A possibilistic approach for automatic word sense disambiguation

Oussama Ben Khiroun[1], Bilel Elayeb[1,2], Ibrahim Bounhas[3], Fabrice Evrard[2],
and Narjès Bellamine Ben Saoud[1]

[1]RIADI Research Laboratory, ENSI Manouba University, 2010 Tunisia.
oussama.ben.khiroun@gmail.com, Bilel.Elayeb@riadi.rnu.tn, Narjes.Bellamine@ensi.rnu.tn
[2]IRIT-ENSEEIHT, 02 Rue Camichel, 31071 Toulouse Cedex 7, France.
Fabrice.Evrard@enseeiht.fr
[3]Department of Computer Science, Faculty of Sciences of Tunis, 1060 Tunis, Tunisia
Bounhas.Ibrahim@yahoo.fr

Abstract

This paper presents and experiments a new approach for automatic word sense disambiguation (WSD) applied for French texts. First, we are inspired from possibility theory by taking advantage of a double relevance measure (possibility and necessity) between words and their contexts. Second, we propose, analyze and compare two different training methods: judgment and dictionary based training. Third, we summarize and discuss the overall performance of the various performed tests in a global analysis way. In order to assess and compare our approach with similar WSD systems we performed experiments on the standard ROMANSEVAL test collection.

Keywords: Word Sense Disambiguation, Semantic Dictionary of Contexts, Possibility Theory.

1. Introduction

Word Sense Disambiguation (WSD) is the ability to identify the meaning of a word in its context in a computational manner. A lexical semantic disambiguation allows to select in a predefined list the significance of a word given its context. In fact, the task of semantic disambiguation requires enormous resources such as labeled corpora, dictionaries, semantic networks or ontologies. This task is important in many fields such as optical character recognition, lexicography, speech recognition, natural language comprehension, accent restoration, content analysis, content categorization, information retrieval and computer aided translation [13] [14].

The problem of WSD has been considered as a difficult task in the field of Natural Language Processing. In fact, a reader is frequently faced to problems of ambiguity in information retrieval or automatic translation tasks. Indeed, the main idea on which were based many researches in this field is to find relations between an occurrence of a word and its context

which will help identify the most probable sense of this occurrence [1][2].

We discuss in this paper the contribution of a new approach for WSD. We presuppose that combining knowledge extracted from corpora and traditional dictionaries will improve disambiguation rates. We also show that this approach may perform satisfactory results even without using manually labeled corpora for training. We also propose to apply possibility theory as an efficient framework to solve the WSD problem seen as a case of imprecision. Indeed, WSD approaches need training and matching models which compute the similarities (or the relevance) between senses and contexts. Existing models for WSD are based on poor, uncertain and imprecise data. Whereas, possibility theory is naturally designed to this kind of applications; because it makes it possible to express ignorance and to take account of the imprecision and uncertainty at the same time. For example a recent work of Ayed et al. (2012) [23][24] which have proposed possibilistic approach for the morphological disambiguation of arabic texts showed the contribution of possibilistic models compared to probabilistic ones. That is, we evaluate the relevance of a word sense given a polysemous sentence proposing two types of relevance: plausible relevance and necessary relevance.

This paper is structured as follows. First, we give an overview of the main existing WSD approaches in section 2. Section 3 briefly recalls possibility theory. Our approach is detailed in section 4. Subsequently, a set of experimentations and comparison results are discussed in section 5. Finally, we summarize our findings in the conclusion and propose some directions for future research.

2. Related Works

In this literature review, we briefly cite the most important methods which allowed to clarify the main issues in WSD. We mainly focus on the limits of traditional dictionaries in WSD process. In fact, the most popular WSD approaches are based on traditional dictionaries or thesauruses (such as WordNet), which are quite similar in terms of sense organization. Indeed, dictionaries were made for a human use and are not suitable for automatic treatments, thus missing accurate information useful for WSD. This fact is confirmed by Véronis [16][17] who argues that it is not possible to progress in WSD while dictionaries do not include in their definitions distributional criteria or surface indices (syntax, collocations, etc). In addition, the inconsistency of the dictionaries is well-known for lexicographers.

For these multiple reasons, many researchers proposed to build new types of dictionaries or to restructure traditional dictionaries. For example, Reymond [22] proposed to build a "distributional" dictionary based on differential criteria. The idea is to organize words in lexical items having coherent distributional properties. This dictionary contained initially the detailed description of 20 common nouns, 20 verbs and 20 adjectives. It enabled him to manually label each of the 53000 occurrences of these 60 terms in the corpus of the project SyntSem (Corpus of approximately 5.5 million words, composed of texts of various kinds). This corpus is a starting resource to study the criteria of automatic semantic disambiguation

since it helps implement and evaluate algorithms of WSD.

Audibert [15] worked on Reymond's dictionary to study different criteria of disambiguation (co-occurrence, domain information, synonyms of co-occurring words and so on). In the same perspective, Véronis, [17] used a graph of co-occurrence to automatically determine the various usages of a word in a textual base. His algorithm searches high density zones in the graph of co-occurrence and allows to isolate non frequent usages. Thus, Véronis applied the advice of Wittgenstein: "Don't look for the meaning, but for the use". In fact, co-occurrence-based approaches generate much noise since unrelated words may occur in the same sentence. We also find that none of these methods treated in a sufficient manner the problem of lexicon organization. Even the methods based on computing the similarities do not seek to represent the semantic distances between senses and do not manage to correctly organize the obtained senses. However, several research works tried to resolve the problem of polysemy on the level of dictionary. Gaume et al. (2004) [18] used a dictionary as information source to discover relations between lexical items. His work is based on an algorithm which computes the semantic distance between the words of the dictionary by taking into account the complete topology of the dictionary, which gives him a greater robustness. This algorithm makes it possible to solve the problem of polysemy which exists in the definitions of the dictionary. He started to test this approach on the disambiguation of the definitions of the dictionaries themselves. But this work is limited to disambiguate nouns, using only nouns or nouns and verbs.

