Multilingual Pseudo-Relevance Feedback: Performance Study of Assisting Languages

Manoj K. Chinnakotla Karthik Raman Pushpak Bhattacharyya
Department of Computer Science and Engineering
Indian Institute of Technology, Bombay,
Mumbai, India
{manoj,karthikr,pb}@cse.iitb.ac.in

Abstract

In a previous work of ours Chinnakotla et al. (2010) we introduced a novel framework for Pseudo-Relevance Feedback (PRF) called *MultiPRF*. Given a query in one language called *Source*, we used English as the *Assisting Language* to improve the performance of PRF for the source language. MulitiPRF showed remarkable improvement over plain Model Based Feedback (MBF) uniformly for 4 languages, *viz., French, German, Hungarian* and *Finnish* with *English* as the assisting language. This fact inspired us to study the effect of *any source-assistant* pair on MultiPRF performance from out of a set of languages with widely different characteristics, *viz., Dutch, English, Finnish, French, German* and *Spanish*. Carrying this further, we looked into the effect of using *two assisting languages together* on PRF.

The present paper is a report of these investigations, their results and conclusions drawn therefrom. While performance improvement on MultiPRF is observed whatever the assisting language and whatever the source, observations are mixed when two assisting languages are used simultaneously. Interestingly, the performance improvement is more pronounced when the source and assisting languages are *closely related*, e.g., French and Spanish.

1 Introduction

The central problem of Information Retrieval (IR) is to satisfy the user's information need, which is typically expressed through a short (typically 2-3 words) and often ambiguous query. The problem of matching the user's query to the documents is rendered difficult by natural language phenomena like *morphological variations*, *polysemy* and *synonymy*. Relevance Feedback (RF) tries to overcome these problems by eliciting user feedback on the relevance of documents obtained from the initial ranking and then uses it to automatically refine the query. Since user input is hard to obtain, Pseudo-Relevance Feedback (PRF) (Buckley et al., 1994; Xu and Croft, 2000; Mitra et al., 1998) is used as an alternative, wherein RF is performed by *assuming* the top k documents from the initial retrieval as being *relevant* to the query. Based on the above assumption, the terms in the feedback document set are analyzed to choose the most distinguishing set of terms that characterize the feedback documents and as a result the relevance of a document. Query refinement is done by adding the terms obtained through PRF, along with their weights, to the actual query.

Although PRF has been shown to improve retrieval, it suffers from the following drawbacks: (a) the type of term associations obtained for query expansion is restricted to co-occurrence based relationships in the feedback documents, and thus other types of term associations such as lexical and semantic relations (morphological variants, synonyms) are not explicitly captured, and (b) due to the inherent assumption in PRF, *i.e.*, relevance of top k documents, performance is sensitive to that of the initial retrieval algorithm and as a result is not robust.

Multilingual Pseudo-Relevance Feedback (MultiPRF) (Chinnakotla et al., 2010) is a novel framework for PRF to overcome both the above limitations of PRF. It does so by taking the help of a different language called the *assisting language*. In MultiPRF, given a query in source language L_1, the query is automatically translated into the assisting language L_2 and PRF performed in the assisting language. The resultant terms are translated back into L_1 using a probabilistic bi-lingual dictionary. The translated feedback

model, is then combined with the original feedback model of L_1 to obtain the final model which is used to re-rank the corpus. MulitiPRF showed remarkable improvement on standard CLEF collections over plain Model Based Feedback (MBF) uniformly for 4 languages, *viz., French, German, Hungarian* and *Finnish* with *English* as the assisting language. This fact inspired us to study the effect of *any source-assistant* pair on PRF performance from out of a set of languages with widely different characteristics, *viz., Dutch, English, Finnish, French, German* and *Spanish*. Carrying this further, we looked into the effect of using *two assisting languages together* on PRF.

The present paper is a report of these investigations, their results and conclusions drawn therefrom. While performance improvement on PRF is observed whatever the assisting language and whatever the source, observations are mixed when two assisting languages are used simultaneously. Interestingly, the performance improvement is more pronounced when the source and assisting languages are *closely related*, e.g., French and Spanish.

The paper is organized as follows: Section 2, discusses the related work. Section 3, explains the Language Modeling (LM) based PRF approach. Section 4, describes the MultiPRF approach. Section 5 discusses the experimental set up. Section 6 presents the results, and studies the effect of varying the assisting language and incorporates multiple assisting languages. Finally, Section 7 concludes the paper by summarizing and outlining future work.

2 Related Work

PRF has been successfully applied in various IR frameworks like vector space models, probabilistic IR and language modeling (Buckley et al., 1994; Jones et al., 2000; Lavrenko and Croft, 2001; Zhai and Lafferty, 2001). Several approaches have been proposed to improve the performance and robustness of PRF. Some of the representative techniques are (i) Refining the feedback document set (Mitra et al., 1998; Sakai et al., 2005), (ii) Refining the terms obtained through PRF by selecting good expansion terms (Cao et al., 2008) and (iii) Using selective query expansion (Amati et al., 2004; Cronen-Townsend et al., 2004) and (iv) Varying the importance of documents in the feedback set (Tao and Zhai, 2006). Another direction of work, often reported in the TREC Robust Track, is to use a large external collection like Wikipedia or the Web as a source of expansion terms (Xu et al., 2009; Voorhees, 2006). The intuition behind the above approach is that if the query does not have many relevant documents in the collection then any improvements in the modeling of PRF is bound to perform poorly due to query drift.

Several approaches have been proposed for including different types of lexically and semantically related terms during query expansion. Voorhees (1994) use Wordnet for query expansion and report negative results. Recently, random walk models (Lafferty and Zhai, 2001; Collins-Thompson and Callan, 2005) have been used to learn a rich set of term level associations by combining evidence from various kinds of information sources like WordNet, Web *etc*. Metzler and Croft (2007) propose a feature based approach called *latent concept expansion* to model term dependencies.

All the above mentioned approaches use the resources available *within* the language to improve the performance of PRF. However, we make use of a *second language* to improve the performance of PRF. Our proposed approach is especially attractive in the case of resource-constrained languages where the original retrieval is bad due to poor coverage of the collection and/or inherent complexity of query processing (for example *term conflation*) in those languages.

Jourlin et al. (1999) use parallel blind relevance feedback, *i.e.* they use blind relevance feedback on a larger, more reliable parallel corpus, to improve retrieval performance on imperfect transcriptions of speech. Another related idea is by Xu et al. (2002), where a statistical thesaurus is learned using the probabilistic bilingual dictionaries of Arabic to English and English to Arabic. Meij et al. (2009) tries to expand a query in a different language using language models for domain-specific retrieval, but in a very different setting. Since our method uses a corpus in the assisting language from a similar time period, it can be likened to the work by Talvensaari et al. (2007) who used comparable corpora for Cross-Lingual Information Retrieval (CLIR). Other work pertaining to document alignment in comparable corpora, such as Braschler and Schäuble (1998), Lavrenko et al. (2002), also share certain common themes with our approach. Recent work by Gao et al.

