

**Exeter at CLEF 2003: Experiments with Machine
Translation for Monolingual, Bilingual and
Multilingual Retrieval**

Adenike M. Lam-Adesina

Department of Computer Science, University of Exeter, U.K.

Gareth J. F. Jones

School of Computing, Dublin City University, Ireland

Main Objective of Participation

Explore effectiveness of monolingual, bilingual and multilingual retrieval for non-English documents using only English retrieval.

Overview

- Non-English documents and topics translated into English using commercial machine translation.
 - Documents translated using *SYSTRAN 3.0*.
 - Topics translated separately using *SYSTRAN 3.0* and *Globalink Power Translator Pro 6.4*.

CLIR Strategy

- Retrieval using the City University research distribution of the Okapi system.
 - Around 260 stop words removed from the texts, Porter stemming applied, small set of standard synonyms plus list of U.S. and U.K. English spellings used.
- Okapi augmented with summary-based pseudo relevance feedback (Lam-Adesina & Jones SIGIR 2001).
- Pilot searching incorporating TREC-8 ad hoc document collection.

Monolingual German Runs

	Systran MT		Pro MT	
	Av.P.	% chg.	Av.P.	% chg.
Baseline	0.488	—	0.441	—
TC $ow(i)$, TC $cfw(i)$	0.568*	+16.4%	0.511*	+15.9%
PC $ow(i)$, TC $cfw(i)$	0.512*	+4.9%	0.457	+3.6%
PC $ow(i)$, PC $cfw(i)$	0.458	+6.1%	0.431	-2.3%
CC $ow(i)$, TC $cfw(i)$	0.550	+12.7%	0.494	+12.0%
CC $ow(i)$, CC $cfw(i)$	0.551	+12.9%	0.512	+16.1%

All runs use Topic Title and Description fields.

Monolingual French Runs

	Systran MT		Pro MT	
	Av.P.	% chg.	Av.P.	% chg.
Baseline	0.487	—	0.422	—
TC $ow(i)$, TC $cfw(i)$	0.521*	+6.9%	0.457*	+8.3%
PC $ow(i)$, TC $cfw(i)$	0.491*	+0.8%	0.403	+4.5%
PC $ow(i)$, PC $cfw(i)$	0.489	+0.4%	0.426	+0.9%
CC $ow(i)$, TC $cfw(i)$	0.519	+6.6%	0.446	+5.7%
CC $ow(i)$, CC $cfw(i)$	0.553	+13.6%	0.467	+10.7%

All runs use Topic Title and Description fields.

Monolingual Italian Runs

	Systran MT		Pro MT	
	Av.P.	% chg.	Av.P.	% chg.
Baseline	0.419	—	0.387	—
TC $ow(i)$, TC $cfw(i)$	0.494*	+17.9%	0.449*	+16.0%
PC $ow(i)$, TC $cfw(i)$	0.432*	+3.1%	0.402	+3.9%
PC $ow(i)$, PC $cfw(i)$	0.393	-6.2%	0.387	0.0%
CC $ow(i)$, TC $cfw(i)$	0.456	+8.8%	0.452	+16.8%
CC $ow(i)$, CC $cfw(i)$	0.454	+8.4%	0.481	+24.3%

All runs use Topic Title and Description fields.

Monolingual Spanish Runs

	Systran MT		Pro MT	
	Av.P.	% chg.	Av.P.	% chg.
Baseline	0.422	—	0.393	—
TC $ow(i)$, TC $cfw(i)$	0.470*	+11.3%	0.452*	+15.0%
PC $ow(i)$, TC $cfw(i)$	0.426*	+0.9%	0.415	+5.6%
PC $ow(i)$, PC $cfw(i)$	0.372	-11.8%	0.397	+1.0%
CC $ow(i)$, TC $cfw(i)$	0.462	+9.5%	0.466	+18.6%
CC $ow(i)$, CC $cfw(i)$	0.470	+11.3%	0.462	+17.6%

All runs use Topic Title and Description fields.

Observations

- General trends are consistent for the different languages.
- Systran topic translation gives better results than Power Translator Pro.
- Pseudo-relevance feedback on the test collection only is consistently effective.
- Pilot searching using only the English document pilot collection is generally poor.
- Pilot searching using the merged multilingual collection generally gives similar results to test collection only retrieval.

Bilingual German to Italian

	Systran MT		Pro MT	
	Av.P.	% chg.	Av.P.	% chg.
Baseline	0.311	—	0.314	—
TC $ow(i)$, TC $cfw(i)$	0.370	+18.9%	0.359	+14.3%
PC $ow(i)$, TC $cfw(i)$	0.339	+9.0%	0.334	+6.4%
PC $ow(i)$, PC $cfw(i)$	0.327	+5.1%	0.335*	+6.7%
TopicC $ow(i)$, TC $cfw(i)$	0.365	+17.4%	0.355*	+13.1%
TopicC $ow(i)$, TopicC $cfw(i)$	0.415*	+33.4%	0.397*	+26.4%
TopicC + TC $ow(i)$, TopicC $cfw(i)$	0.433	+39.2%	0.418	+33.1%

Bilingual Italian to Spanish

	Systran MT		Pro MT	
	Av.P.	% chg.	Av.P.	% chg.
Baseline	0.327	—	0.349	—
TC $ow(i)$, TC $cfw(i)$	0.376	+14.9%	0.417	+19.5%
PC $ow(i)$, TC $cfw(i)$	0.331	+1.2%	0.365	+4.6%
PC $ow(i)$, PC $cfw(i)$	0.339	+3.7%	0.365*	+4.3%
TopicC $ow(i)$, TC $cfw(i)$	0.389	+19.0%	0.417*	+19.5%
TopicC $ow(i)$, TopicC $cfw(i)$	0.391*	+19.6%	0.385*	+10.3%
TopicC + TC $ow(i)$, TopicC $cfw(i)$	0.389	+19.0%	0.379	+8.6%

Observations

- Pilot searching is again less effective than using the test collection.
- Pre- and post-“translation” feedback can be effective, but this is not always the case using this approach.
- Use of $cfw(i)$ from a collection in the topic language can be effective, but again this is not always the case.

Multilingual Retrieval

	Prec.10	Av.P.	% chg.	Rel. Ret.
Baseline	0.593	0.383	—	4613
Raw Score Merge	0.532	0.356*	-7.0%	4457
Global Max. Merge	0.518	0.356*	-7.0%	4428
CC $ow(i)$, CC $cfw(i)$	0.617	0.425	+10.9%	4853
PC $ow(i)$, CC $cfw(i)$	0.623	0.438*	+14.3%	4828
CC $ow(i)$, Global Max. Merge	0.593	0.400	+4.4%	4675

Title and Description fields from English topics.

Observations

- Combination of the document sets into a single collection looks the more effective strategy.
- In this case, pilot collection topic expansion is more effective than test collection expansion for the combined collection, but not the individual collections. Why?

Further Work

- Additional experiments and investigation to explore unresolved term weighting and topic expansions issues.
- Detailed investigation of fusion methods and behaviour.
- Indexing and retrieval in original document language.