Our approach is supported by a semantic space where the various senses of a word are organized and exploited. Indeed, computing the sense of a sentence is a dynamic process during which the senses of the various words are mutually influenced and which leads simultaneously to the determination of the sense of each word and the global sense of the sentence. A distance between contexts and word senses is used to find the correct sense in a given sentence. Our work uses possibilistic networks to compute a preliminary rate of ambiguity of each sentence and to match senses to contexts. That is, we start by recalling principles of possibility theory in the following section.

3. Possibility Theory

The possibility theory introduced by Zadeh (1978) [10] and developed by several authors, handles uncertainty in the interval [0,1] called the possibility scale, in a qualitative or quantitative way. This section briefly reviews basic elements of possibility theory, for more details see [3][4][21].

3.1 Possibility distribution

Possibility theory is based on possibility distributions. The latter, denoted by π, are mappings from Ω (the universe of discourse) to the scale [0,1] encoding partial knowledge on the world. The possibility scale is interpreted in two ways. In the ordinal case, possibility values only

reflect an ordering between possible states; in the numerical scale, possibility values often account for upper probability bounds [3][4][21].

Probability distribution mainly differs from possibility distribution because it requires that the probability sum of elements in the universe of discourse is equal to 1, but this restriction is not necessary in the case of possibility theory. Besides, the probability of the complement of a given event is relevant to provide the probability of this event in probability theory. But, it is not the same thing in possibility theory, which involves non-additive measures. When we use probabilities in uncertainty representation, it is required to list an exhaustive set of mutually exclusive alternatives. This is the fundamental difficulty to use probabilities in this case. In reality, an expert cannot provide events that are exhaustive and mutually exclusive due to the increasing of his/her knowledge along time, and so uncertainty about the situation decreases.

Furthermore possibility distributions may be more expressive in some situations and is able to distinguish between problems, *ambiguity* and *ignorance* whereas probability distributions can only represent *ambiguity*. In particular, the distribution $\pi(\omega) = 1; \forall\ \omega \in \Omega$ express a total ignorance which reflects the absence of any relevant information. However in probability theory, complete ignorance is modeled by a uniform distribution which results in assigning *equal* weights $p(\omega) = 1/n; \forall\ \omega \in \Omega$ for each event although no justification can explain this arbitrary assignment. For more reading, we can refer to [3][4].

3.2 Possibility and necessity measures

While other approaches provide a unique relevance value, the possibility theory defines two measures. A possibility distribution π on Ω enables events to be qualified in terms of their plausibility and their certainty, in terms of possibility and necessity measures respectively. In our context of WSD, the possible relevance allows rejecting non relevant senses. The necessary relevance permits to reinforce possibly relevant senses.

- The possibility of an event A relies on the most normal situation in which A is true.

$$\Pi(A) = \max_{x \in A} \pi(x) \tag{1}$$

- The necessity of an event A reflects the most normal situation in which A is false.

$$N(A) = \min_{x \notin A}(1 - \pi(x)) = 1 - \Pi(\neg A) \tag{2}$$

The width of the gap between $N(A)$ and $\Pi(A)$ evaluates the amount of ignorance about A. Note that $N(A) > 0$ implies $\Pi(A) = 1$. When A is a fuzzy set this property no longer holds but the inequality $N(A) \leq \Pi(A)$ remains valid [3][4][21].

3.3 Possibilistic Networks

A directed possibilistic network (PN) on a variable set V is characterized by a graphical and a numeric component. The first one is a directed acyclic graph. The graph structure encodes independence relation sets just like Bayesian nets [19][20]. The second component quantifies distinct links of the graph and consists of the conditional possibility matrix of each node in

the context of its parents. These possibility distributions should respect normalization. For each variable V:

- If V is a root node and $dom(V)$ the domain of V, the prior possibility of V should satisfy: $\max_{v \in dom(V)} \Pi(v) = 1$; $\hspace{4cm}$ (3)

- If V is not a root node, the conditional distribution of V in the context of its parents context satisfy: $\max_{v \in dom(V)} \Pi(v|Par_V) = 1$; $\quad Par_V \in dom(Par_V)$ $\hspace{2cm}$ (4)

Where: $dom(V)$: domain of V; Par_V : value of parents of V; $dom(Par_V)$: domain of parent set of V.

In this paper, possibilistic networks are exploited to compute relevance of a correct sense of a polysemous word given the context.

4. The Proposed approach

Our approach tries to avoid the limits of traditional dictionaries by combining them with knowledge extracted from corpora and organized as a Semantic Dictionary of Contexts (SDC). Thus, the richness of traditional dictionaries is improved by contextual knowledge linking words to their contexts. WSD is also seen as a classification task where we have training and testing steps. In the training step, we need to learn dependencies between senses of words and contexts. This may be performed in labeled corpora (Judgment-based training) leading to a semi-automatic approach. We may also weight these dependencies directly from a traditional dictionary (Dictionary-based training), what may be considered as an automatic approach. In this case, we need to organize all the instances in such a way that improves classification rates. In this paper, we propose to sort the instances by computing an ambiguity rate (sf. section 4.2). In the testing step, the distance between the context of an occurrence of a word and its senses is computed in order to select the best sense.

We present in the next sections the formulae for computing the DPR and the ambiguity rate.

