A Discriminative Possibilistic Approach for Query Translation Disambiguation

Wiem Ben Romdhane[1], Bilel Elayeb[1,2(✉)], and Narjès Bellamine Ben Saoud[1]

[1] RIADI Research Laboratory, ENSI Manouba University, Manouba, Tunisia
br.wiem@yahoo.fr, Bilel.Elayeb@riadi.rnu.tn, Narjes.Bellamine@ensi.rnu.tn
[2] Emirates College of Technology, P.O. Box: 41009 Abu Dhabi, United Arab Emirates

Abstract. We propose, assess and compare in this paper a new discriminative possibilistic query translation (QT) disambiguation approach using both a bilingual dictionary and a parallel text corpus in order to overcome some drawbacks of the dictionary-based techniques. In this approach, the translation relevance of a given source query term is modeled by two measures: the possible relevance allows rejecting irrelevant translations, whereas the necessary relevance makes it possible to reinforce the translations not eliminated by the possibility. We experiment this new approach using the French-English parallel text corpus *Europarl* and the *CLEF-2003* French-English CLIR test collection. Our experiments highlighted the performance of our new discriminative possibilistic approach compared to both the probabilistic and the probability-to-possibility transformation-based approaches, especially for short queries and using different assessment metrics.

Keywords: Cross-Language information retrieval (CLIR) · Query translation disambiguation · Probabilistic model · Possibilistic model · Relevance

1 Introduction

Nowadays, the number of online non-English documents on the Internet is continuously increased, which requires the availability of high-performance cross-language information retrieval (CLIR) systems satisfying the Internet users' needs. Query translation (QT) techniques are the main research task in the domain of CLIR [12]. Besides, the usefulness of the simple dictionary-based translation approaches in QT has been improved due to the availability of machine readable bilingual dictionaries for several languages. However, these approaches are lacked by many major challenges such as: (i) the translation disambiguation problem known as the difficulty to choose the correct translation corresponding to each source query term among all the possible translations existing in the dictionary; and (ii) the poor coverage of the available dictionaries suffering from the missing of many translations corresponding to new terminologies. Nevertheless, many research works in the literature [20, 22, 25] are dedicated to manually or automatically collect larger lexical resources in order to

F. Frasincar et al. (Eds.): NLDB 2017, LNCS 10260, pp. 366–379, 2017.
DOI: 10.1007/978-3-319-59569-6_45

increase the coverage of these dictionaries. Moreover, translation ambiguity can decrease IR effectiveness. In order to overcome this limit, some approaches have used a phrase dictionary to firstly select noun phrases in the source query and secondly translate them as units.

We propose, assess and compare in this paper a new discriminative possibilistic QT approach dedicated to solve the problem of QT ambiguity using both a bilingual dictionary and a parallel text corpus in order to overcome some drawbacks of the dictionary-based QT techniques. In this approach, the relevance of a source query term translation is modeled by two measures: The possible relevance allows rejecting irrelevant translations, whereas the necessary relevance makes it possible to reinforce the translations not eliminated by the possibility. We compare this new approach to both the probabilistic and the probability-to-possibility transformation-based approaches [11], in which we start by identifying noun phrases (NPs) using the *Stanford Parser*[1] and translating them as units using translation patterns and a language model. Then, remaining source query terms are translated using a probabilistic word-by-word translation technique. Indeed, the suitable translation of each source query term or NP has a tendency to co-occur in the target language documents unlike unsuitable ones. Besides, additional words and their translations are automatically generated from a parallel bilingual corpus in order to increase the coverage of the bilingual dictionary. We assessed our approach using the French-English parallel text corpus *Europarl*[2] and the *CLEF-2003* French-English CLIR test collection. Our results confirmed the performance of our new discriminative possibilistic approach compared to both the probabilistic and the probability-to-possibility transformation-based approaches [11], principally for short queries and using various evaluation scenarios and metrics.

This paper is organized as follows. We present in Sect. 2 related works in the field of QT and discuss the solutions aiming at solving the problem of translation ambiguity. In Sect. 3, we recall the necessary of possibility theory. The new discriminative possibilistic QT approach is given in Sect. 4. Section 5 details our experimentations, expose and discuss a comparative study between QT approaches. Section 6 concludes our work in this paper and suggests some perspectives for future research.

2 Related Work

Since the early 1990's, many researchers (e.g. [19]) showed that the usefulness of a manually translating phrases have performed better results in CLIR effectiveness than a word-by-word dictionary-based translation. Moreover, CLIR performance was increased by [6] when they used their phrase dictionary generated from a set of parallel sentences in French and English. In the same way, Ballesteros and Croft [2] confirmed that translations of multi-word concepts as phrases are more perfect than word-by-word translations. Indeed, phrases translations were achieved using information existing in phrase and word usage available in the Collins machine readable dictionary.

