Machine Translation of Canadian Court Decisions

L. Langlois, M. Simard, E. Macklovitch

AMTA 2016

Austin TX, October 2016

This presentation is dedicated to Lucie Langlois

Nous dédions cette présentation à Lucie Langlois

Courts Administration Service

- Created in 2003 to rationalize services offered to four Canadian tribunals:

    - Federal Court of Appeal, Federal Court, Court Martial Appeal Court, Tax Court of Canada

- Responsible for meeting courts' administrative needs & ensuring public access to all court records & decisions

- OLA: court decisions *must* be published in both official languages

Federal Court   Cour fédérale

Ottawa, Ontario, November 27, 2015

PRESENT: The Honourable Mr. Justice Locke

BETWEEN:

**ALCON CANADA INC.,
ALCON LABORATORIES, INC.,
ALCON PHARMACEUTICALS LTD.,
and ALCON RESEARCH, LTD.**

Plaintiffs/
Defendants by Counterclaim

and

ACTAVIS PHARMA COMPANY

Defendant/
Plaintiff by Counterclaim

ORDER AND REASONS

I.   Overview

[1]   The plaintiffs, Alcon Canada Inc., Alcon Laboratories, Inc., Alcon Pharmaceuticals Ltd.,

and Alcon Research, Ltd. (collectively referred to as Alcon), appeal from an Order of

Prothonotary Martha Milczynski dated September 24, 2015, dismissing Alcon's motion to strike

CAS and Translation

- CAS responsible for ensuring the timely translation & publication of all court decisions
  - approx. 8 million words/year; mostly Eng > Fr
  - all outsourced; revised internally by jurilinguists
  - requirement for high quality; both linguistic versions have equal force before the law
  - simultaneous publication on Web
- Long translation delays; traditional workflow unable to cope

A pilot project in MT

- Launched at the initiative of L. Langlois, DG, Judicial Services
  - extensive experience in NLP and translation
  - in her view, solution could only come from MT
- LL contacts NRC re: MT pilot (early 2015)
  - contacts EM to act as independent consultant
  - imposing translation workload, but CAS has assets
  - NRC begins by analysing available corpora

NRC at CAS

- Main deliverable: provide CAS with the best possible Machine Translation

- Strategy: Build specialized MT engines for each of the four tribunals

- MT technology: NRC's Portage
  - Phrase-based MT technology
  - Continuous development since 2004
  - Participated in numerous shared tasks: WMT, NIST, etc.
  - Commercially available since 2010

Building Specialized Engines

General procedure for building specialized MT:

- Collect domain translations

- Process corpus

- Train engines

- Test and evaluate

- Repeat

Collecting CAS Data

- Historically, all translation was outsourced
  → no structured Translation Memory (TM)

However...

- All CAS court decisions are on the Web since the mid-1990s
  → all decisions of the last 20 years available in HTML format

Collecting CAS Data

		CMAC	FC	FCA	TCC	Total
Documents	Paired	142	25.3k	6.5k	8.8k	**40.7k**
	Orphan	1	1.9k	348	561	**2813**
TU's		28k	3.4M	888k	1.8M	**6.6M**
Words	EN	600k	89M	17M	35M	**141.6M**
	FR	600k	103M	19M	41M	**163.6M**

CMAC = Court Martial Appeals Court
FC = Federal Court
FCA = Federal Court of Appeal
TCC = Tax Court of Canada

CAS Data Analysis: Linguistic Complexity

Corpus	Court	Type-Token Ratio (@100k words)	Growth Rate (@100k words)	BLEU
References	"Rich"	0.141	1/13	33.2
	"Medium"	0.109	1/18	42.9
	"Poor"	0.078	1/26	50.7
	Weather Reports	0.018	1/200	↑
CAS	CMAC	0.079	1/29	
	FC	0.103	1/19	?
	FCA	0.101	1/19	
	TCC	0.094	1/20	

CAS Data Analysis: Translation Memory

TM coverage (% source words)

Court-specific TMs	Court	70%+	85%+	Exact
	CMAC	7.2	5.9	3.9
	FC	13.6	11.6	9.0
	FCA	11.9	10.3	7.7
	TCC	12.5	9.9	6.8