(2008) uses English to improve the performance over a subset of Chinese queries whose translations in English are unambiguous. They use inter-document similarities across languages to improve the ranking performance. However, cross language document similarity measurement is in itself known to be an hard problem and the scale of their experimentation is quite small.

3 PRF in the LM Framework

The Language Modeling (LM) Framework allows PRF to be modelled in a principled manner. In the LM approach, documents and queries are modeled using multinomial distribution over words called *document language model* $P(w|D)$ and *query language model* $P(w|\Theta_Q)$ respectively. For a given query, the document language models are ranked based on their proximity to the query language model, measured using KL-Divergence.

$$KL(\Theta_Q||D) = \sum_w P(w|\Theta_Q) \cdot log \frac{P(w|\Theta_Q)}{P(w|D)}$$

Since the query length is short, it is difficult to estimate Θ_Q accurately using the query alone. In PRF, the top *k* documents obtained through the initial ranking algorithm are assumed to be relevant and used as feedback for improving the estimation of Θ_Q. The feedback documents contain both relevant and noisy terms from which the feedback language model is inferred based on a Generative Mixture Model (Zhai and Lafferty, 2001).

Let $D_F = \{d_1, d_2, \ldots, d_k\}$ be the top k documents retrieved using the initial ranking algorithm. Zhai and Lafferty (Zhai and Lafferty, 2001) model the feedback document set D_F as a mixture of two distributions: (a) the *feedback language model* and (b) the *collection model* $P(w|C)$. The feedback language model is inferred using the EM Algorithm (Dempster et al., 1977), which iteratively accumulates probability mass on the most *distinguishing* terms, *i.e.* terms which are more frequent in the feedback document set than in the entire collection. To maintain query focus the final converged feedback model, Θ_F is interpolated with the initial query model Θ_Q to obtain the final query model Θ_{Final}.

$$\Theta_{Final} = (1 - \alpha) \cdot \Theta_Q + \alpha \cdot \Theta_F$$

Θ_{Final} is used to re-rank the corpus using the KL-Divergence ranking function to obtain the final ranked list of documents. Henceforth, we refer to the above technique as *Model Based Feedback (MBF)*.

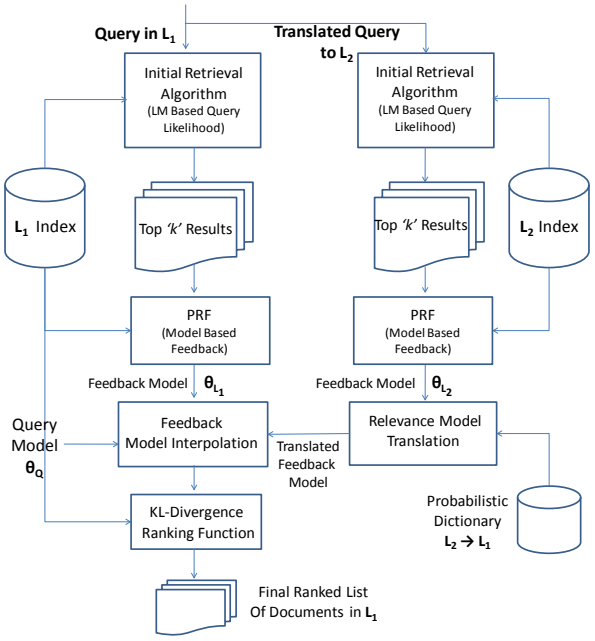


Figure 1: Schematic of the Multilingual PRF Approach

Symbol	Description
Θ_Q	Query Language Model
$\Theta^F_{L_1}$	Feedback Language Model obtained from PRF in L_1
$\Theta^F_{L_2}$	Feedback Language Model obtained from PRF in L_2
$\Theta^{Trans}_{L_1}$	Feedback Model Translated from L_2 to L_1
$t(f\|e)$	Probabilistic Bi-Lingual Dictionary from L_2 to L_1
β, γ	Interpolation coefficients coefficients used in MultiPRF

Table 2: Glossary of Symbols used in explaining MultiPRF

4 Multilingual PRF (MultiPRF)

The schematic of the MultiPRF approach is shown in Figure 1. Given a query Q in the source language L_1, we automatically translate the query into the assisting language L_2. We then rank the documents in the L_2 collection using the query likelihood ranking function (John Lafferty and Chengxiang Zhai, 2003). Using the top k documents, we estimate the feedback model using MBF as described in the previous section. Similarly, we also estimate a feedback model using the original query and the top k documents retrieved from the initial ranking in L_1. Let the resultant feedback models be $\Theta^F_{L_2}$ and $\Theta^F_{L_1}$ respectively. The feedback model estimated in the assisting language $\Theta^F_{L_2}$ is translated back into language L_1 using a probabilistic bi-lingual dictionary $t(f|e)$ from $L_2 \rightarrow L_1$ as follows:

$$P(f|\Theta^{Trans}_{L_1}) = \sum_{\forall\ e \text{ in } L_2} t(f|e) \cdot P(e|\Theta^F_{L_2}) \quad (1)$$

The probabilistic bi-lingual dictionary $t(f|e)$ is

Language	CLEF Collection Identifier	Description	No. of Documents	No. of Unique Terms	CLEF Topics (No. of Topics)
English	EN-00+01+02	LA Times 94	113005	174669	-
	EN-03+05+06	LA Times 94, Glasgow Herald 95	169477	234083	-
	EN-02+03	LA Times 94, Glasgow Herald 95	169477	234083	91-200 (67)
French	FR-00	Le Monde 94	44013	127065	1-40 (29)
	FR-01+02	Le Monde 94, French SDA 94	87191	159809	41-140 (88)
	FR-02+03	Le Monde 94, French SDA 94-95	129806	182214	91-200 (67)
	FR-03+05	Le Monde 94, French SDA 94-95	129806	182214	141-200,251-300 (99)
	FR-06	Le Monde 94-95, French SDA 94-95	177452	231429	301-350 (48)
German	DE-00	Frankfurter Rundschau 94, Der Spiegel 94-95	153694	791093	1-40 (33)
	DE-01+02	Frankfurter Rundschau 94, Der Spiegel 94-95, German SDA 94	225371	782304	41-140 (85)
	DE-02+03	Frankfurter Rundschau 94, Der Spiegel 94-95, German SDA 94-95	294809	867072	91-200 (67)
	DE-03	Frankfurter Rundschau 94, Der Spiegel 94-95, German SDA 94-95	294809	867072	141-200 (51)
Finnish	FI-02+03+04	Aamulehti 94-95	55344	531160	91-250 (119)
	FI-02+03	Aamulehti 94-95	55344	531160	91-200 (67)
Dutch	NL-02+03	NRC Handelsblad 94-95, Algemeen Dagblad 94-95	190604	575582	91-200 (67)
Spanish	ES-02+03	EFE 94, EFE 95	454045	340250	91-200 (67)

Table 1: Details of the CLEF Datasets used for Evaluating the MultiPRF approach. The number shown in brackets of the final column CLEF Topics indicate the actual number of topics used during evaluation.