[1] http://nlp.stanford.edu/software/lex-parser.shtml.

[2] http://www.statmt.org/europarl/.

Unfortunately, it is hard to find or build an exhaustive phrase dictionary in CLIR since we have until now many missing phrases in these available lexicons. Therefore, it is not easy to select all missing phrases in the queries and suitably translate them. Actually, many unfamiliar phrases suffer from the problem of their identification and translation since they cannot be identified in any dictionaries. Hence, the problem of the lexicon coverage is one of the limits of this approach since we cannot know until now: How can one build an exhaustive phrase dictionary?

On the other hand, many authors have tackled the problem of translation ambiguity using word sense disambiguation (WSD) techniques. For example, Hull [18] used structured queries for disambiguation in QT task. Besides, co-occurrence statistics from corpora are used by Ballesteros and Croft [13] in order to decrease translation ambiguity in CLIR. Then, many disambiguation strategies are suggested and evaluated by Hiemstra and Jong [61] in CLIR tasks. Later, a technique-based statistical term similarity for word sense disambiguation was suggested and tested by [1] in order to enhance CLIR effectiveness. Even if the difficulty of translation ambiguity was significantly decreased, Xu and Weischedel [24] showed that it is not evident to accomplish perfect enhancement in CLIR performance. Recently, Lefever and Hoste [21] have argued the advantage of moving from traditional monolingual WSD task into a cross-lingual one. In fact, this new technique, namely "Cross-Lingual WSD" (CLWSD) has been involved in CLIR since SemEval-2010 and SemEval-2013 competitions. The participant in these exercises confirmed that CLIR effectiveness has been enhanced due to the new CLWSD technique which significantly resolves the problem of translation ambiguity.

QT techniques require training and matching models in order to compute a score of relevance (or similarities) between source query terms/phrases and their possible translations. Existing QT approaches in the literature are based on poor, uncertain and imprecise data, while possibility theory is naturally dedicated to this type of applications, since it takes into account of the imprecision and uncertainty at the same time and it makes it possible to express ignorance. However, the main challenge of our approach is that the context used in the translation disambiguation process of a given source query term can be also ambiguous. Consequently, we consider this phenomenon as a case of imprecision. That's why we are inspired from the possibility theory which naturally applies to this kind of imperfection. By cons, the probability theory is not suitable to deal with such type of data. Besides, and given that the possibility theory is the best framework suitable for imprecision treatment, we have taken advantage of possibility distributions in order to solve the problem of translation ambiguity in CLIR task. Recently, we have proposed and tested in [11] a possibilistic QT approach derived from the probabilistic one using a probability-to-possibility transformation as a mean to introduce further tolerance in QT process. This approach has achieved a statistically significant improvement compared to the probabilistic one using both long and short queries.

3 Possibility Theory

We briefly present in the following sections the basic elements of possibility theory such as the possibility distribution (cf. Sect. 3.1), the possibility and necessity measures (cf. Sect. 3.2) and the possibilistic networks (cf. Sect. 3.3). More details and discussions about possibility theory are available in [7–10].

3.1 Possibility Distribution

The fundamental element of the possibility theory is the *possibility distribution*. Given the universe of discourse $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, we symbolised by π the basic concept corresponding to a function which associates to each element $\omega_i \in \Omega$ a value from a bounded and linearly ordered valuation set (L, <). Moreover, the *possibility degree* is defined as the value in which our knowledge on the real world is encoded. Indeed, this scale has two interpretations: (i) when the handled values reflect only an ordering between the different states of the world, it is the *qualitative* setting which can be applied using the *min* operator; and (ii) when the handled values have a real sense, it is the *quantitative* setting which can be applied using the *product* operator. Flexibility is modelled by allowing providing a possibility degree from the interval [0, 1]. We note that $\pi(\omega_i) = 1$ means that it is fully possible that ω_i is the real world, and $\pi(\omega_i) = 0$ means that it is impossible that ω_i is the real world. The possibility theory provides different significant exacting cases of knowledge as the following: (i) *Complete knowledge* ($\exists\ \omega_i \in \Omega$, $\pi(\omega_i) = 1$ *and* $\forall \omega_j \neq \omega_i$, $\pi(\omega_j) = 0$); (ii) *Partial ignorance* ($\forall\ \omega_i \in A \subseteq \Omega$, $\pi(\omega_i) = 1$, $\forall \omega_i \notin A$, $\pi(\omega_i) = 0$; when A is not a singleton); and (iii) *Total ignorance:* all values in Ω are possible ($\forall\ \omega_i \in \Omega$, $\pi(\omega_i) = 1$).

