TLR at CLEF-2004: Capturing the translation of compounds and other experiments

Isabelle Moulinier
TLR, Research and Development
isabelle.moulinier@thomson.com

September 15, 2004

Setting up the context

- Our production environment: mostly Boolean queries
- Interest in word-by-word translation
- Learning bilingual lexicons from corpora using statistical MT

↪ Can we use information beyond translation probabilities?

Our CLEF-2004 effort: focus on compound translation

- Observation 1: German compounds typically translate into French phrases

- Observation 2: IBM translation model 3 introduces fertility and distortion. Fertility probabilities capture 1-to-many term translations.

Two translation approaches

- `lex` select at most n most probable translations $f_1, \ldots, f_n$ ranked according to their translation probabilities $t(f_i|d) \geq p_{min}$

- `fert` select a translation set of the m most probable target terms $f_1, \ldots, f_m$, ranked according to their translation probabilities $t^3(f_i|d)$. The number of selected terms m is given by

$$\mathrm{ArgMax}_{\phi} \begin{cases} n(\phi|d) * p_0 & \text{if } \phi = 0 \\ n(\phi|d) * \sum_{i=1}^{\phi} t^3(f_i|d) & \text{if } \phi > 0 \end{cases}$$

Translation examples

Term: **globale**					
f	$t^3(f\|d)$	ϕ	$n(\phi\|d)$	`lex` trans.	`fert` trans.
globale	0.306778	1	0.746871	globale	globale
mondiale	0.152177	0	0.165741	mondiale	
global	0.115814	2	0.0617001	global	
mondial	0.0928475	3	0.0207158		
chelle	0.0456918		. . .		
Term: **Klimaveränderungen**					
f	$t^3(f\|d)$	ϕ	$n(\phi\|d)$	`lex` trans.	`fert` trans.
climatiques	0.269569	2	0.589625	climatiques	climatiques
changements	0.258488	1	0.105312	changements	changements
changement	0.105622	3	0.0936477	changement	
climatique	0.103034	4	0.07117		
climat	0.0250892		. . .		

Term: **Treibhauseffektes**					
f	$t^3(f\|d)$	ϕ	$n(\phi\|d)$	`lex` trans.	`fert` trans.
effet	0.265273	2	0.283692	effet	effet
serre	0.26525	1	0.246126	serre	serre
venir	0.0380016	3	0.174969		
mes	0.0191118	9	0.0651408		
Term: **Lawinenunglücken**					
f	$t^3(f\|d)$	ϕ	$n(\phi\|d)$	`lex` trans.	`fert` trans.
avalanches	0.10976	1	0.404492	avalanches	avalanches
programmer	0.10976	2	0.231625	programmer	programmer
servir	0.10976	3	0.1003	servir	servir
court	0.10976	0	0.0752761		court
interventions	0.10976	9	0.0611943		interventions
diverses	0.109759	4	0.0435146		diverses
série	0.109759				série
pourquoi	0.109759		. . .		pourquoi
zones	0.109624				zones

Handling of non-founds

- Bilingual lexicons contain surface forms
- When no translation is found, translate stem and compound parts

German Term	French translations
Wohnungsbrände	logement, incendie, au, feu
Weltmeisterin	du, champions, monde

Experimental Results

- We use GIZA++ and the Europarl corpus to build translation resources
- We use the WIN (cousin of Inquery) search engine
- We compared
  - translation models
  - query structure for compound translation