Source Term	Top Aligned Terms in Target
French	**English**
américain	american, us, united, state, america
nation	nation, un, united, state, country
eťude	study, research, assess, investigate, survey
German	**English**
flugzeug	aircraft, plane, aeroplane, air, flight
spiele	play, game, stake, role, player
verhältnis	relationship, relate, balance, proportion

Table 3: Top Translation Alternatives for some sample words in Probabilistic Bi-Lingual Dictionary

learned from a parallel sentence-aligned corpora in $L_1 - L_2$ based on word level alignments. Tiedemann (Tiedemann, 2001) has shown that the translation alternatives found using word alignments could be used to infer various morphological and semantic relations between terms. In Table 3, we show the top translation alternatives for some sample words. For example, the French word *américain* (american) brings different variants of the translation like *american, america, us, united, state, america* which are lexically and semantically related. Hence, the probabilistic bi-lingual dictionary acts as a rich source of morphologically and semantically related feedback terms. Thus, during this step, of translating the feedback model as given in Equation 1, the translation model adds related terms in L_1 which have their source as the term from feedback model $\Theta^F_{L_2}$. The final MultiPRF model is obtained by interpolating the above translated feedback model with the original query model and the feedback model of language L_1 as given below:

$$\Theta^{Multi}_{L_1} = (1 - \beta - \gamma) \cdot \Theta_Q + \beta \cdot \Theta^F_{L_1} + \gamma \cdot \Theta^{Trans}_{L_1} \quad (2)$$

Since we want to retain the query focus during back translation the feedback model in L_2 is interpolated with the translated query before translation of the L_2 feedback model. The parameters β and γ control the relative importance of the original query model, feedback model of L_1 and the translated feedback model obtained from L_1 and are tuned based on the choice of L_1 and L_2.

5 Experimental Setup

We evaluate the performance of our system using the standard CLEF evaluation data in six languages, widely varying in their familial relationships - Dutch, German, English, French, Spanish and Finnish using more than 600 topics. The details of the collections and their corresponding topics used for MultiPRF are given in Table 1. Note that, in each experiment, we choose assisting collections such that the topics in the source language are covered in the assisting collection so as to get meaningful feedback terms. In all the topics, we only use the *title* field. We ignore the topics which have no relevant documents as the true performance on those topics cannot be evaluated.

We demonstrate the performance of MultiPRF approach with French, German and Finnish as source languages and Dutch, English and Spanish as the assisting language. We later vary the assisting language, for each source language and study the effects. We use the Terrier IR platform (Ounis et al., 2005) for indexing the documents. We perform standard tokenization, stop word removal and stemming. We use the Porter Stemmer for English and the stemmers available through the Snowball package for other languages. Other than these, we do not perform any language-specific processing on the languages. In case of French,

Collection	Assist. Lang	P@5			P@10			MAP			GMAP		
		MBF	MultiPRF	% Impr.	MBF	MultiPRF	% Impr.	MBF	MultiPRF	% Impr.	MBF	MultiPRF	% Impr.
FR-00	EN	0.4690	0.5241	**11.76‡**	0.4000	0.4000	0.00	0.4220	0.4393	4.10	0.2961	0.3413	**15.27**
	ES		0.5034	**7.35‡**		0.4103	2.59		0.4418	4.69		0.3382	**14.22**
	NL		0.5034	7.35		0.4103	2.59		0.4451	5.47		0.3445	**16.34**
FR-01+02	EN	0.4636	0.4818	3.92	0.4068	0.4386	**7.82‡**	0.4342	0.4535	**4.43‡**	0.2395	0.2721	**13.61**
	ES		0.4977	**7.35‡**		0.4363	**7.26‡**		0.4416	1.70		0.2349	-1.92
	NL		0.4818	3.92		0.4409	**8.38‡**		0.4375	0.76		0.2534	**5.80**
FR-03+05	EN	0.4545	0.4768	**4.89‡**	0.4040	0.4202	**4‡**	0.3529	0.3694	**4.67‡**	0.1324	0.1411	**6.57**
	ES		0.4727	4.00		0.4080	1.00		0.3582	1.50		0.1325	0.07
	NL		0.4525	-0.44		0.4010	-0.75		0.3513	0.45		0.1319	-0.38
FR-06	EN	0.4917	0.5083	3.39	0.4625	0.4729	2.25	0.3837	0.4104	6.97	0.2174	0.2810	**29.25**
	ES		0.5083	3.39		0.4687	1.35		0.3918	2.12		0.2617	**20.38**
	NL		0.5083	3.39		0.4646	0.45		0.3864	0.71		0.2266	**4.23**
DE-00	EN	0.2303	0.3212	**39.47‡**	0.2394	0.2939	**22.78‡**	0.2158	0.2273	5.31	0.0023	0.0191	**730.43**
	ES		0.3212	**39.47‡**		0.2818	**17.71‡**		0.2376	10.09		0.0123	**434.78**
	NL		0.3151	**36.82‡**		0.2818	**17.71‡**		0.2331	8.00		0.0122	**430.43**
DE-01+02	EN	0.5341	0.6000	**12.34‡**	0.4864	0.5318	**9.35‡**	0.4229	0.4576	**8.2‡**	0.1765	0.2721	**9.19**
	ES		0.5682	**6.39‡**		0.5091	**4.67‡**		0.4459	5.43		0.2309	**30.82**
	NL		0.5773	**8.09‡**		0.5114	**5.15‡**		0.4498	**6.35‡**		0.2355	**33.43**
DE-03	EN	0.5098	0.5412	6.15	0.4784	0.4980	4.10	0.4274	0.4355	1.91	0.1243	0.1771	**42.48**
	ES		0.5647	**10.77‡**		0.4980	4.10		0.4568	**6.89‡**		0.1645	**32.34**
	NL		0.5529	**8.45‡**		0.4941	3.27		0.4347	1.72		0.1490	**19.87**
FI-02+03+04	EN	0.3782	0.4034	**6.67‡**	0.3059	0.3319	**8.52‡**	0.3966	0.4246	**7.06‡**	0.1344	0.2272	**69.05**
	ES		0.3879	2.58		0.3267	6.81		0.3881	-2.15		0.1755	**30.58**
	NL		0.3948	4.40		0.3301	7.92		0.4077	2.79		0.1839	**36.83**