3.2 Possibility and Necessity Measures

The two dual measures in which a possibility distribution π on Ω enables events to be qualified in terms of their *plausibility* and their *certainty* are respectively known as the *Possibility* (Π) and the *Necessity* (N) [7]. Given a possibility distribution π on the universe of discourse Ω, the corresponding possibility and necessity measures of any event $A \subseteq 2^{\Omega}$ are respectively determined by the Eqs. (1) and (2):

$$\prod(A) = \max_{w \in A} \pi(w) \tag{1}$$

$$N(A) = \min_{w \notin A}(1 - \pi(w)) = 1 - \prod(\overline{A}) \tag{2}$$

In fact, $\Pi(A)$ provides an assessment similar to a degree of *non-emptiness* of the intersection of the fuzzy set having π as membership function with the classical subset A. Thus, $\Pi(A)$ measures at which level A is *consistent* with our knowledge represented by π. While, $N(A)$ assesses at which level A is *certainly* inferred by our knowledge represented by π; since it is a degree of inclusion of the fuzzy set corresponding to π into the subset A.

3.3 Possibilistic Networks (PN)

The numerical and graphical components are the main characteristic of a directed possibilistic network on a variable set V. These two components are defined as follow: (i) the distinct links in the graph are quantified via the *numerical component*. Indeed, it represents conditional possibility matrix of every node given the context of its parents. (ii) The *graphical component* is a directed acyclic graph (DAG). The DAG enables representing conditional dependency between dependent or independent variables. Every link denotes a dependency between two variables and every node in the graph denotes a domain variable. The graph structure encodes independence relation sets between nodes. Moreover, these possibility distributions should respect the normalization feature [4]. For each variable V:

- If V is a root node and *Dom(V)* the domain of V, the prior possibility of V should satisfy:

$$max_{v \in Dom(V)} \Pi(v) = 1 \tag{3}$$

- If V is not a root node, the conditional distribution of V in the context of its parents denoted U_V should satisfy:

$$max_{v \in Dom(V)} \Pi(v|u_V) = 1;\; u_V \in Dom(U_V) \tag{4}$$

Where: $Dom(V)$: domain of V; U_V: value of parents of V; $Dom(U_V)$: domain of parent set of V.

We suggest in this paper a new possibilistic approach for QT disambiguation based on possibilistic network (cf. Sect. 4). We link in this network the possible translations (T_i) to the terms of a source query $SQ = (t_1, t_2, \ldots, t_p)$, which represents its context. In this case: $v_i = t_i$; $u_V = T_i$; $Dom(V) = \{t_1, t_2, \ldots, t_p\}$; and $Dom(U_V) = \{T_1, T_2, \ldots, T_N\}$.

We provide in Sect. 4.2 an illustrative example including detailed calculus. The possibilistic graph which associated conditional possibility distribution is based on the product operator. The product-based possibilistic graph (*PPG*), is generally comfortable for the numerical setting where possibility measures represent numerical values in [0, 1]. The possibility distribution of product-based possibilistic networks (π_p) obtained by the associated chain class is calculated via Eq. (5):

$$\pi_p(V_1, V_2, \cdots, V_N) = \prod_{i=1}^{N} \Pi(V_i|U_{V_i}) \tag{5}$$

4 The Discriminative Possibilistic QT Approach

We present in Sect. 4.1 the formulae for calculating the Degree of Possibilistic Relevance (DPR) and an illustrative example with a detailed calculus in Sect. 4.2.

4.1 The Degree of Possibilistic Relevance (DPR)

Let us consider the source query (*SQ*) enclosing *P* terms and denoted as: $SQ = (t_1, t_2, ..., t_P)$. We assume that *SQ* includes only one ambiguous term having several possible translations. We note the Degree of Possibilistic Relevance of a translation T_j given *SQ* by $DPR(T_j|SQ)$. We have taken advantage of a possibilistic matching model of information retrieval (IR) used in [5, 12–16] in order to assess the relevance of a translation T_j given a source query *SQ*. In the case of IR, the matching score is calculated between a query and a document. However, in case of QT disambiguation, we model the relevance of a translation given a source query using double measures: The *possible* relevance and the *necessary* relevance. The irrelevant translations are rejected due to the possible relevance, while the relevance of the remaining translations, which have not been rejected by the possibility, is reinforced due to the necessary relevance.

Figure 1 presents our possibilistic network which links the word translation T_j to the terms of a given source query $SQ = (t_1, t_2, ..., t_P)$. The output of the QT disambiguation process is the target query $TQ = (T_1, T_2, ..., T_P)$. The later will be useful to retrieve a set of relevant documents on the target language.