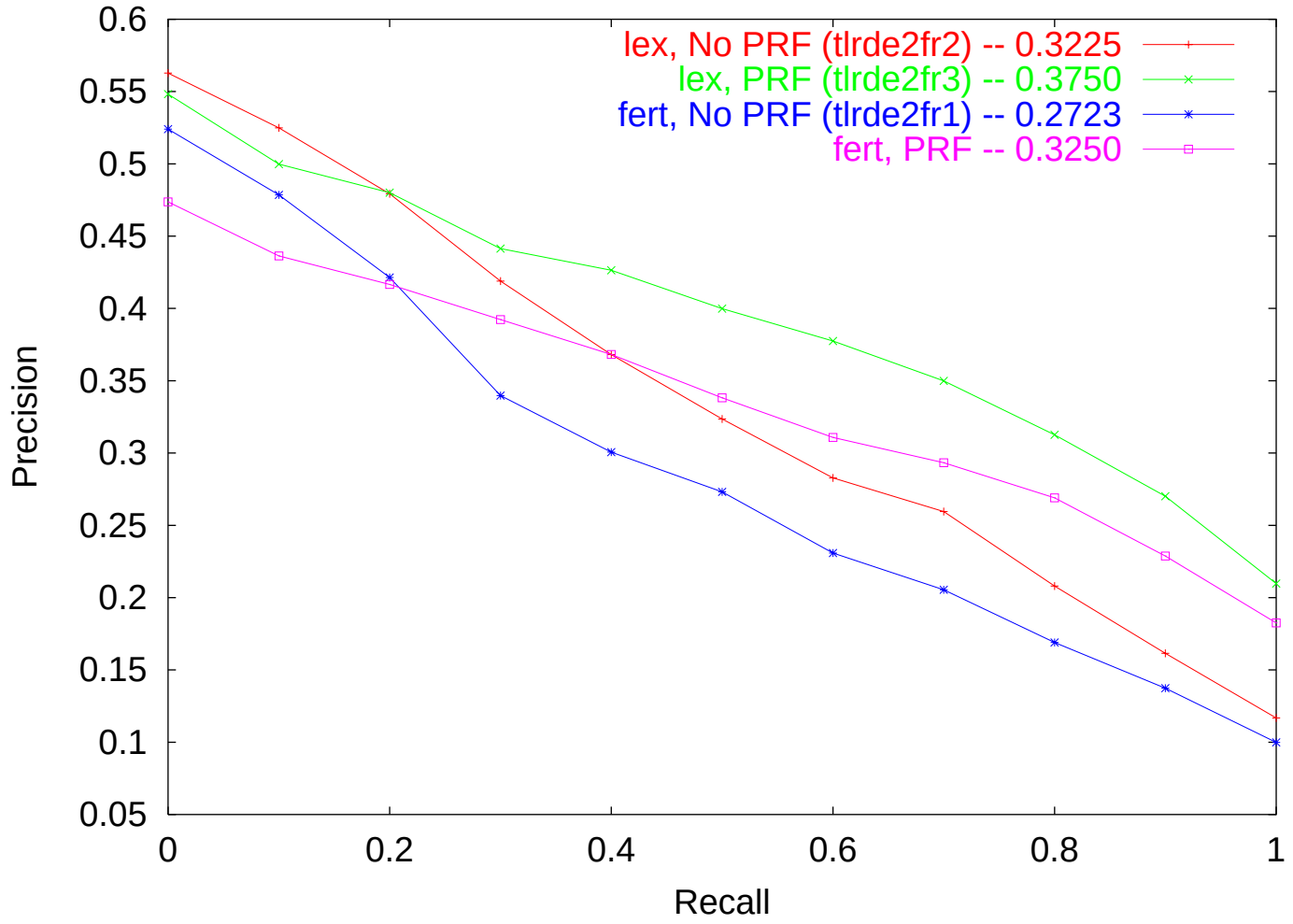
Run	Avg. Prec.	R-Prec.	Prec. at 20 doc.
t^1, `lex`, #SUM, $n = 3$	0.2934	0.2951	0.2224
t^3, `lex`, #SUM, $n = 3$	0.3225	0.3250	0.2541
t^3, `fert`, #SUM, $n = 1$	0.2717	0.2868	0.2133
t^3, `fert`, #NPHR, $n = 1$	0.2708	0.2779	0.2153
t^3, `lex`, #SUM, $n = 1$	0.2641	0.2755	0.2051

Pseudo-relevance feedback

- Expansion using Rocchio-like approach: select 20 terms from top 5 documents as relevant and bottom 20 as non-relevant, $\alpha = 0$, $\beta = 1$, $\gamma = 1$ or $\gamma = 4$

- We observed often-reported behavior
  - Overall improvement in average precision
  - Seesaw effect: over 50% of queries improve, around 40% degrade

PRF experimental results



Can we improve?

- Do translation probabilities improve retrieval with increasing model complexity?
- Model 3 also generates distortion probabilities. Can we integrate them as well?
- Overcoming shortcomings:
  - Training bias versus search bias (cf. *globale* versus *mondiale*) – can we integrate search priors during translation selection?
  - Picking too many translations (cf. *Lawinenunglücken*) – can we reformulate our idea and prevent such behavior?
- Can we better integrate translation models with Win?
- Selective application of relevance feedback