4.1 The Degree of Possibilistic Relevance (DPR)

Supposing that we have only one polysemous word in a sentence *ph*, let us note *DPR(S$_i$|ph)* the Degree of Possibilistic Relevance of a word sense S_i given *ph*. Let us consider that *ph* is composed of *T* words: $ph = (t_1, t_2, ..., t_T)$. We evaluate the relevance of a word sense S_i given a sentence *ph* by a possibilistic matching model of Information Retrieval (IR) used in [5][21]. In this case, the goal is to compute a matching score between a query and a document. In the case of WSD, the relevance of a sense given a polysemous sentence is modeled by a double measurement. The possible relevance makes it possible to reject the irrelevant senses. But, the necessary relevance makes it possible to reinforce relevance of the restored word senses, which have not been rejected by the possibility.

In our case, possibilistic network links the word sense (S_i) to the words of a given a polysemous sentence ($ph_i = (t_1, t_2, ..., t_T)$) as presented in figure 1.
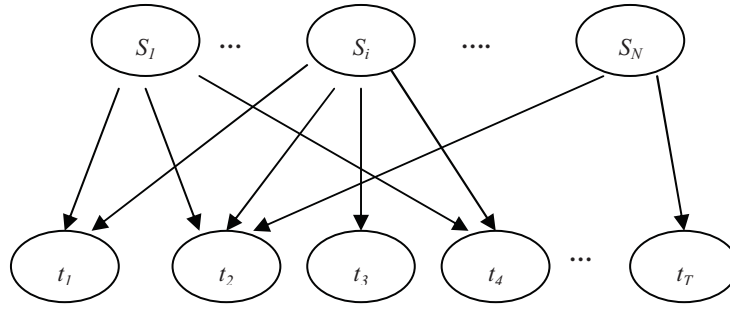
Figure 1. Possibilistic network of WSD approach

The relevance of each word sense (S_j), giving the polysemous sentence (ph_i) is calculated as follows:

According to Elayeb et al. (2009) [5], the possibility $\Pi(S_j|ph)$ is proportional to:

$$\Pi'(S_j|ph) = \Pi(t_1|S_j)*\ldots*\Pi(t_T|S_j) = nft_{1j}*\ldots*nft_{Tj} \quad (5)$$

With $nft_{ij} = tf_{ij}/\max(tf_{kj})$: the normalized frequency of the term t_i in the sense S_j

And tf_{ij} = (number of occurrence of the term t_i in S_j/number of terms in S_j)

The necessity to restore a relevant sense S_j for the sentence ph, denoted $N(S_j|ph)$, calculated as the following:

$$N(S_j|ph) = 1 - \Pi(\neg S_j|ph) \quad (6)$$

Where: $\Pi(\neg S_j|ph) = (\Pi(ph|\neg S_j)*\Pi(\neg S_j))/\Pi(ph) \quad (7)$

At the same way $\Pi(\neg S_j|ph)$ is proportional to:

$$\Pi'(\neg S_j|ph) = \Pi(t_1|\neg S_j)*\ldots*\Pi(t_T|\neg S_j) \quad (8)$$

This numerator can be expressed by:

$$\Pi'(\neg S_j|ph) = (1-\phi S_{1j})*\ldots*(1-\phi S_{Tj}) \quad (9)$$

Where: $\phi S_{ij} = \log_{10}(nCS/nS_i)*(nft_{ij}) \quad (10)$

With: nCS = Number of senses of the word in the dictionary.

nS_i = Number of senses of the word containing the term t_j. This includes only senses which are in the SDC and does not cover all the senses of t_i which are in the traditional dictionary.

We define the Degree of Possibilistic Relevance (*DPR*) of each word sense S_j, giving a polysemous sentence ph by the following formula:

$$DPR(S_j|ph) = \Pi(S_j|ph) + N(S_j|ph) \quad (11)$$

The preferred senses are those which have a high value of *DPR($S_j|ph$)*.

4.2 The Ambiguity rate of a polysemous sentence

We compute the ambiguity rate of a polysemous sentence *ph* using the possibility and necessity values as follow: (i) We index the definitions of all the possible senses of the ambiguous word; (ii) We use the index of each sense as a query; (iii) We evaluate relevance of the sentence given this query using a possibilistic matching model; and (iv) A sentence is considered as very ambiguous if it is relevant for many senses or if it is not relevant for any

one. In other words, the relevance degrees of the sentence for all the senses are almost equal. Therefore, the ambiguity rate is inversely proportional to standard deviation value:

$$\text{Ambiguity\_rate}(ph) = 1 - \sigma(ph) \tag{12}$$

Where $\sigma(ph)$: standard deviation of $DPR(S_i|ph)$ values corresponding to each sense of ambiguous word contained in the polysemous sentence ph.

$$\sigma(ph) = \sqrt{1/N * \sum_i (DPR(S_i \mid ph) - S)^2} \tag{13}$$

Where S is the average of $DPR(S_i|ph)$ and N is the number of possible senses in the dictionary.

4.3 Illustrative example

Let us consider the polysemous word M, which has two senses S_1 and S_2 such as:

S_1 is indexed by the three terms $\{t_1, t_2, t_3\}$ and S_2 is indexed by $\{t_1, t_4, t_5\}$.

Let us consider also the polysemous sentence $ph = (M, t_2, t_4, t_5)$, which contains only one polysemous word (M) in order to simplify the calculus.

We have : $\Pi(S_1|ph) = nf_{(M, S_1)} * nf_{(t_2, S_1)} * nf_{(t_4, S_1)} * nf_{(t_5, S_1)} = 0*(1/3)*0*0 = 0$

With: $nf_{(M, S_1)}$ is the normalized frequency of M in the first sense S_1.

$\Pi(S_2|ph) = nf_{(M, S_2)} * nf_{(t_2, S_2)} * nf_{(t_4, S_2)} * nf_{(t_5, S_2)} = 0*(1/3)*0*0 = 0$

We have frequently $\Pi(S_j|ph) = 0$, except if all the words of the sentence exist in the index of the sense.