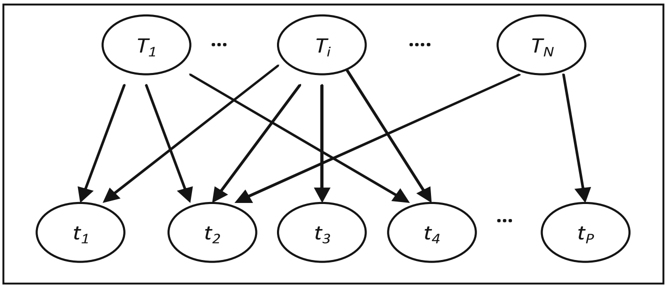

Fig. 1. Possibilistic network of the QT disambiguation approach

Giving the source query *SQ*, the relevance of every word translation T_j is computed as the following:

Analogically to the IR matching model proposed in [46–49], the possibility $\Pi(T_j|SQ)$ is proportional to:

$$\Pi'(T_j|SQ) = \Pi(t_1|T_j) * \ldots * \Pi(t_P|T_j) = nft_{1j} * \ldots * nft_{Pj} \tag{6}$$

Where:

- $nft_{ij} = tf_{ij}/max(tf_{kj})$: the normalized frequency of the source term t_i in the parallel text of the translation T_j.
- tf_{ij} is the number of occurrence of the source term t_i in the parallel text of the translation T_j divided by the number of terms in the parallel text of the translation T_j.

We compute the necessity to restore a relevant translation T_j given the source query SQ, denoted $N(T_j|SQ)$, as the following:

$$\mathrm{N}\left(T_j|SQ\right) = 1 - \Pi\left(\neg T_j|SQ\right) \tag{7}$$

Where:

$$\Pi\left(\neg T_j|SQ\right) = \left(\Pi\left(SQ|\neg T_j\right) * \Pi\left(\neg T_j\right)\right)/\Pi(SQ) \tag{8}$$

At the same way $\Pi(\neg T_j| SQ)$ is proportional to:

$$\Pi'\left(\neg T_j|SQ\right) = \Pi\left(t_1|\neg T_j\right) * \ldots * \Pi\left(t_P|\neg T_j\right) \tag{9}$$

This numerator can be expressed as the following:

$$\Pi'\left(\neg T_j|SQ\right) = \left(1 - \phi T_{1j}\right) * \ldots * \left(1 - \phi T_{Pj}\right) \tag{10}$$

Where:

$$\phi T_{ij} = Log_{10}\left(nCT/nT_j\right) * \left(nft_{ij}\right) \tag{11}$$

Where: nCT is the number of possible translations in the bilingual dictionary. But, nT_j is the number of parallel texts of the translation T_j containing the source term t_i. This includes all possible translations existing in the bilingual dictionary.

We compute the Degree of Possibilistic Relevance (DPR) of each word translation T_j giving a source query SQ via the following Eq. (12):

$$DPR\left(T_j|SQ\right) = \Pi\left(T_j|SQ\right) + \mathrm{N}\left(T_j|SQ\right) \tag{12}$$

Finally, the suitable translations are those which have a high score of $DPR(T_j|SQ)$.

4.2 Illustrative Example

We provide here a numerical calculation example for reasons of brevity. But, we have already detailed in [11] some data/corpus-based examples in which we have showed the difference between the probabilistic and the possibilistic QT approaches.

Let us consider the source query $SQ = (W, t_2, t_4, t_5, t_7)$, which contains only one polysemous source query term W in order to simplify the calculus in this example. We assume that W has two possible translations T_1 and T_2 in the bilingual dictionary. We suppose also that the parallel text of T_1 is indexed by the three terms $\{t_1, t_2, t_3, t_4\}$ and the parallel text of T_2 is indexed by $\{t_1, t_4, t_5, t_6, t_7\}$. We have:

$\Pi(T_1|SQ) = nf_{(W,\ T1)} * nf_{(t2,\ T1)} * nf_{(t4,\ T1)} * nf_{(t5,\ T1)} * nf_{(t7,\ T1)} = 0*(1/4)*(1/4)*0*0 = 0$. Where: $nf_{(W,\ T1)}$ is the normalized frequency of W in the parallel text of the first translation T_1.

$\Pi(T_2|\ SQ) = nf_{(W,\ T2)} * nf_{(t2,\ T2)} * nf_{(t4,\ T2)} * nf_{(t5,\ T2)} * nf_{(t7,\ T2)} = 0*0*(1/5)*(1/5)*(1/5) = 0$. We have frequently $\Pi(T_j|\ SQ) = 0$; except if all the words of the source query exist in the index of the parallel text of the translation.

