other NooJ dictionaries -- adjectives, determinants and verbs -- are used (Fehri et al., 2016).

7.1 Experimentation of disambiguation phase

To evaluate the disambiguation phase, we applied our resources to a corpus of journalistic articles published in *Le Monde Diplomatique* newspaper. The various topics that are covered deal with current world politics, economics, culture, sports, etc. The purpose of using a wide variety of subject areas is to have a broad coverage of the complex structures of sentences that constitute the context of the verbs to be disambiguated. This corpus contains 1,009 articles giving a total of 2,006,631 graphic words including 1,544 occurrences of locative verbs. These verbs are manually identified using NooJ queries.

Before	Seq.	After
ent que	les ouvriers baignaient dans une atmosphère/<V+N1j> Verbe= baignaient sens=immerger	saturé
allées à	l'hôtel Président de Yamoussoukro, qui abritait des militaires/<V+T3100> Verbe= abritait sens=loger	des fo
iriennes	. Les autorités sanitaires avaient fondé leur décision sur l'absence/<V+T13g0> Verbe= avaient fondé sens=baser	de ris
mission	. Plusieurs étudiantes assistaient à mes cours/<V+N1a> Verbe= assistaient sens=se présenter	chaqu
onomie	. Jason, 22 ans, habitait chez ses parents/<V+A11> Verbe= habitait sens=demeurer	à Bor
ordeaux	. Son nom avait figuré dans des agendas de personnes/<V+A31> Verbe= avait figuré sens=se trouver	impliq
rait que	le bateau ne finira pas en Méditerranée/<V+A31> Verbe= ne finira pas sens=aboutir	La sol
terrané.	La soldate, qui figurait sur bon nombre des photographies/<V+A11> Verbe= figurait sens=se trouver	de sév
Nord.	Trois milliards de gens vivaient dans les régions/<V+A11> Verbe= vivaient sens=habiter	côtier
angères	. Laila Freivalds, assistait tranquillement à une pièce de théâtre/<V+N1a> Verbe= assistait sens=se présenter	à Stoc
Jacob.	Beaucoup de réfugiés gravèrent des histoires d'une lourdeur infinie dans leurs mémoires/<V+T13j0> Verbe= gravèrent sens=imprimer	Les fer
moires.	Les femmes jetaient un voile sur leur tête/<V+T13g0> Verbe= jetaient sens=placer	une ét
voisins	. Sa petite maison se trouvait à une trentaine de mètres du bord de l'eau/<V+P3001> Verbe= se trouvait sens=se situer	. L'anci
de l'eau	. L'ancien chef de la délégation palestinienne à la conférence de Madrid apparaissait sur d'innombrables affiches/<V+A11> Verbe= apparaissait sens=se présenter	aux cc
. Depuis	. les frères Finaly vivaient en Israël/<V+A11> Verbe= vivaient sens=habiter	Il s'agit
ndiale (500 soldats japonais se trouvaient actuellement dans le sud de l'Irak/<V+P1001> Verbe= se trouvaient sens=se tenir	une ini
pout de	ce camp, qui abritait quelque 3 000 réfugiés/<V+T3100> Verbe= abritait sens=loger	sur les
le. Mais	Tony Blair, qui cachait sa passion sous ses bonnes manières/<V+T1308> Verbe= cachait sens=masquer	et sa t
l'adage	Les quelque 190 familles qui campaient dans les bois sous des toiles de plastique/<V+A11> Verbe= campaient sens=bivouaquer	ont toi
toutes	les personnes qui se trouvaient dans le village/<V+P1001> Verbe= se trouvaient sens=se tenir	à ce r

Figure 4. Concordance table of locative verb disambiguation

Each line of the concordance table shown in Figure 4 displays the sequence identified by the disambiguation transducer. For example, in the first line of Figure 4, the verb is assigned the meaning of the construction <N1j>. Consequently, the context of the verb *baignaient* (< *baigner* “to bathe”) takes on the meaning of *immerger* “immerse” in this particular context. To measure the performance of our tool, the following evaluation metrics are used: precision, recall and F-measure. The results obtained are described in Table 2.