Table 4: Results comparing the performance of MultiPRF over baseline MBF on CLEF collections with English (EN), Spanish (ES) and Dutch (NL) as assisting languages. Results marked as ‡ indicate that the improvement was found to be statistically significant over the baseline at 90% confidence level ($\alpha = 0.01$) when tested using a paired two-tailed t-test.

since some function words like *l'*, *d' etc.,* occur as prefixes to a word, we strip them off during indexing and query processing, since it significantly improves the baseline performance. We use standard evaluation measures like *MAP*, *P@5* and *P@10* for evaluation. Additionally, for assessing robustness, we use the Geometric Mean Average Precision (GMAP) metric (Robertson, 2006) which is also used in the TREC Robust Track (Voorhees, 2006). The probabilistic bi-lingual dictionary used in MultiPRF was learnt automatically by running GIZA++: a word alignment tool (Och and Ney, 2003) on a parallel sentence aligned corpora. For all the above language pairs we used the *Europarl Corpus* (Philipp, 2005). We use Google Translate as the query translation system as it has been shown to perform well for the task (Wu et al., 2008). We use the MBF approach explained in Section 3 as a baseline for comparison. We use two-stage Dirichlet smoothing with the optimal parameters tuned based on the collection (Zhai and Lafferty, 2004). We tune the parameters of MBF, specifically λ and α, and choose the values which give the optimal performance on a given collection. We uniformly choose the top ten documents for feedback. Table 4 gives the overall results.

6 Results and Discussion

In Table 4, we see the performance of the MultiPRF approach for three assisting languages, and how it compares with the baseline MBF methods. We find MultiPRF to consistently outperform the baseline value on all metrics, namely MAP (where significant improvements range from 4.4% to 7.1%); P@5 (significant improvements range from 4.9% to 39.5% and P@10 (where MultiPRF has significant gains varying from 4% to 22.8%). Additionally we also find MultiPRF to be more robust than the baseline, as indicated by the GMAP score, where improvements vary from 4.2% to 730%. Furthermore we notice these trends hold across different assisting languages, with Spanish and Dutch outperforming English as the assisting language on some of the French and German collections. On performing a more detailed study of the results we identify the main reason for improvements in our approach is the ability to obtain good feedback terms in the assisting language coupled with the introduction of lexically and semantically related terms during the back-translation step.

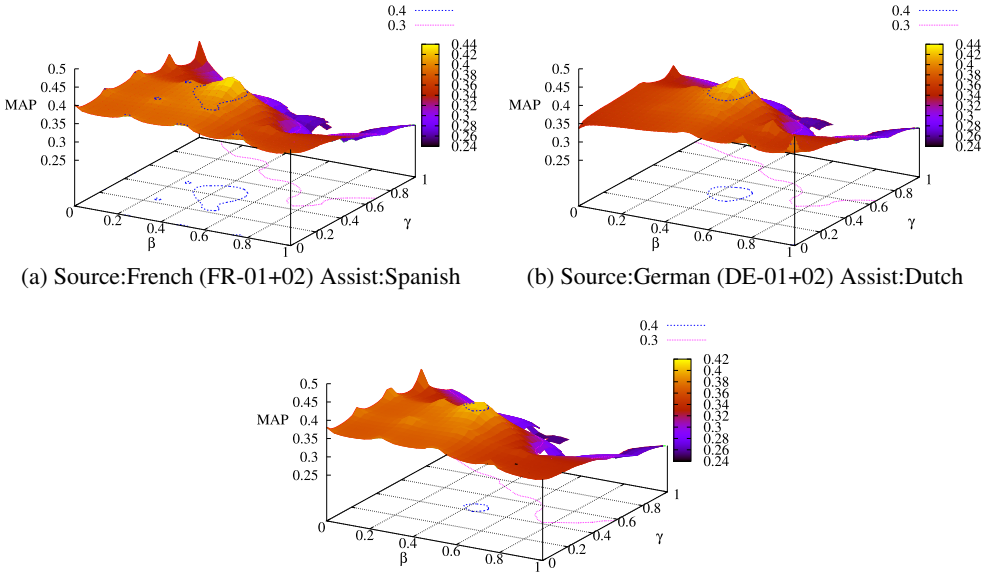
In Table 5, we see some examples, which illustrates the feedback terms brought by the MultiPRF method. As can be seen by these example, the gains achieved by MultiPRF are primarily due to one of three reasons: (a) Good Feedback in Assisting Language: If the feedback model in the assisting language contains good terms, then the back-translation process will introduce the corresponding feedback terms in the source language, thus leading to improved performance. As an example of this phenomena, consider the French Query *"Maladie de Creutzfeldt-Jakob"*. In this case the original feedback model also performs

TOPIC NO	ASSIST LANG.	SOURCE LANGUAGE QUERY	TRANSLATED QUERY	QUERY MEANING	MBF MAP	MPRF MAP	MBF- Top Representative Terms (With Meaning) Excl. Query Terms	MultiPRF- Top Representative Terms (With Meaning) Excl. Query Terms
GERMAN '01: TOPIC 61	EN	Ölkatastrophe in Sibirien	Oil Spill in Siberia	Siberian Oil Catastrophe	0.618	0.812	exxon, million, ol (oil), tonn, russisch (russian), olp (oil), moskau (moscow), us	olverschmutz (oil pollution), ol, russisch, erdol (petroleum), russland (russia), olunfall(oil spill), olp
GERMAN '02: TOPIC 105	ES	Bronchialasthma	El asma bronquial	Bronchial Asthma	0.062	0.636	chronisch (chronic), pet, athlet (athlete), ekrank (ill), gesund (healthy), tuberkulos (tuberculosis), patient, reis (rice), person	asthma, allergi, krankheit (disease), allerg (allergenic), chronisch, hauterkrank (illness of skin), arzt (doctor), erkrank (ill)
FRENCH '02: TOPIC 107	NL	Ingénierie génétique	Genetische Manipulatie	Genetic Engineering	0.145	0.357	dévelopg (developed), évolu (evolved), product, produit (product), moléculair (molecular)	genetic, gen, engineering, dévelopg, product
FRENCH '06: TOPIC 256	EN	Maladie de Creutzfeldt-Jakob	Creutzfeldt-Jakob	Creutzfeldt-Jakob Disease	0.507	0.688	malad (illness), produit (product), animal (animal), hormon (hormone)	malad, humain (human), bovin (bovine), encéphalopath (suffering from encephalitis), scientif, recherch (research)
GERMAN '03: TOPIC 157	EN	Siegerinnen von Wimbledon	Champions of Wimbledon	Wimbledon Lady Winners	0.074	0.146	telefonbuch (phone book), sieg (victory), titelseit (front page), telekom (telecommunication), graf	gross (large), verfecht (champion), sampra (sampras), 6, champion, steffi, verteidigt (defendending), martina, jovotna, navratilova
GERMAN '01: TOPIC 91	ES	AI in Lateinamerika	La gripe aviar en América Latina	AI in Latin America	0.456	0.098	international, amnesty, strassenkind (street child), kolumbi (Columbian), land, brasili (Brazil), menschenrecht (human rights), polizei (police)	karib (Caribbean), land, brasili, schuld (blame), amerika, kalt (cold), welt (world), forschung (research)
GERMAN '03: TOPIC 196	EN	Fusion japanischer Banken	Fusion of Japanese banks	Merger of Japanese Banks	0.572	0.264	daiwa, tokyo, filial (branch), zusammenschluss (merger)	kernfusion (nuclear fusion), zentralbank (central bank), daiwa, weltbank (world bank), investitionsbank (investment bank)
FRENCH '03: TOPIC 152	NL	Les droits de l'enfant	De rechten van het kind	Child Rights	0.479	0.284	convent (convention), franc, international, onun (united nations), réserv (reserve)	per (father), convent, franc, jurid (legal), homm (man), cour (court), biolog