On the other hand, we have not null values of $N(T_j|\ SQ)$:

$N(T_1|\ SQ) = 1-((1-\phi(T_1, W)) * (1-\phi(T_1, t_2)) * (1-\phi(T_1, t_4)) * (1-\phi(T_1, t_5) * (1-\phi(T_1, t_7)))$; $nf_{(T1,\ W)} = 0$, so $\phi(T_1, W) = 0$; $\phi(T_1, t_2) = log_{10}(2/1)*(1/4) = 0{,}075$; $\phi(T_1, t_4) = log_{10}(2/2)*(1/4) = 0$; $\phi(T_1, t_5) = 0$; $\phi(T_1, t_7) = 0$. So: $N(T_1|\ SQ) = 1-((1-0) * (1-0.075) * (1-0) * (1-0) * (1-0)) = 1-(1* 0.925* 1* 1*1) = 0.075$. Thus, $DPR(T_1|\ SQ) = 0.075$. $N(T_2|SQ) = 1-((1-\phi(T_2, W))* (1-\phi(T_2, t_2))* (1-\phi(T_2, t_4)) * (1-\phi(T_2, t_5) * (1-\phi(T_2, t_7)))$. Where: $\phi(T_2, W) = 0$ because $nf(T_2, W) = 0$; $\phi(T_2, t_2) = 0$; $\phi(T_2, t_4) = log_{10}(2/2)*(1/5) = 0$; $\phi(T_2, t_5) = log_{10}(2/1)*(1/5) = 0.06$; $\phi(T_2, t_7) = log_{10}(2/1)*(1/5) = 0.06$. So: $N(T_2|SQ) = 1-((1-0)* (1-0)* (1-0)*(1-0.06)*(1-0.06)) = 1-(1* 1*1*0.94* 0.94) = 1-0.8836 = 0.1164$. Thus, $DPR(T_2|SQ) = 0.1164 > DPR(T_1|SQ) = 0.075$.

We notice that the source query SQ is more relevant for T_2 than T_1; because it encloses three terms (t_4, t_5, t_7) of the index of the parallel text of T_2 and only two terms (t_2, t_4) of the index of the parallel text of T_1.

5 Experiments and Discussion

We assess, compare and discuss in this section the discriminative possibilistic approach for QT disambiguation. Indeed, we suggest various evaluation scenarios and metrics using the CLEF-2003 standard CLIR test collection. We compare the performance and the efficiency of the discriminative approach to both the most known efficient probabilistic and the probability-to-possibility transformation-based ones [11]. Moreover, our evaluation is performed according to the TREC protocol. We used the IR matching model OKAPI-BM25 existing in *Terrier*[3] platform in order to retrieve English relevant document. We are focused on performance metrics such as *Recall* and *Precision*, mostly used in the evaluation of CLIR tools. The evaluation assumes that there is an ideal set the system is supposed to search. This ideal set is useful to define these two metrics as follows. The recall is the percentage of documents in the ideal set that were retrieved by the system, while the precision is the percentage of documents retrieved by the system, which are also in the ideal set. Moreover, we used the precision (Y-axis) over 11 points of recall in the X-axis (0.0, 0.1,…, 1.0) to draw all recall-precision curves.

Besides, we evaluated these approaches using both the Mean Average Precision (*MAP*) and the exact precision (*R-Precision*). The latter is defined as the precision at rank R; where R is the total number of relevant documents, while the MAP is the mean of the average precision scores for each query. The Equations of calculating the *MAP* and the *R-Precision* are given in [11]. We compute also the improvement percentage between two variations of the model variables. This percentage is obtained in generally for two variables A and B measuring the percentage of C as $\%C = [(B-A)/A]*100$. Our 54 test queries enclose 717 French words having 2324 possible English

[3] http://terrier.org/.

translations in the bilingual dictionary. Indeed, we firstly generate our bilingual dictionary from the *Europarl* parallel corpus using all French words with their possible translations existing in this corpus in order to enlarge our lexicon coverage. Then, we have benefited from the online intelligent speller and grammar checker *Reverso*[4] in order to check this dictionary. Finally, the free online Google translate[5] has been used to enrich and check this bilingual dictionary.

Firstly, we provide in Fig. 2 the Recall-Precision curves comparing monolingual (English queries provided by CLEF-2003), discriminative, probabilistic and probability-to-possibility transformation-based (possibilistic) runs using the *title* part of the source queries as input data to our CLIR tool described in [11]. Secondly, we provide the precision values at different top documents P@5, P@10,…, P@1000, the *MAP* and the *R-Precision*. For example, the precision in point 5, namely P@5, is the ratio of relevant documents among the top 5 returned documents. The goal of the following experiments and discussion is to show and assess our contributions compared to these competitors QT disambiguation approaches mainly investigated and tested in [11].