	Nb. of occurrences (locative verbs)	Precision	Recall	F-measure
Newspaper <i>Le Monde diplomatique</i> (1009 texts)	1544	0.75	0.85	0.79

Table 2. Obtained results

The results shown in Table 2 reflect the existence of some issues that have not been resolved by our tool. These problems lie in the inability to disambiguate the verbs that appear in sentences whose subject is implicitly defined and so it is not possible to identify its nature. Take the example of the following phrase: *most being located near the islands*. In this example, the subject is represented by the noun *most*. This noun is not followed by anything that specifies the subject, such as *most people*. As a result, we cannot decide on the nature of the subject (human, animal or thing) and subsequently the verb *to locate* cannot be disambiguated.

In addition, the LVF construction codes describe the semantic traits of the arguments (subject, object, etc.) of each verbal use, as human, thing, animal, etc. For example, to distinguish the use *shelter 3* in *The embassy shelters the refugees* from *shelter 2* used in *The ambassador shelters refugees at home*, one must take into account that the noun *embassy* is a thing, while *ambassador* and *refugee* are human nouns. But such a purely lexical characterization of the arguments of a verb does not allow researchers to analyze many sentences that include semantic mechanisms, such as metaphor or metonymy, or to modify the function of nominal groups. Thus, for example, the noun *embassy*, principally lexicalized as a thing, can play the role of a human by metonymy: the phrase *The embassy shelters the refugees* must then be considered as ambiguous, depending on whether *embassy* designates the building itself or the embassy staff.

7.2 Disambiguation phase

The translation phase is applied to the disambiguated locative verbs returned by the disambiguation phase. Thus, as pointed out before, this verb translation from French into Arabic takes into account the verbal tense used and the inflected form of the verb. The results obtained are illustrated in Figure 5.

Before	Seq.	After
ceptions du travail relèvent que les ouvriers	بَاغَاوْا بِتَشْمُوْرُوْنَ / <كُنْ بِتَشْمُوْرُوْنَ>	dans une atmosphère saturée de
s à l'hôtel Président de Yamoussoukro, qui	كَاْنْ يَأْوِي / <كَاْنَتْ تَأْوِي>	des militaires des forces armées
ériennes ivoiriennes. Les autorités sanitaires	كَاْوَا قَدْ بَلَّوْا / <كُنْ قَدْ بَلَّوْا>	leur décision sur l'absence
faits par la commission. Plusieurs étudiantes	كَاْوَا يَحْضُرُوْنَ / <كُنْ يَحْضُرُوْنَ>	à mes cours chaque semaine
aine. En licence d'économie, Jason, 22 ans,	كَاْنْ يَسْكُنْ / <كَاْنَتْ تَسْكُنْ>	chez ses parents à Bordeaux
était chez ses parents à Bordeaux. Son nom	كَاْنْ قَدْ وَرَدَتْ / <كَاْنَتْ قَدْ وَرَدَتْ>	dans des agendas de personnes
sonnes impliquées. Il espérait que le bateau	حَاْنْ يَمِيْلْ / <حَاْنْ تَمِيْلْ>	en Méditerranée. La soldate, qui
ne finira pas en Méditerranée. La soldate, qui	كَاْنْ يَجِدْ / <كَاْنَتْ تَجِدْ>	sur bon nombre des photograph
Caroline du Nord. Trois milliards de gens	كَاْوَا يَطْمَئِنُوْنَ / <كُنْ يَطْمَئِنُوْنَ>	dans les régions côtières qui
stre des affaires étrangères, Laila Freivalds,	كَاْنْ يَحْضُرْ / <كَاْنَتْ تَحْضُرْ>	tranquillement à une pièce de
lm...' Antoine Jacob. Beaucoup de réfugiés	قَلَّدَ طَلَبُوْا / <قَلَّدَ طَلَبُوْا>	des histoires d'une lourdeur
leur infinie dans leurs mémoires. Les femmes	كَاْوَا يَضَعُوْنَ / <كُنْ يَضَعُوْنَ>	un voile sur leur tête
es des habitants voisins. Sa petite maison se	كَاْنْ يَجِدْ / <كَاْنَتْ تَجِدْ>	à une trentaine de mètres
on palestinienne à la conférence de Madrid	كَاْنْ يَحْضُرْ / <كَاْنَتْ تَحْضُرْ>	sur d'innombrables affiches, au
e guerre mondiale (500 soldats japonais se	كَاْوَا يَلْتَمِزُوْنَ / <كُنْ يَلْتَمِزُوْنَ>	actuellement dans le sud de
la télévision nationale. Mais Tony Blair, qui	كَاْنْ يَخْفِي / <كَاْنَتْ تَخْفِي>	sa passion sous ses bonnes
dans l'adage Les quelque 190 familles qui	كَاْوَا يُخَيِّمُوْنَ / <كُنْ يُخَيِّمُوْنَ>	dans les bois sous des