Table 5: Qualitative comparison of feedback terms given by MultiPRF and MBF on representative queries where positive and negative results were observed in French and German collections.

quite strongly with a MAP score of 0.507. Although there is no significant topic drift in this case, there are not many relevant terms apart from the query terms. However the same query performs very well in English with all the documents in the feedback set of the English corpus being relevant, thus resulting in informative feedback terms such as {*bovin, scientif, recherch*}. (b) Finding Synonyms/Morphological Variations: Another situation in which MultiPRF leads to large improvements is when it finds semantically/lexically related terms to the query terms which the original feedback model was unable to. For example, consider the French query *"Ingénierie g´n´tique"*. While the feedback model was unable to find any of the synonyms of the query terms, due to their lack of co-occurence with the query terms, the MultiPRF model was able to get these terms, which are introduced primarily during the back-translation process. Thus terms like {*genetic, gen, engineering*}, which are synonyms of the query words, are found thus resulting in improved performance. (c) Combination of Above Factors: Sometimes a combination of the above two factors causes improvements in the performance as in the German query *"Ölkatastrophein Sibirien"*. For this query, MultiPRF finds good feedback terms such as {*russisch, russland*} while also obtaining semantically related terms such as {*olverschmutz, erdol, olunfall*}.

Although all of the previously described examples had good quality translations of the query in the assisting language, as mentioned in (Chinnakotla et al., 2010), the MultiPRF approach is robust to suboptimal translation quality as well. To see how MultiPRF leads to improvements even with errors in query translation consider the German Query *"Siegerinnen von Wimbledon"*. When this is translated to English, the term "Lady" is dropped, this causes only "Wimbledon Champions" to remain. As can be observed, this causes terms like *sampras* to come up in the MultiPRF model. However, while the MultiPRF model has some terms pertaining to Men's Winners of Wimbledon as well, the original feedback model suffers from severe topic drift, with irrelevant terms such as {*telefonbuch, telekom*} also amongst the top terms. Thus we notice that despite the error in query translation MultiPRF still manages to correct the drift of the original feedback model, while also introducing relevant terms such as {*verfecht, steffi, martina, novotna, navratilova*} as well. Thus as shown in (Chinnakotla et al., 2010), having a better query translation system can only lead to better performance. We also perform a detailed error analysis and found three main reasons for MultiPRF failing: (i) Inaccuracies in query translation (including the presence of out-of-vocabulary terms). This is seen in the German Query *AI in Lateinamerika*, which wrongly translates to *Avian Flu in Latin America* in Spanish thus affecting performance. (ii) Poor retrieval in Assisting Language. Consider the French query *Les droits de l'enfant*, for which due to topic drift in English, MultiPRF performance reduces. (iii) In a few rare cases inaccuracy in the back transla-



(a) Source:French (FR-01+02) Assist:Spanish

(b) Source:German (DE-01+02) Assist:Dutch

(c) Source:Finnish (FI-02+03+04) Assist:English

Figure 2: Results showing the sensitivity of MultiPRF performance to parameters β and γ for French, German and Finnish.

tion affects performance as well.

6.1 Parameter Sensitivity Analysis

The MultiPRF parameters β and γ in Equation 2 control the relative importance assigned to the original feedback model in source language L_1, the translated feedback model obtained from assisting language L_2 and the original query terms. We varied the β and γ parameters for French, German and Finnish collections with English, Dutch and Spanish as assisting languages and studied its effect on MAP of MultiPRF. The results are shown in Figure 2. The results show that, in all the three collections, the optimal value of the parameters almost remains the same and lies in the range of 0.4-0.48. Due to the above reason, we arbitrarily choose the parameters in the above range and do not use any technique to learn these parameters.

6.2 Effect of Assisting Language Choice

In this section, we discuss the effect of varying the assisting language. Besides, we also study the inter and intra familial behaviour of source-assisting language pairs. In order to ensure that the results are comparable across languages, we indexed the collections from the years 2002, 2003 and use common topics from the topic range 91-200 that have relevant documents across all the six languages. The number of such common topics were 67. For each source language, we use the other languages as assisting collections and study the performance of MultiPRF. Since query translation quality varies across language pairs, we analyze the behaviour of MultiPRF in the following two scenarios: (a) Using ideal query translation (b) Using Google Translate for query translation. In ideal query translation setup, in order to eliminate its effect, we skip the query translation step and use the corresponding original topics for each target language instead. The results for both the above scenarios are given in Tables 6 and 7.