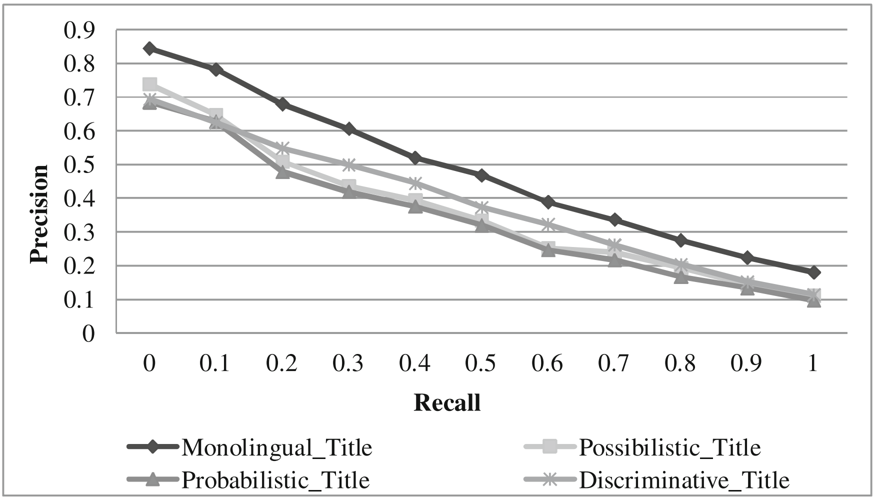

Fig. 2. Recall-Precision curves comparing monolingual, discriminative, probabilistic and probability-to-possibility transformation-based (possibilistic) runs

If we focus only on the *title* part of the query, the context is limited to a few numbers of terms in which the identification of the NP is not frequent in this case. Therefore, the discriminative possibilistic approach significantly outperforms both the probabilistic and the probability-to-possibility transformation-based (possibilistic)

[4] http://www.reverso.net/spell-checker/english-spelling-grammar/.

[5] https://translate.google.fr/?hl=fr.

approaches; except in some low-levels points of recall (0 and 0.1). Unfortunately, the monolingual run is still outperformed these approaches in all points of recall.

On the other hand, and in order to further confirm our conclusions made above, we provide a comparative study between the monolingual, the discriminative, the possibilistic and the probabilistic runs using the precision at different top documents, the *MAP* and the *R-Precision* metrics (cf. Fig. 3). It is trivial that the precision decreases when the number of returned documents increases. Moreover, short queries using only *title* are more efficient for the discriminative approach for all top returned documents; except for some rare cases such as P@100 and P@1000 where the number of returned documents is important, which increases the noise in the retrieved results. Unfortunately, the monolingual run is normally upper bound of CLIR performance in all our experiments, because we don't involve in our tests any query expansion step before and/or after the translation process and no close phrases/words have enriched the source and/or the target queries before returning search results.

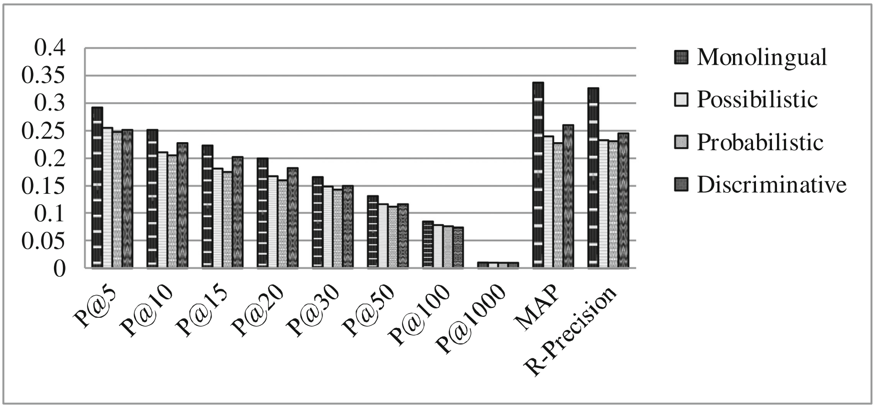

Fig. 3. Results using the precision at different top documents, MAP and R-Precision

We remark that the discriminative approach outperformed both of them in terms of the *MAP* and the *R-Precision*. In fact, these two metrics confirm again that short query using *title* is still suitable for the discriminative approach compared to its two competitors QT techniques. We present in Table 1 the improvement percentage of the discriminative approach compared to both the probabilistic and the probability-to-possibility transformation-based approaches using the precision at different top documents, the *MAP* and the *R-Precision*. Firstly, and compared to the probabilistic, the discriminative approach has performed a significant improvement in terms of precision of documents returned to the top of list using the *title* of the source query. For example, we have registered an improvement percentage more than 15% for P@15 and almost 14% for P@20, while the average improvement for all top documents is about 6%. Besides, if we focus on the *MAP* metric, the improvement of the MAP is about 14.4% and about 6% for the R-Precision metric. Secondly, and compared to the probability-to-possibility transformation-based approach, the discriminative achieved

Table 1. The improvement percentage of the discriminative compared to both the probabilistic and the probability-to-possibility transformation-based approaches

Precision metrics	% improvement Discriminative *vs.* Probabilistic	% improvement Discriminative *vs.* Possibilistic
P@5	1.53	−1.45
P@10	10.8	7.91
P@15	15.52	11.57
P@20	13.86	8.83
P@30	4.75	0.81
P@50	4.01	0
P@100	−2.88	−5.59
P@1000	0	−2.91
MAP	14.4	8.54
R-Precision	6.14	5.27

an improvement percentage more than 7.9% for P@10 and more than 11.5% for P@15, whereas the average improvement for all top documents is about 2.4%. Moreover, the MAP is about 8.5% and the R-Precision is about 5.3%. These results confirm our deductions concluded above about the efficiency of the discriminative approach in case of short queries. It is the case when the user provided to the CLIR tool a few number of terms in his/her source query due to his/her lack of language or his/her limit knowledge about the retrieved domain.