Figure 5. Concordance table of locative verb translation

Figure 5 shows the Arabic equivalent assigned to the ambiguous verbs shown in Figure 4 of the previous section. As a result, we have translated the verb according to the context in which it appears. Take

the example of the verb *baignaient* (< *baigner* “to bathe”) on the first line of the concordance table. This verb can have four different meanings, corresponding to four separate translations. Here, in this particular context, the verb *baigner* “to bathe” has been disambiguated with the meaning *immerger* “immerse” which is then translated into Arabic by **انغمَرَ** enghamara “immersed,” thus eliminating the other three translations.

Our translation tool gives satisfactory results, with a precision of 87%, which is comparable to well-known translators such as Google and Babylon that support the translation from French to Arabic.

French Locative Verbs	Obtained result (Our tool)	Obtained result (Google)	Obtained result (Babylon)
Les ouvriers baignaient dans une atmosphère	كَانُوا يَنْغَمِرُونَ	و استحم العمال في جو	baignaient في جو العمال
Les autorités sanitaires avaient fondé leur décision sur des preuves	كُنْ قَدْ بَيَّنَّ	و استندت السلطات الصحية الى قراراتها بشأن الأدلة	السلطات الصحية التي تقوم على قرار بشأن الأدلة
Son nom figurait dans des agendas	وَرَدَ	و كان اسمه في يوميات	اسمه وارد في مذكرات
Le bateau finira en Méditerrané	سَيَنْتَهِي	فإن القارب ينتهي في البحر الأبيض المتوسط	مركب ستنجز بالفعل في Méditerrané

Table3. Experimental results

As shown in Table 3, it is noted that Google did not use the correct sense of the verb. Moreover, the tense and the form of the verb have not been respected, such as for the verb *baigner* which appears in the sentence *Les ouvriers baignaient dans une atmosphère* “the workers were immersed in an atmosphere.” Our tool gives as result **كانوا ينغمرون** Kânû yanghamirûna “have immersed.” This translation corresponds to the meaning of the verb *baigner* in this particular context. Also, the translated verb respects the tense and the form of the source verb. However, Google translates the verb *baignaient* into **استحم** estahama “has bathed” where the verbal tense and the inflected form of the verb have not been respected. Babylon does not even produce any translation of the verb *baigner*.

Note also that there are verbs that are not recognized by either Google or Babylon and therefore no translation is provided for the verb *appear* in the sentence *His name appeared in diaries*. The results obtained reflect the effectiveness of our translation tool. Indeed, a specific treatment for the locative verbs allows us to understand the meaning of the verb according to the statements produced in the source language and thus to propose a satisfactory translated version that takes into account the characteristics of the verb.

8 Conclusion and future perspectives

In the present work, we have developed a rule-based system for disambiguating French locative verbs and their translation into Arabic. The proposed disambiguation approach is based on the LVF dictionary. We have shown the effectiveness of this dictionary as a relevant resource for the task of disambiguation. As for the translation, it consists in translating the already disambiguated French locative verbs into Arabic. This translation takes into account the verbal time used as well as the inflected form of the verb. The experimentation and the evaluation were done using the linguistic platform NooJ, with satisfactory results.