On the other hand, we have a not null values of $N(S_j|ph)$:

$N(S_1|ph) = 1 - [(1-\phi(S_1, M)) * (1-\phi(S_1, t_2)) * (1-\phi(S_1, t_4)) * (1-\phi(S_1, t_5))]$

$nf(S_1, M) = 0$, so $\phi(S_1, M) = 0$; $\phi(S_1, t_2) = \log_{10}(2/1)*1/3 = 0,1$; $\phi(S_1, t_4) = \log_{10}(2/1)*0 = 0$; $\phi(S_1, t_5) = 0$

So: $N(S_1|ph) = 1 - [(1-0)* (1-0,1)* (1-0)* (1-0)] = 1 - [1* 0,9* 1* 1] = 0,1$.

And $DPR(S_1|ph) = 0,1$

$N(S_2|ph) = 1 - [(1-\phi(S_2, M)) * (1-\phi(S_2, t_2)) * (1-\phi(S_2, t_4)) * (1-\phi(S_2, t_5))]$

With: $\phi(S_2, M) = 0$ because $nf_{(S_2, M)} = 0$; $\phi(S_2, t_2) = 0$; $\phi(S_2, t_4) = \log_{10}(2/1)*1/3 = 0,1$; $\phi(S_2, t_5) = 0,1$.

So: $N(S_2|ph) = 1 - [(1-0)* (1-0)* (1-0,1)* (1-0,1)] = 1 - [1* 0,9* 0,9* 1] = 0,19$.

$DPR(S_2|ph) = 0,19 > DPR(S_1|ph)$

We remark that the polysemous sentence ph is more relevant for S_2 than S_1 because it contains two terms of the second sense S_2 (t_4, t_5) and only one term of the sense S_1 (t_2).

The average is $S = (0,1 + 0,19)/2 = 0,145$. The Standard Deviation $= (1/2 *((0,1 - 0,145)^2 + (0,19 - 0,145)^2))^{1/2} = 0,045$ and the Ambiguity Rate $= (1-$ Standard Deviation$) = 0,955$.

Let us notice in this example that the polysemous sentence ph is very ambiguous because two values 0,1 and 0,19 are very close.

5. Experimentation and results

This section introduces the test collection used in our experiments (cf. section 5.1). To improve our assessment, we performed two types of evaluation in the training step: the judgment-based training and the dictionary-based training (cf. sections 5.3 and 5.4 respectively). We analyze and interpret our results in section 5.5.

5.1 ROMANSEVAL test collection

We used in our experiments the ROMANSEVAL standard test collection which provides necessary tools for WSD including: (1) a set of documents (issued from the Official Journal of the European Commission); and (2) a list of test sentences including ambiguous words. The set of documents consists of parallel texts in 9 languages part of the Official Journal of the European Commission (Series C, 1993). Texts (numbering several thousand) consist of written questions on a wide range of topics and corresponding responses from the European Commission. The total size of the corpus is approximately 10.2 million words (about 1.1 million words per language), which were collected and prepared within MULTEXT-MLCC projects [6].

These texts were prepared in order to obtain a standard test collection. The corpus was split into words labeled with, in particular, categorical labels to distinguish the names N, adjectives A and verbs V. Then the 600 most frequent words (200 N, 200 A, 200 V) were extracted, and their contexts of occurrence. These words were annotated in parallel by 6 students in Linguistics, in accordance with the sense of the French dictionary "Le Petit Larousse", each occurrence of a word that can receive a label of a sense, several or none. After this first step, the 60 most polysemous words have been preserved (20 N, 20 A, 20 V). The body offered to participants for the experiment was therefore made up of 60 words and 3624 contexts in which they appear each with about 60 word occurrences.

5.2 Experimental scenarios

We performed three stages of tests as explained below. For each test, we prepared an XML Semantic Dictionary of Contexts (SDC). It is used as a training subset from the sentences to be evaluated in ROMANSEVAL corpus. For each parsed sentence S and given a polysemous word W, we link words of S with the correct sense of W. The "correct sense" may be identified from the tags of the corpus or using context-independent knowledge from the traditional dictionary. Thus, two subset selection methods for building the SDC are described in the following (cf. section 5.3 and section 5.4). To assess our system, we compute the accuracy rate for each word be using the *agree* and *kappa* [11][12] metrics which are computed as follows:

$$Agree = \frac{|\{S_i \in \Delta, \ where S_i^{system} = S_i^{judges}\}|}{|\{S_i \in \Delta\}|} \tag{14}$$

Where: Δ : The set of judged senses corresponding to test sentences. S_i^{system} : The selected

sense by DPR measure (computed by the system). S_i^{judges} : Sense attributed by judges.

The *Kappa* measure is based on the difference between how much agreement is actually present ("observed" agreement) compared to how much agreement would be expected to be present by chance alone ("expected" agreement) as follow [7]:

$$k = \frac{Pobserved - Pexpected}{1 - Pexpected} \qquad (15)$$

Kappa measure takes into account the agreement occurring by chance and is considered as a refined value. According to Landis and Koch [8], *Kappa* values between 0–0.2 are considered slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement.