On the other hand, we need to more investigate on the statistical significance of the improvement achieved by the discriminative possibilistic approach compared to its competitors in terms of precision at different top documents, the *MAP* and the R-*Precision* scores using short queries. To do this, we use the Wilcoxon Matched-Pairs Signed-Ranks Test as suggested by [17]. This statistical test is a non-parametric alternative to the paired *t*-test that enables us to decide whether the improvement by method 1 over method 2 is significant. Indeed, the *t*-test computes a *p*-value based on the performance data of both methods 1 and 2. The improvement is more significant for the smaller *p*-value. Generally, the improvement is statistically significant if the *p*-value is small enough (p-value < 0.05).

The improvement of the discriminative possibilistic approach compared to the probabilistic one is statistically significant (p-value $= 0.010793 < 0.05$), while it is not statistically significant (p-value $= 0.085831 > 0.05$) compared to the probability-to-possibility transformation-based approach. Globally, these tests confirms again the performance of our discriminative possibilistic approach in the disambiguation of short queries using *title* compared to both the known efficient probabilistic and to the probability-to-possibility transformation-based approaches using different assessment metrics. Finally, and in order to provide an objective evaluation of our approach, we are limited in our empirical comparative study to these two approaches (probabilistic and possibilistic); because the other state-of-the-art QT techniques detailed in Sect. 2 are assessed using both different linguistic resources (dictionary and parallel corpora) and different CLIR test collection for different pair of languages.

6 Conclusion and Future Work

The translations' ambiguities in a QT process are considered as cases of imprecision since many possible translations are available for each ambiguous source query term. The disambiguation process consists of using the context of the source query which can be also ambiguous. Consequently, we propose in this paper a new discriminative possibilistic approach for QT disambiguation dedicated to improve the dictionary-based QT ones. This new technique has taken advantage of both a parallel text corpus and a bilingual dictionary in order to overcome some weaknesses of the-state-of-the-art QT approaches. Indeed, we have modelled the relevance of possible translations within two measures: the possible relevance allows eliminating irrelevant translations, whereas the necessary relevance makes it possible to reinforce the translations not rejected by the possibility. We assessed and compared the discriminative possibilistic approach using the French-English parallel text corpus *Europarl* and the CLEF-2003 French-English CLIR test collection. Our experiments highlighted the performance of our new discriminative possibilistic approach compared to both the known efficient probabilistic and the probability-to-possibility transformation-based approaches [11] using different assessment metrics. Indeed, the improvement of the discriminative approach compared to the probabilistic one is statistically significant in terms of precision at different top documents, the MAP and the R-Precision.

In spite of its significant efficiency in translating short queries, the discriminative possibilistic approach suffers from some weaknesses in case of specific domain queries. Consequently, the identification of the suitable translations requires a domain-specific translation [23] and a language model. Nevertheless, combining a language model with a domain-specific translation and their integration in the discriminative approach are not easy tasks. Moreover, the evaluation processes of our approach should be done in real contexts by allowing the users to contribute in its assessment. It is also relevant to assess the impact of QT disambiguation on CLIR efficiency before and/or after query expansion process using our recent techniques in [3, 13, 15].

References

1. Adriani, M.: Using statistical term similarity for sense disambiguation in cross-language information retrieval. Inf. Retr. **2**(1), 67–78 (2000)
2. Ballesteros, L., Croft, W.B.: Resolving ambiguity for cross-language retrieval. In: Proceedings of ACM-SIGIR, pp. 64–71 (1998)
3. Ben Khiroun, O., Elayeb, B., Bounhas, I., Evrard, F., Bellamine Ben Saoud, N.: Improving query expansion by automatic query disambiguation in intelligent information retrieval. In: Proceedings of ICAART 2014, pp. 153–160 (2014)
4. Benferhat, S., Dubois, D., Garcia, L., Prade, H.: Possibilistic logic bases and possibilistic graphs. In: Proceedings of the UAI 1999, pp. 57–64 (1999)
5. Boughanem, M., Brini, A., Dubois, D.: Possibilistic networks for information retrieval. Int. J. Approx. Reason. **50**(7), 957–968 (2009)