In the future, we intend to study the syntactic and semantic phenomena in more detail. We also plan to generalize our phase of disambiguation to all the verbs. As for translation, we plan to take into account the linguistic context of the verb in order to specify the most appropriate verbal tense with the syntax of the Arabic sentence.

References

- Marianna Apidaniaki. 2008. *Acquisition automatique de sens pour la désambiguïsation et la sélection lexicale en traduction*. PhD thesis, PARIS University. DIDEROT (Paris 7).
- Laurent Audibert. 2003. *Outils d'exploration de corpus et désambiguïsation lexicale automatique*. Autre [cs.OH]. University of Provence - Aix-Marseille I, French.
- Fatiha Boubekeur-Amirouche. 2008. *Contribution à la définition de modèles de recherche d'information flexibles basés sur les CP-Nets*. PhD thesis in computer science, University of Toulouse.
- Caroline Brun, Bernard Jacquemin, and Frédérique Segond. 2001. *Exploitation de dictionnaires électroniques pour la désambiguïsation sémantique lexicale*. Automatic Language Processing, ATALA, 42 (3): 677-691.
- Vimal Dixit, Kamlesh Dutta and Pardeep Singh. 2015. *Word Sense Disambiguation and Its Approaches*. CPUH-Research Journal: 2015, 1(2): 54-58.
- Héla Fehri. 2012. *Reconnaissance automatique des entités nommées arabes et leurs traductions vers le français*. PhD thesis in computer science, University of Sfax.
- Héla Fehri, Kais Haddar, and Abdelmajid Ben Hamadou. 2011. *Recognition and Translation of Arabic Named Entities with NooJ Using a New Representation Model*. Proceedings of the 9th International Workshop on Finite State Methods and Natural Language Processing, Blois, France, pages 134–142.
- Héla Fehri, Mohamed Zaidi, and Kamel Boudhina. 2016. *Création d'un dictionnaire des verbes NooJ open source pour la langue Arabe*. Report of the end of studies project. Higher Institute of Management of Gabes.
- Gaston Gross and André Clas. 1997. *Synonymie, polysémie et classes d'objets*. Meta 421 (1997.DOI: 10.7202/002977ar): 147–154.
- Mohammad Nessereddine, Andon Tchechmedjiev, Hervé Blanchon, and Didier Schwab. 2015. *Création rapide et efficace d'un système de désambiguïsation lexicale pour une langue peu dotée*. 22nd Automatic Processing of Natural Languages, Caen. <http://talnarchives.atala.org/TALN/TALN-2015/taln-2015-long-008.pdf>.
- Hinrich Schütze. 1998. Automatic word sense discrimination. *Computational Linguistics: Special Issue on Word Sense Disambiguation*, 24 (1): 97–123.
- Morgane Sénéchal and Dominique Willems. 2007. Classes verbales et régularités polysémiques : le cas des verbes trivalenciels locatifs. *French language* (no. 153): 92-110.
- Max Silberztein. 2010. La formalisation du dictionnaire LVF avec NooJ et ses applications pour l'analyse automatique de corpus. *Languages* (No. 179-180), DOI 10.3917 / lang.179.0221: 221-241.
- Andon Tchechmedjiev. 2012. État de l'art : mesures de similarité sémantique locales & algorithmes globaux pour la désambiguïsation lexicale à base de connaissances. *Proceedings of the joint JEP-TALN-RECITAL conference volume 3: RECITAL*, Grenoble, pages 295–308.
- D.Yu Turdakov. 2010. Word Sense Disambiguation Methods. *Programming and Computer Software*. Vol. 36, No. 6: 309–326.
- Florentina Vasilescu. 2003. *Désambiguïsation de corpus monolingues par des approches de type Lesk*. Master of Science in Computer Science, University of Montréal.