5.3 Judgment-based training

To fill the XML SDC, we have applied the cross validation method. In each test case of the 10 iterations, we select 90% of sentences randomly and enlarge the training semantic dictionary by voted contexts. The 10% remaining ones are used in test by searching the most suitable context from the trained data. We applied there the DPR measure described in section 4.1. Averages agree values are presented in the following figures 2, 3 and 4.
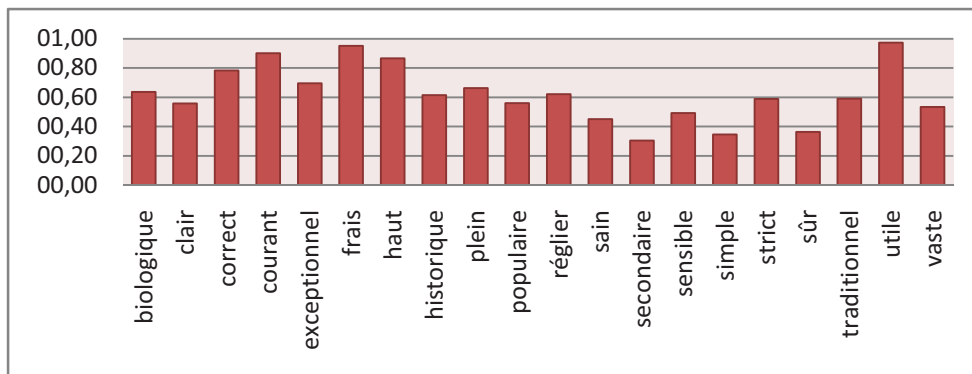

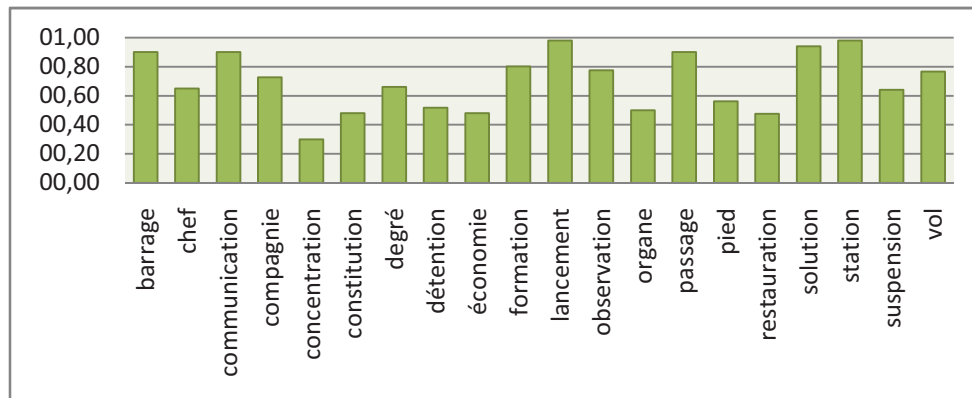Figure 2. Adjectives mean agrees for judgment-based training WSD method

Figure 3. Nouns mean agrees for judgment-based training WSD method
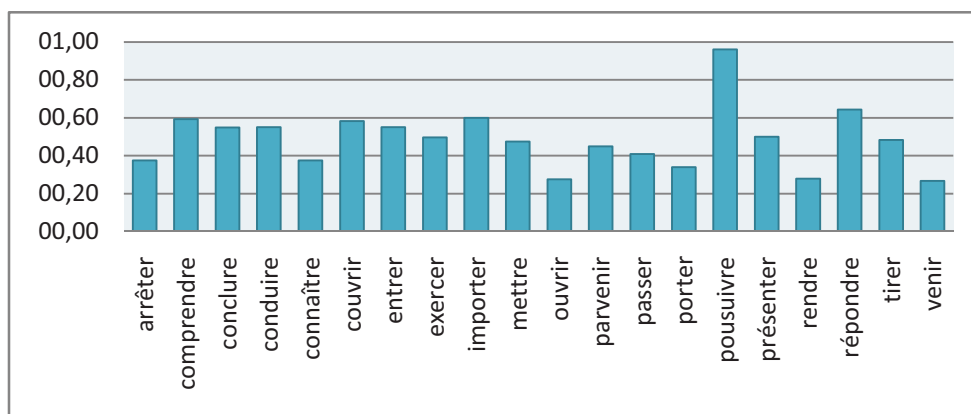
Figure 4. Verbs mean agrees for judgment-based training WSD method

As a first interpretation of these histograms, we conclude that more a word is frequent in the corpus and has few senses, the more is its accuracy rate. Thus, verbs represent the most ambiguous words, because they have fewer occurrences in the corpus. On the other hand, nouns (except for some ones) are less ambiguous, because they are more frequent. The accuracy rate depends also on the characteristics of the corpus. For example, we discuss the case of "constitution" which has a weak accuracy rate compared to other nouns. This word has many meanings ("constitution" has 6 different meanings: (1) constitution (constitution), (2) mise en place (establishment), (3) incorporation (incorporation), (4) règle (rule), (5) habitude (habit) and (6) code (code)). The legal discussion subjects in ROMANSEVAL articles contribute in increasing ambiguity of such words (the same interpretation is applied on "économie" word (meaning: économie (economy), finances (economics), épargne (saving), élevage (thrift or husbandry)).

5.4 Dictionary-based training

In this training method, senses are associated by the system (no more default judgments as the previous method). For each sentence, to be evaluated that contains an ambiguous word; one sense is attributed after computing the DPR values of each definition entry in the dictionary "Le Petit Larousse". Sense having the greatest DPR is considered as the best one to fit the sentence.

Sentences are therefore sorted in descendant (resp. ascendant) order by ambiguity rate (cf. section 4.2). Having the sorted list of sentences in descendent (resp. ascendant) order, the 80% most (resp. less) ambiguous sentences were used to build the SDC; the 20% remaining ones were used for test purpose.