From the results, we firstly observe that besides English, other languages such as French, Spanish, German and Dutch act as good assisting languages and help in improving performance over monolingual MBF. We also observe that the best assisting language varies with the source language. However, the crucial factors of the assisting language which influence the performance of MultiPRF are: (a) *Monolingual PRF Performance:* The main motivation for using a different language was to get good feedback terms, especially in case of queries which fail in the source language. Hence, an assisting language in which the monolingual feedback performance itself is poor, is unlikely to give any performance gains. This observation is evident in case of Finnish, which has the lowest Monolingual MBF performance. The results show that Finnish is the least helpful of assisting languages, with performance similar to those of the baselines. We also observe that the three best performing assistant languages, i.e. English, French and Spanish, have the highest monolingual performances as well, thus further validating the claim. One possible reason for this is the relative

Source Lang.		Assisting Language						Source Lang.MBF
		English	German	Dutch	Spanish	French	Finnish	
English	MAP	-	0.4464 (-0.7%)	0.4471 (-0.5%)	0.4566 (+1.6%)	0.4563 (+1.5%)	0.4545 (+1.1%)	0.4495
	P@5		0.4925 (-0.6%)	0.5045 (+1.8%)	0.5164 (+4.2%)	0.5075 (+2.4%)	**0.5194 (+4.8%)**	0.4955
	P@10		0.4343 (+0.4%)	0.4373 (+1.0%)	**0.4537 (+4.8%)**	0.4343 (+0.4%)	0.4373 (+1.0%)	0.4328
German	MAP	**0.4229 (+4.9%)**	-	**0.4346 (+7.8%)**	**0.4314 (+7.0%)**	0.411 (+1.9%)	0.3863 (-4.2%)	0.4033
	P@5	**0.5851 (+14%)**		**0.5851 (+14%)**	**0.5791 (+12.8%)**	**0.594 (+15.7%)**	**0.5522 (+7.6%)**	0.5134
	P@10	**0.5284 (+11.3%)**		**0.5209 (+9.8%)**	**0.5179 (+9.1%)**	**0.5149 (+8.5%)**	**0.5075 (+6.9%)**	0.4746
Dutch	MAP	0.4317 (+4%)	**0.4453 (+7.2%)**	-	0.4275 (+2.9%)	0.4241 (+2.1%)	0.3971 (-4.4%)	0.4153
	P@5	**0.5642 (+11.8%)**	**0.5731 (+13.6%)**		**0.5343 (+5.9%)**	**0.5582 (+10.6%)**	0.5045 (0%)	0.5045
	P@10	**0.5075 (+9%)**	**0.4925 (+5.8%)**		**0.4896 (+5.1%)**	**0.5015 (+7.7%)**	0.4806 (+3.2%)	0.4657
Spanish	MAP	0.4667 (-2.9%)	0.4749 (-1.2%)	0.4744 (-1.3%)	-	0.4609 (-4.1%)	0.4311 (-10.3%)	0.4805
	P@5	0.62 (-2.9%)	0.6418 (+0.5%)	0.6299 (-1.4%)		0.6269 (-1.6%)	0.6149 (-3.7%)	0.6388
	P@10	0.5625 (-1.8%)	0.5806 (+1.3%)	0.5851 (+2.1%)		0.5627 (-1.8%)	0.5478 (-4.4%)	0.5731
French	MAP	**0.4658 (+6.9%)**	0.4526 (+3.9%)	0.4374 (+0.4%)	**0.4634 (+6.4%)**	-	0.4451 (+2.2%)	0.4356
	P@5	0.4925 (+3.1%)	0.4806 (+0.6%)	0.4567 (-4.4%)	0.4925 (+3.1%)		0.4836 (+1.3%)	0.4776
	P@10	0.4358 (+3.9%)	0.4239 (+1%)	0.4224 (+0.7%)	**0.4388 (+4.6%)**		0.4209 (+0.4%)	0.4194
Finnish	MAP	0.3411 (-4.7%)	**0.3796 (+6.1%)**	0.3722 (+4%)	0.369 (+3.1%)	0.3553 (-0.7%)	-	0.3578
	P@5	0.394 (+3.1%)	**0.403 (+5.5%)**	**0.406 (+6.3%)**	**0.4119 (+7.8%)**	0.397 (+3.9%)		0.3821
	P@10	**0.3463 (+11.5%)**	**0.3582 (+15.4%)**	**0.3478 (+12%)**	**0.3448 (+11%)**	**0.3433 (+10.6%)**		0.3105

Table 6: Results showing the performance of MultiPRF with different source and assisting languages using Google Translate for query translation step. The intra-familial affinity could be observed from the elements close to the diagonal.

ease of processing in these languages. (b) *Familial Similarity Between Languages:* We observe that the performance of MultiPRF is good if the assisting language is from the same language family. Birch et al. (2008) show that the language family is a strong predictor of machine translation performance. Hence, the query translation and back translation quality improves if the source and assisting languages belong to the same family. For example, in the Germanic family, the source-assisting language pairs German-English, Dutch-English, Dutch-German and German-Dutch show good performance. Similarly, in Romance family, the performance of French-Spanish confirms this behaviour. In some cases, we observe that MultiPRF scores decent improvements even when the assisting language does not belong to the same language family as witnessed in French-English and English-French. This is primarily due to their strong monolingual MBF performance.

6.3 Effect of Language Family on Back Translation Performance

As already mentioned, the performance of MultiPRF is good if the source and assisting languages belong to the same family. In this section, we verify the above intuition by studying the impact of language family on back translation performance. The experiment designed is as follows: Given a query in source language L_1, the ideal translation in assisting language L_2 is used to compute the query model in L_2 using only the query terms. Then, without performing PRF the query model is directly back translated from L_2 into L_1 and finally documents are re-ranked using this translated feedback model. Since the automatic query translation and PRF steps have been eliminated, the only factor which influences the MultiPRF performance is the back-translation step. This means that the source-assisting language pairs for which the back-translation is good will score a higher performance. The results of the above experiment is shown in Table 8. For each source language, the best performing assisting languages have been highlighted.

Source Lang.	Assisting Language							
	FR	ES	DE	NL	EN	FI	MBF	MPRF
French	-	**0.3686**	0.3113	0.3366	**0.4338**	0.3011	0.4342	0.4535
Spanish	**0.3647**	-	0.3440	0.3476	**0.3954**	0.3036	0.5000	0.4892
German	0.2729	0.2736	-	**0.2951**	0.2107	0.2266	0.4229	0.4576
Dutch	0.2663	0.2836	**0.2902**	-	0.2757	0.2372	0.3968	0.3989

Table 8: Effect of Language Family on Back Translation Performance measured through MultiPRF MAP. 100 Topics from years 2001 and 2002 were used for all languages.