6. Davis, M.W., Ogden, W.C.: QUILT: implementing a large-scale cross-language text retrieval system. In: Proceedings of ACM-SIGIR, pp. 92–98 (1997)
7. Dubois, D., Prade, H.: Possibility theory: qualitative and quantitative aspects. In: Gabbay, D. M., Smets, P. (eds.) Quantified Representation of Uncertainty and Imprecision, Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1, pp. 169–226. Klower Academic Publishers, Netherlands (1998)
8. Dubois, D., Prade, H.: An overview of ordinal and numerical approaches to causal diagnostic problem solving, abductive reasoning and learning. In: Gabbay, D.M., Kruse, R. (eds.) Handbooks of Defeasible Reasoning and Uncertainty Management Systems, Drums Handbooks, vol. 4, pp. 231–280 (2000)
9. Dubois, D., Prade, H.: Représentations formelles de l'incertain et de l'imprécis, Concepts et méthodes pour l'aide à la décision - outils de modélisation. In: Bouyssou, D., Dubois, D., Pirlot, M., Prade, H. (eds.) vol. 2, pp. 99–137 (2006)
10. Dubois, D., Prade, H.: Formal representations of uncertainty. In: Bouyssou, D., Dubois, D., Pirlot, M., Prade, H. (eds.) Decision-Making Process, ISTE, pp. 85–156. Wiley, London and USA (2009)
11. Elayeb, B., Ben Romdhane, W., Bellamine Ben Saoud, N.: Towards a new possibilistic query translation tool for cross-language information retrieval. Multimedia Tools Appl. (2017). doi:10.1007/s11042-017-4398-2
12. Elayeb, B., Bounhas, I.: Arabic cross-language information retrieval: a review. ACM Trans. Asian Low-Resour. Lang. Inf. Process. **15**(3), Article 18, 44 (2016)
13. Elayeb, B., Bounhas, I., Khiroun, O.B., Bellamine Ben Saoud, N.: Combining semantic query disambiguation and expansion to improve intelligent information retrieval. In: Duval, B., Herik, J., Loiseau, S., Filipe, J. (eds.) ICAART 2014. LNCS, vol. 8946, pp. 280–295. Springer, Cham (2015). doi:10.1007/978-3-319-25210-0_17
14. Elayeb, B., Bounhas, I., Ben Khiroun, O., Evrard, F., Bellamine Ben Saoud, N.: Towards a possibilistic information retrieval system using semantic query expansion. Int. J. Intell. Inf. Technol. **7**(4), 1–25 (2011)
15. Elayeb, B., Bounhas, I., Ben Khiroun, O., Evrard, F., Bellamine Ben Saoud, N.: A comparative study between possibilistic and probabilistic approaches for monolingual word sense disambiguation. Knowl. Inf. Syst. **44**(1), 91–126 (2015)
16. Elayeb, B., Evrard, F., Zaghdoud, M., Ben Ahmed, M.: Towards an intelligent possibilistic web information retrieval using multiagent system. Interact. Technol. Smart Educ. Spec. Iss.: New Learn. Support Syst. **6**(1), 40–59 (2009)
17. Hull, D.A.: Using statistical testing in the evaluation of retrieval experiments. In: Proceedings of ACM-SIGIR, pp. 329–338 (1993)
18. Hull, D.A.: Using structured queries for disambiguation in cross-language information retrieval. In: Hull, D., Oard, D. (eds.) Proceedings of the AAAI Symposium on Cross-Language Text and Speech Retrieval, California, USA, pp. 84–98 (1997)
19. Hull, D.A., Grefenstette, G.: Querying across languages: a dictionary-based approach to multilingual information retrieval. In: Proceedings of ACM-SIGIR, pp. 46–57 (1996)
20. Kowk, K.L.: Exploiting a Chinese-English bilingual wordlist for English-Chinese cross-language information retrieval. In: Proceedings of the 5th International Workshop on Information Retrieval with Asian Languages, Hong Kong, pp. 173–179. ACM Press (2000)
21. Lefever, E., Hoste, V.: SemEval-2013 Task 10: cross-lingual word sense disambiguation. In: Proceedings of SemEval 2013, Atlanta, Georgia, pp. 158–166 (2013)
22. Nie, J.Y., Simard, M., Isabelle, P., Durand, R.: Cross-language information retrieval based on parallel texts and automatic mining of parallel texts from the web. In: Proceedings of ACM-SIGIR, pp. 74–81 (1999)

23. Ture, F., Boschee, E.: Learning to translate: a query-specific combination approach for cross-lingual information retrieval. In: Proceedings of EMNLP, pp. 589–599 (2014)
24. Xu, J., Weischedel, R.: Cross-lingual information retrieval using hidden markov models. In: Proceedings of the 2000 Joint SIGDAT Conference, pp. 95–103 (2000)
25. Xu, J., Fraser, A., Weischedel, R.: TREC 2001: cross-lingual retrieval at BBN. Proceedings of the TREC **2001**, 68–77 (2002)