Figures 5, 6 and 7 present the mean agrees for dictionary-based training WSD methods (descendant and ascendant sentences ambiguity rate) for respectively adjectives, nouns and verbs.
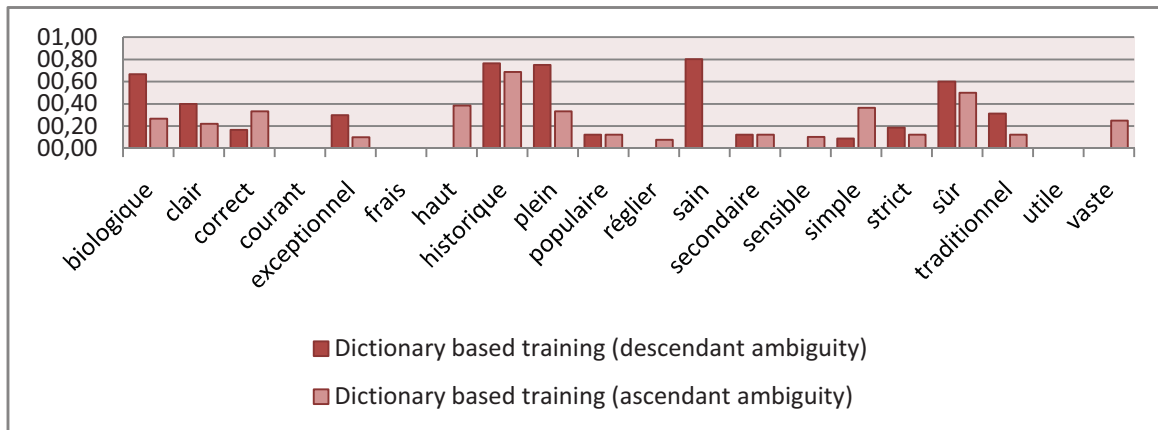
Figure 5. Adjectives mean agrees for dictionary based training WSD methods (descendant and ascendant sentences ambiguity)
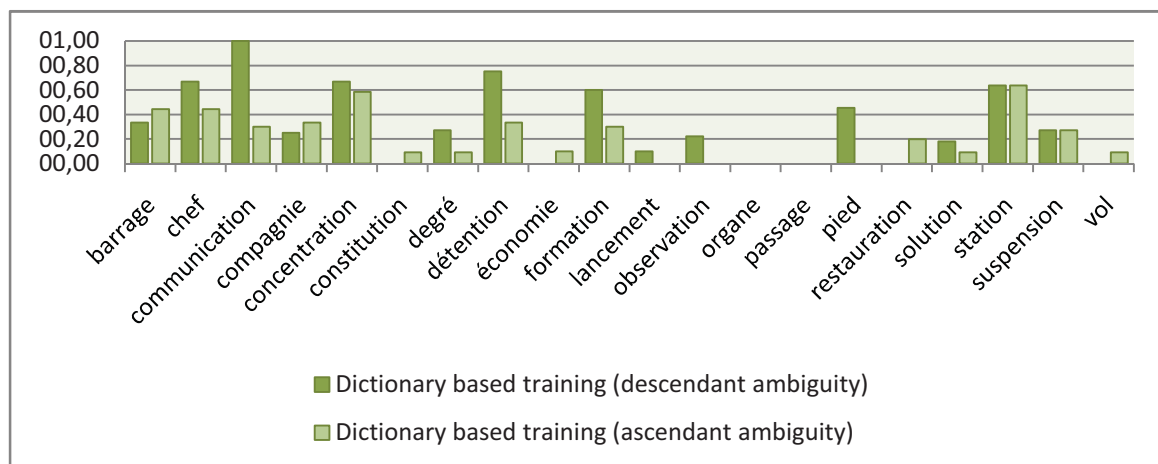

Figure 6. Nouns mean agrees for dictionary based training WSD methods (descendant and ascendant sentences ambiguity)
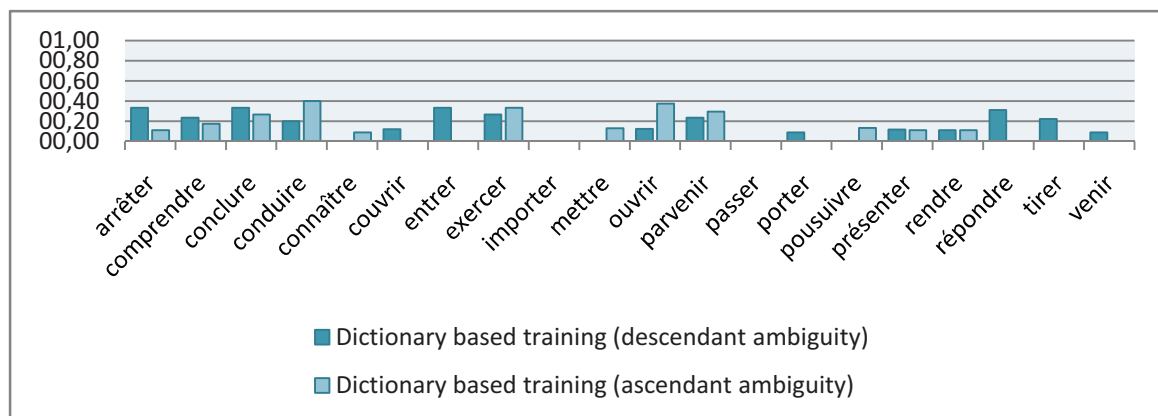

Figure 7. Verbs mean agrees for dictionary-based training WSD methods (descendant and ascendant sentences ambiguity)

These experiments confirm that training data should start from the most ambiguous sentences to the less ones (descendent ambiguity rate order). We should notice that the small accuracy rates are caused by the system selection of senses while building the SDC in the training step. However, this constitutes a first attempt for full automatic WSD.