The results show that the performance of closely related languages like French-Spanish and German-Dutch is more when compared to other source-assistant language pairs. This shows that in case of closely related languages, the back-translation step succeeds in adding good terms which are relevant like morphological variants, synonyms and other semantically related terms. Hence, familial closeness of the assisting language helps in boosting the MultiPRF performance. An exception to this trend is English as assisting lan-

Source Lang.		Assisting Language						Source Lang.MBF
		English	German	Dutch	Spanish	French	Finnish	
English	MAP	-	0.4513 (+0.4%)	0.4475 (-0.4%)	**0.4695 (+4.5%)**	0.4665 (+3.8%)	0.4416 (-1.7%)	0.4495
	P@5		0.5104 (+3.0%)	0.5104 (+3.0%)	**0.5343 (+7.8%)**	**0.5403 (+9.0%)**	0.4806 (-3.0%)	0.4955
	P@10		0.4373 (+1.0%)	0.4358 (+0.7%)	**0.4597 (+6.2%)**	**0.4582 (+5.9%)**	0.4164 (-3.8%)	0.4328
German	MAP	**0.4427 (+9.8%)**	-	**0.4306 (+6.8%)**	**0.4404 (+9.2%)**	0.4104 (+1.8%)	0.3993 (-1.0%)	0.4033
	P@5	**0.606 (+18%)**		**0.5672 (+10.5%)**	**0.594 (+15.7%)**	**0.5761 (+12.2%)**	**0.5552 (+8.1%)**	0.5134
	P@10	**0.5373 (+13.2%)**		**0.503 (+6.0%)**	**0.5299 (+11.7%)**	0.494 (+4.1%)	**0.5 (+5.4%)**	0.4746
Dutch	MAP	**0.4361 (+5.0%)**	**0.4344 (+4.6%)**	-	0.4227 (+1.8%)	0.4304 (+3.6%)	0.4134 (-0.5%)	0.4153
	P@5	**0.5761 (+14.2%)**	**0.5552 (+10%)**		**0.5403 (+7.1%)**	**0.5463 (+8.3%)**	**0.5433 (+7.7%)**	0.5045
	P@10	**0.5254 (+12.8%)**	**0.497 (+6.7%)**		0.4776 (+2.6%)	**0.5134 (+10.2%)**	**0.4925 (+5.8%)**	0.4657
Spanish	MAP	0.4665 (-2.9%)	0.4773 (-0.7%)	0.4733 (-1.5%)	-	0.4839 (+0.7%)	0.4412 (-8.2%)	0.4805
	P@5	0.6507 (+1.8%)	0.6448 (+0.9%)	0.6507 (+1.8%)		0.6478 (+1.4%)	0.597 (-6.5%)	0.6388
	P@10	0.5791 (+1.0%)	0.5791 (+1.0%)	0.5761 (+0.5%)		0.5866 (+2.4%)	0.5567 (-2.9%)	0.5731
French	MAP	**0.4591 (+5.4%)**	0.4514 (+3.6%)	0.4409 (+1.2%)	**0.4712 (+8.2%)**	-	0.4354 (0%)	0.4356
	P@5	0.4925 (+3.1%)	0.4776 (0%)	0.4776 (0%)	**0.4995 (+4.6%)**		0.4955 (+3.8%)	0.4776
	P@10	**0.4463 (+6.4%)**	0.4313 (+2.8%)	0.4373 (+4.3%)	**0.4448 (+6.1%)**		0.4209 (+0.3%)	0.4194
Finnish	MAP	0.3733 (+4.3%)	0.3559 (-0.5%)	0.3676 (+2.7%)	0.3594 (+0.4%)	0.371 (+3.7%)	-	0.3578
	P@5	**0.4149 (+8.6%)**	0.385 (+0.7%)	0.388 (+1.6%)	0.388 (+1.6%)	0.3911 (+2.4%)		0.3821
	P@10	**0.3567 (+14.9%)**	0.31 (-0.2%)	**0.3253 (+4.8%)**	0.32 (+3.1%)	0.3239 (+4.3%)		0.3105

Table 7: Results showing the performance of MultiPRF without using automatic query translation *i.e.* by using corresponding original queries in assisting collection. The results show the potential of MultiPRF by establishing a performance upper bound.

guage which shows good performance across both families.

6.4 Multiple Assisting Languages

So far, we have only considered a single assisting language. However, a natural extension to the method which comes to mind, is using multiple assisting languages. In other words, combining the evidence from all the feedback models of more than one assisting language, to get a feedback model which is better than that obtained using a single assisting language. To check how this simple extension works, we performed experiments using a pair of assisting languages. In these experiments for a given source language (from amongst the 6 previously mentioned languages) we tried using all pairs of assisting languages (for each source language, we have 10 pairs possible). To obtain the final model, we simply interpolate all the feedback models with the initial query model, in a similar manner as done in MultiPRF. The results for these experiments are given in Table 9. As we see, out of the 60 possible combinations of source language and assisting language pairs, we obtain improvements of greater than 3% in 16 cases. Here the improvements are with respect to the best model amongst the two MultiPRF models corresponding to each of the two assisting languages, with the same source language. Thus we observe that a simple linear interpolation of models is not the best way of combining evidence from multiple assisting languages. We also observe than when German or Spanish are used as one of the two assisting languages, they are most likely to lead to improvements. A more detailed study of this observation needs to be done to explain this.

Source Language	Assisting Language Pairs with Improvement >3%
English	FR-DE (4.5%), FR-ES (4.8%), DE-NL (+3.1%)
French	EN-DE (4.1%), DE-ES (3.4%), NL-FI (4.8%)
German	None
Spanish	None
Dutch	EN-DE (3.9%), DE-FR (4.1%), FR-ES (3.8%), DE-ES (3.9%)
Finnish	EN-ES (3.2%), FR-DE (4.6%), FR-ES (6.4%), DE-ES (11.2%), DE-NL (4.4%), ES-NL (5.9%)
Total - 16	EN – 3 Pairs; FR – 6 Pairs; DE – 10 Pairs; ES - 8 Pairs; NL – 4 Pairs; FI – 1 Pair

Table 9: Summary of MultiPRF Results with Two Assisting Languages. The improvements described above are with respect to maximum MultiPRF MAP obtained using either L_1 or L_2 alone as assisting language.

7 Conclusion and Future Work

We studied the effect of different *source-assistant* pairs and multiple assisting languages on the performance of MultiPRF. Experiments across a wide range of language pairs with varied degree of familial relationships show that MultiPRF improves performance in most cases with the performance improvement being more pronounced when the source and assisting languages are *closely related*. We also notice that the results are mixed when two assisting languages are used simultaneously. As part of future work, we plan to vary the model interpolation parameters dynamically to improve the performance in case of multiple assisting languages.

Acknowledgements

The first author was supported by a fellowship award from Infosys Technologies Ltd., India. We would like to thank Mr. Vishal Vachhani for his help in running the experiments.

References

Giambattista Amati, Claudio Carpineto, and Giovanni Romano. 2004. Query Difficulty, Robustness, and Selective Application of Query Expansion. In *ECIR '04*, pages 127–137.

Alexandra Birch, Miles Osborne and Philipp Koehn. 2008. Predicting Success in Machine Translation. In *EMNLP '08*, pages 745-754, ACL.

Martin Braschler and Carol Peters. 2004. Cross-Language Evaluation Forum: Objectives, Results, Achievements. *Inf. Retr.*, 7(1-2):7–31.

Martin Braschler and Peter Schäuble. 1998. Multilingual Information Retrieval based on Document Alignment Techniques. In *ECDL '98*, pages 183–197, Springer-Verlag.