5.5 Discussion and interpretation

This section summarizes and discusses the overall performance of the various performed tests. Figure 8 shows the mean agree rates over the three methods by Part-Of-Speech.
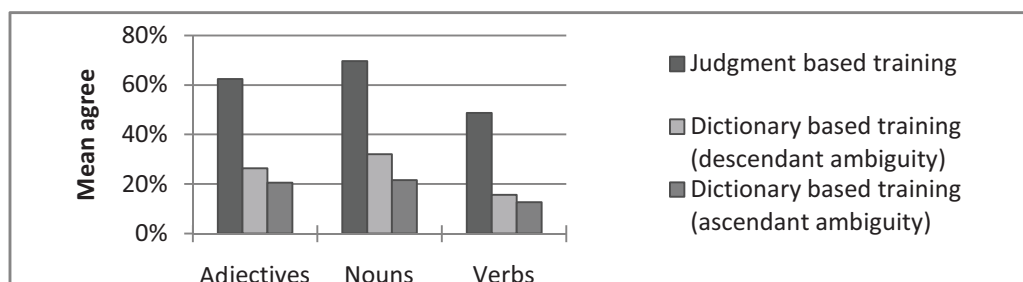

Figure 8. Mean agree rates over the three possibilistic WSD methods by Part-Of-Speech

We remark that the judgment-based approach performed better than dictionary-based approaches because it exploits human knowledge to build the SDC. However dictionary-based is a full automatic approach which may be used when labeled corpora are unavailable. In this case, it is more suitable to start from the most ambiguous sentences.

Then, we compare the performance of the best possibilistic method (judgment-based training) with five other WSD systems participating in the French exercise [6]. These systems are developed respectively by **EPFL** (Ecole Polytechnique Fédérale de Lausanne), **IRISA** (Institut de recherche en informatique et Systèmes Aléatoire, Rennes), **LIA-BERTIN** (Laboratoire d'informatique, Université d'Avignon, and BERTIN, Paris), and **XRCE (**Xerox Research Centre Europe, Grenoble). A comparative study between these systems is available at [6]. Figure 9 shows the values of *agree* and *Kappa* metrics (often used to evaluate WSD approaches) for these five systems and our approach (POSS).
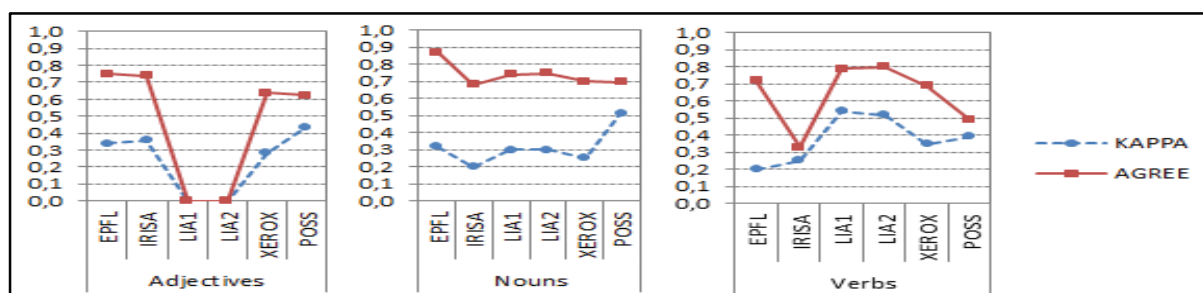

Figure 9. Mean *agree* and *Kappa* results by Part-Of-Speech

According to figure 9, the agree performance using POSS (especially for verbs) is worse than the other systems. We should also recognize that the *agree* metric does not provide alone accurate evaluation of WSD systems. Studying the agreement between two or more observers should include a statistic that takes into account the fact that observers will sometimes agree or disagree simply by chance [12]. The kappa statistic is the most commonly used statistic for this purpose. When focusing on the results over all Part-Of-Speech (cf. Figure 10), our system is distinguished from other systems for the *Kappa* value: in spite of having a medium

agree mean in comparison with other systems, agreement between our system and other judges is not a stroke of chance according to a moderate *Kappa* value (0.45).
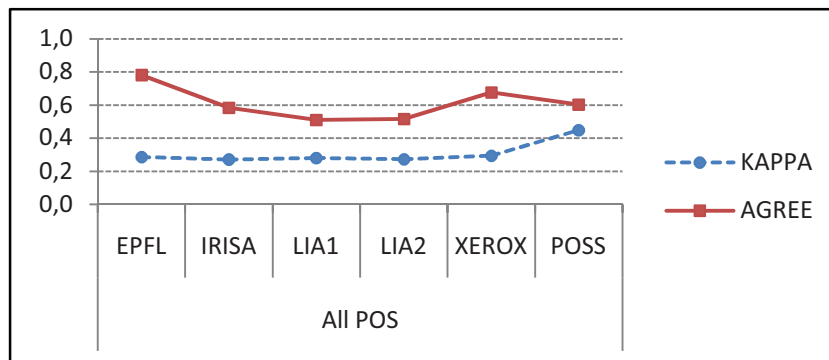

Figure 10. Mean *agree* and *Kappa* results for all Part-Of-Speech

According to *Kappa* results, the good agreement performance of the probabilistic WSD is by chance in many words: for example mean agree of the word "*pied*" (foot) is about 0.68 while *Kappa* measure is under 0.2. Thus, we notice that the possibilistic approach is finer than the probabilistic state-of-the-art systems. This explained by the fact possibility and necessity measures increase the relevance of correct senses and penalize the scores the remaining ones.

We should here notice that disagreement among the human judges who prepared sense tagging of the ROMANSEVAL benchmark is so important according to [9]: *Kappa* ranges between 0.92 (noun "detention") and 0.007 (adjective "correct"). In other terms, there is no more agreement than chance for some words. If human annotators do not agree much more than chance on many words, it seems that systems that produce random sense tags for these words should be considered as satisfactory.

5. Conclusion and future works

In this paper, we proposed and evaluated a new possibilistic approach for word sense disambiguation. In fact, in spite of their advantages, the traditional dictionaries suffer from a lack of accurate information useful for WSD. Moreover, there exists a lack of semantically labeled corpora on which methods of learning could be trained. For these multiple reasons, it became important to use a semantic dictionary of contexts ensuring the machine learning in a semantic platform of WSD. Our approach combines traditional dictionaries and labeled corpora to build a semantic dictionary and identifies the sense of a word by using a possibilistic matching model.