Chris Buckley, Gerald Salton, James Allan, and Amit Singhal. 1994. Automatic Query Expansion using SMART : TREC 3. In *TREC-3*, pages 69–80.

Guihong Cao, Jian-Yun Nie, Jianfeng Gao, and Stephen Robertson. 2008. Selecting Good Expansion Terms for Pseudo-Relevance Feedback. In *SIGIR '08*, pages 243–250. ACM.

Manoj K. Chinnakotla, Karthik Raman, and Pushpak Bhattacharyya. 2010. Multilingual PRF: English Lends a Helping Hand. In *SIGIR '10*, ACM.

Kevyn Collins-Thompson and Jamie Callan. 2005. Query Expansion Using Random Walk Models. In *CIKM '05*, pages 704–711. ACM.

Steve Cronen-Townsend, Yun Zhou, and W. Bruce Croft. 2004. A Framework for Selective Query Expansion. In *CIKM '04*, pages 236–237. ACM.

Ido Dagan, Alon Itai, and Ulrike Schwall. 1991. Two Languages Are More Informative Than One. In *ACL '91*, pages 130–137. ACL.

A. Dempster, N. Laird, and D. Rubin. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39:1–38.

T. Susan Dumais, A. Todd Letsche, L. Michael Littman, and K. Thomas Landauer. 1997. Automatic Cross-Language Retrieval Using Latent Semantic Indexing. In *AAAI '97*, pages 18–24.

Wei Gao, John Blitzer, and Ming Zhou. 2008. Using English Information in Non-English Web Search. In *iNEWS '08*, pages 17–24. ACM.

David Hawking, Paul Thistlewaite, and Donna Harman. 1999. Scaling Up the TREC Collection. *Inf. Retr.*, 1(1-2):115–137.

Hieu Hoang, Alexandra Birch, Chris Callison-burch, Richard Zens, Rwth Aachen, Alexandra Constantin, Marcello Federico, Nicola Bertoldi, Chris Dyer, Brooke Cowan, Wade Shen, Christine Moran, and Ondej Bojar. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *ACL '07*, pages 177–180.

P. Jourlin, S. E. Johnson, K. Spärck Jones and P. C. Woodland. 1999. Improving Retrieval on Imperfect Speech Transcriptions (Poster Abstract). In *SIGIR '99*, pages 283–284. ACM.

John Lafferty and Chengxiang Zhai. 2003. Probabilistic Relevance Models Based on Document and Query Generation. *Language Modeling for Information Retrieval*, pages 1–10. Kluwer International Series on IR.

K. Sparck Jones, S. Walker, and S. E. Robertson. 2000. A Probabilistic Model of Information Retrieval: Development and Comparative Experiments. *Inf. Process. Manage.*, 36(6):779–808.

John Lafferty and Chengxiang Zhai. 2001. Document Language Models, Query Models, and Risk Minimization for Information Retrieval. In *SIGIR '01*, pages 111–119. ACM.

Victor Lavrenko and W. Bruce Croft. 2001. Relevance Based Language Models. In *SIGIR '01*, pages 120–127. ACM.

Victor Lavrenko, Martin Choquette, and W. Bruce Croft. 2002. Cross-Lingual Relevance Models. In *SIGIR '02*, pages 175–182, ACM.

Edgar Meij, Dolf Trieschnigg, Maarten Rijke de, and Wessel Kraaij. 2009. Conceptual Language Models for Domain-specific Retrieval. *Information Processing & Management*, 2009.

Donald Metzler and W. Bruce Croft. 2007. Latent Concept Expansion Using Markov Random Fields. In *SIGIR '07*, pages 311–318. ACM.

Mandar Mitra, Amit Singhal, and Chris Buckley. 1998. Improving Automatic Query Expansion. In *SIGIR '98*, pages 206–214. ACM.

Franz Josef Och and Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19–51.

I. Ounis, G. Amati, Plachouras V., B. He, C. Macdonald, and Johnson. 2005. Terrier Information Retrieval Platform. In *ECIR '05*, volume 3408 of *Lecture Notes in Computer Science*, pages 517–519. Springer.

Koehn Philipp. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *MT Summit '05*.

Stephen Robertson. 2006. On GMAP: and Other Transformations. In *CIKM '06*, pages 78–83. ACM.

Tetsuya Sakai, Toshihiko Manabe, and Makoto Koyama. 2005. Flexible Pseudo-Relevance Feedback Via Selective Sampling. *ACM TALIP*, 4(2):111–135.

Tao Tao and ChengXiang Zhai. 2006. Regularized Estimation of Mixture Models for Robust Pseudo-Relevance Feedback. In *SIGIR '06*, pages 162–169. ACM.

Tuomas Talvensaari, Jorma Laurikkala, Kalervo Järvelin, Martti Juhola, and Heikki Keskustalo. 2007. Creating and Exploiting a Comparable Corpus in Cross-language Information Retrieval. *ACM Trans. Inf. Syst.*, 25(1):4, 2007.

Jrg Tiedemann. 2001. The Use of Parallel Corpora in Monolingual Lexicography - How word alignment can identify morphological and semantic relations. In *COMPLEX '01*, pages 143–151.

Ellen M. Voorhees. 1994. Query Expansion Using Lexical-Semantic Relations. In *SIGIR '94*, pages 61–69. Springer-Verlag.

Ellen Voorhees. 2006. Overview of the TREC 2005 Robust Retrieval Track. In *TREC 2005*, Gaithersburg, MD. NIST.

Dan Wu, Daqing He, Heng Ji, and Ralph Grishman. 2008. A Study of Using an Out-of-Box Commercial MT System for Query Translation in CLIR. In *iNEWS '08*, pages 71–76. ACM.

Jinxi Xu and W. Bruce Croft. 2000. Improving the Effectiveness of Information Retrieval with Local Context Analysis. *ACM Trans. Inf. Syst.*, 18(1):79–112.

Jinxi Xu, Alexander Fraser, and Ralph Weischedel. 2002. Empirical Studies in Strategies for Arabic Retrieval. In *SIGIR '02*, pages 269–274. ACM.

Yang Xu, Gareth J.F. Jones, and Bin Wang. 2009. Query Dependent Pseudo-Relevance Feedback Based on Wikipedia. In *SIGIR '09*, pages 59–66. ACM.

Chengxiang Zhai and John Lafferty. 2001. Model-based Feedback in the Language Modeling approach to Information Retrieval. In *CIKM '01*, pages 403–410. ACM.

Chengxiang Zhai and John Lafferty. 2004. A Study of Smoothing Methods for Language Models applied to Information Retrieval. *ACM Transactions on Information Systems*, 22(2):179–214.