To evaluate our approach, we used the ROMANSEVAL collection and we compared our results to some existing systems. Experiments showed an encouraging improvement in terms of disambiguation rates of French words. This disambiguation performed better on nouns as they are most frequent among the existing words in the context. These results reveal the contribution of possibilistic theory, as it provided good accuracy rates in this first experiment. However, our WSD approach needs to be investigated in a practical case of application.

Indeed, the long term goal of our work is to improve the performance of a cross-lingual information retrieval system by introducing a step of queries and documents disambiguation in a multilingual context. Thus, this work will be wide towards other languages such as English and Arabic. Moreover, our tools and data structures are reusable components that may be integrated in other fields such as information extraction, machine translation, content analysis, word processing, lexicography and the semantic Web applications.

References

[1] X. Zhou and H. Han, "Survey of word sense disambiguation approaches," in *Proceedings of the 18th International Florida AI Research Society Conference*, Clearwater Beach, Florida, USA, pp. 307-313, 2005.

[2] R. Navigli, "Word sense disambiguation: A survey," *ACM Computing Surveys (CSUR)*, vol. 41, no. 2, p. 10, 2009.

[3] D. Dubois and H. Prade, *Théorie des Possibilités : Application à la Représentation des Connaissances en Informatique*. Paris: MASSON, 1987.

[4] D. Dubois and H. Prade, *Possibility Theory: An Approach to computerized Processing*. New York, USA: Plenum Press, 2004.

[5] B. Elayeb, F. Evrard, M. Zaghdoud, and M. Ben Ahmed, "Towards An Intelligent Possibilistic Web Information Retrieval using Multiagent System," *International Journal of Interactive Technology and Smart Education*, vol. 6, no. 1, pp. 40-59, 2009.

[6] F. Segond, "Framework and Results for French," *Computers and the Humanities*, vol. 34, no. 1, pp. 49-60, 2000.

[7] A.J. Viera and J.M. Garrett, "Understanding interobserver agreement: the kappa statistic" *Family Medecine*, vol. 37, no. 5, pp. 360-363, 2005.

[8] J.R. Landis and G.G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159-174, 1977.

[9] J. Véronis, "A study of polysemy judgements and inter-annotator agreement," in *Programme and advanced papers of the Senseval workshop,* Sussex, England, pp. 2-4, 1998.

[10] L. A. Zadeh, "Fuzzy Sets as a basis for a theory of Possibility", *Fuzzy Sets and Systems*, vol. 1, no. 1, pp. 3-28, 1978.

[11] J. Cohen, "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit", *Psychological Bulletin*, vol. 70, no. 4, pp. 213-220, 1968.

[12] B. Di Eugenio, "On the usage of Kappa to evaluate agreement on coding tasks", In *Proceedings of LREC,* Athens, Greece, pp. 441-444, 2000.

[13] D. Yarowsky, "Decision lists for lexical ambiguity resolution: Application to accent restoration in Spanish and French". In *The 32nd Annual Meeting of the Association for*

Computational Linguistics, New Mexico, USA, pp. 88–95, 1994.

[14] N. Ide, and J. Véronis, "Word sense disambiguation: The state of the art", *Computational Linguistics: Special Issue on Word Sense Disambiguation*, vol. 24 , no. 1, pp. 1-40, 1998.

[15] L. Audibert, "Outils d'exploration de corpus et désambiguïsation lexicale automatique", Ph.D. Thesis, Université d'Aix-Marseille I – Université de Provence, 2003.

[16] J. Véronis, "Sense tagging : Does it makes sense", Corpus Linguistics, Lancaster, United Kingdom, p. 599, 2001.

[17] J. Véronis, "Les dictionnaires traditionnels sont-ils adaptés au traitement du sens en T.A.L. ? ", *Journée d'étude de l'ATALA, Les dictionnaires électroniques*, Paris, 2002.

[18] B. Gaume, N. Hathout and P. Muller, P. "Word sense disambiguation using a dictionary for sens similarity measure", In *The 20th International Conference on Computational Linguistics*, Stroudsburg, PA, USA, pp. 1194-1200, 2004.

[19] S. Benferhat, D. Dubois, L. Garcia, and H. Prade, "Possibilistic logic bases and possibilistic graphs", In t*he 15th Conference on Uncertainty in Artificial Intelligence*, Stockholm, Sweden, pp. 57–64, 1999.

[20] C. Borgelt, J. Gebhardt and R. Kruse, "Possibilistic Graphical Models" Computational Intelligence in Data Mining (Proceedings of the 3rd International Workshop, Udine, Italy 1998) CISM Courses and Lectures 408, Springer, Wien, Austria, pp. 51-68, 2000.

[21] A. Brini, M. Boughanem, and D. Dubois, "Towards a Possibilistic Approach for Information Retrieval", In *Data and Knowledge Engineering Proceedings EUROFUSE*, Warszawa, Poland, pp. 92-102, 2004.

[22] D. Reymond, " Méthodologie pour la création d'un dictionnaire distributionnel dans une perspective d'étiquetage lexical semi-automatique", In *6ème Rencontre des étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues*, vol. 1, pp. 405–414, 2002.

[23] R. Ayed, I. Bounhas, B. Elayeb, F. Evrard, and N. Bellamine Ben Saoud, "Arabic Morphological Analysis and Disambiguation Using a Possibilistic Classifier", In *The 8th International Conference on Intelligent Computing*, Huangshan, China, Springer-Verlag Berlin Heidelberg, LNAI 7390, pp. 274–279, 2012.

[24] R. Ayed, I. Bounhas, B. Elayeb, F. Evrard, and N. Bellamine Ben Saoud, "A Possibilistic Approach for the Automatic Morphological Disambiguation of Arabic Texts", In *The 13th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, August 08-10, 2012, Kyoto, Japan, IEEE Computer Society, 2012 (to appear).