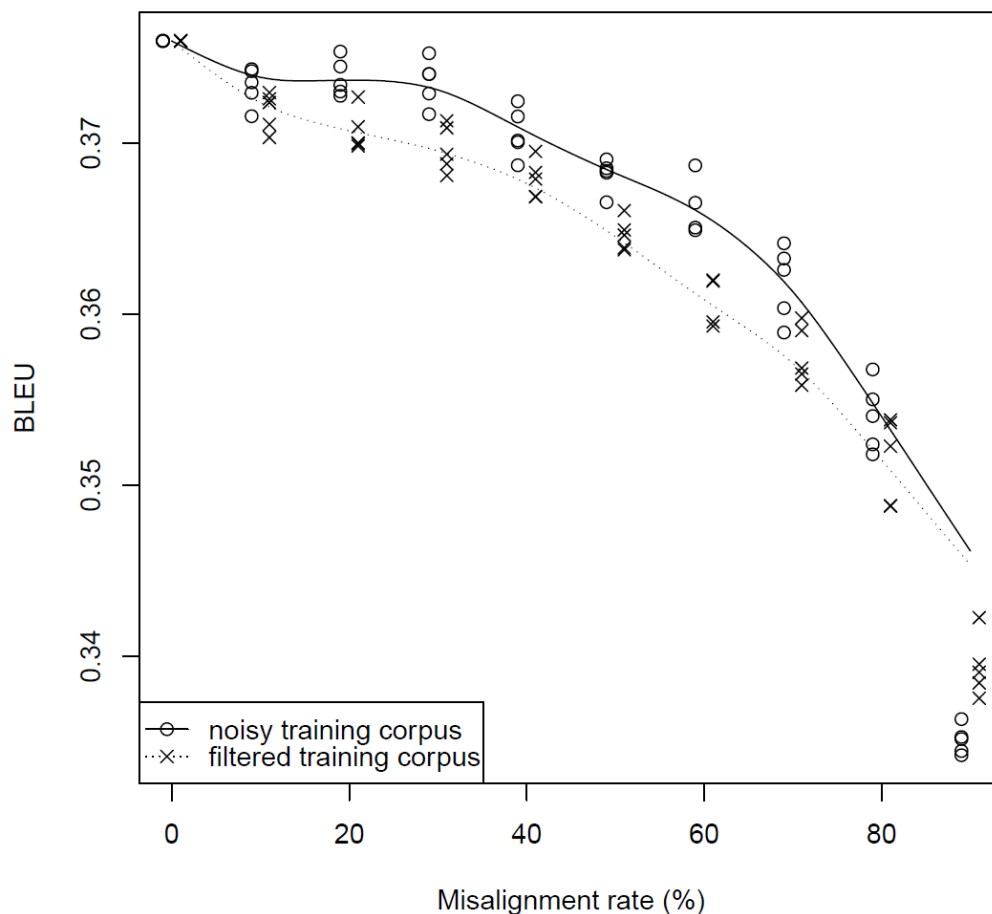| Global TM	Court	70%+	85%+	Exact
	CMAC	8.4	6.8	4.5
	FC	13.9	11.9	9.3
	FCA	16.5	14.8	10.8
	TCC	13.1	10.7	7.2

Processing CAS Data

A 3-step process:

1. Pair up documents
2. Extract text, segment (into translation units), normalize
3. Align segments

- Initially done using NRC tools
- Recently: *AlignFactory* (Terminotix)

Sentence (mis)alignment

Impact of noise on BLEU

- SMT highly tolerant to "noise" in alignment

[Cyril Goutte, Marine Carpuat, George Foster (2012).
The Impact of Sentence Alignment Errors on Phrase-Based Machine Translation Performance.

AMTA 2012]

True only when noise is "uniform"!

Sentence (mis)alignment

Court	Basic Alignment		Improved Alignment	
	Accuracy (%)		Accuracy (%)	
CMAC	89.5		96.5	
FC	89.0		93.5	
FCA	88.5		97.0	
TCC	90.0		99.5	

- To measure alignment accuracy: sample 100 random pairs (A,B), assign labels:

Label	Description	Accuracy
Good	A is a translation of B	1
Partial	Part of A is a translation of part of B	½
Bad	A not a translation of B	0
Unusable	Something is weird	0

Sentence (mis)alignment

Court	Basic Alignment		Improved Alignment	
	Accuracy (%)	MT (BLEU)	Accuracy (%)	MT (BLEU)
CMAC	89.5	40.4	96.5	**41.2**
FC	89.0	46.5	93.5	**49.3**
FCA	88.5	42.7	97.0	**47.1**
TCC	90.0	44.0	99.5	**47.1**

- Obviously, alignment errors are not "uniform"
  → systematic bias is hurting quality of MT
  → Better alignments mean Portage has more "meaningful" data to learn from

Data Filtering

- Untranslated quotation in text
  → same language in both versions

[5]   Dans sa décision, la Commission rappelle d'abord les circonstances ayant entouré le

double meurtre commis par le demandeur et ce qui a pu pousser celui-ci à les commettre,

circonstances qu'elle décrit de la manière suivante :

> The victims were 15 and 17 year-old adolescents. The youngest victim was the brother of your ex-girlfriend and the other victim was one of his friends. On February 28, 1989, you entered the residence of the youngest victim and hid in the basement with a loaded rifle. When the two boys arrived after school, you shot both of them in the head. Each victim was shot twice. They were murdered in cold blood, with planning and deliberation.
>
> […]

According to your file, those violent crimes were committed in

Data Filtering

- Bilingual quotation in both versions of text

[3] The Tax Court found that the lump sum payment was properly included in Mr. Butler's income under paragraph 56(1)(v) of the *Income Tax Act*, R.S.C. 1985, c. 1 (5th Supp.) [the *ITA*]. This paragraph provides:

56. (1) Without restricting the generality of section 3, there shall be included in computing the income of a taxpayer for a taxation year,	56. (1) Sans préjudice de la portée générale de l'article 3, sont à inclure dans le calcul du revenu d'un contribuable pour une année d'imposition :
[...]	[...]
(v) compensation received under an employees' or workers' compensation law of Canada or a province in respect of an injury, a disability or death;	v) une indemnité reçue en vertu d'une loi sur les accidents du travail du Canada ou d'une province à l'égard d'une blessure, d'une invalidité ou d'un décès;

[4] The Tax Court held that the term "in respect of an injury" is to be broadly interpreted to mean all amounts paid in relation to a compensable injury, relying on the

Data Filtering

- We applied a simple filter, based on short lists of frequent French and English words
- Filters out 1-5% of training data
- BLEU gains between 0.3 and 1.5

Court	Unfiltered	Filtered
CMAC	41.5	**41.8**
FC	49.0	**50.4**
FCA	47.4	**48.9**
TCC	47.3	**47.6**

System Combinations

- Portage allows different combination strategies

- Best results are obtained with *mixture models,* that assign different weights to each component, to optimize performance on a certain type of text
  - mixLM: mixture target language model
  - mixTM: mixture translation model ("phrasetable")

System Combinations

CMAC   FC   FCA   TCC
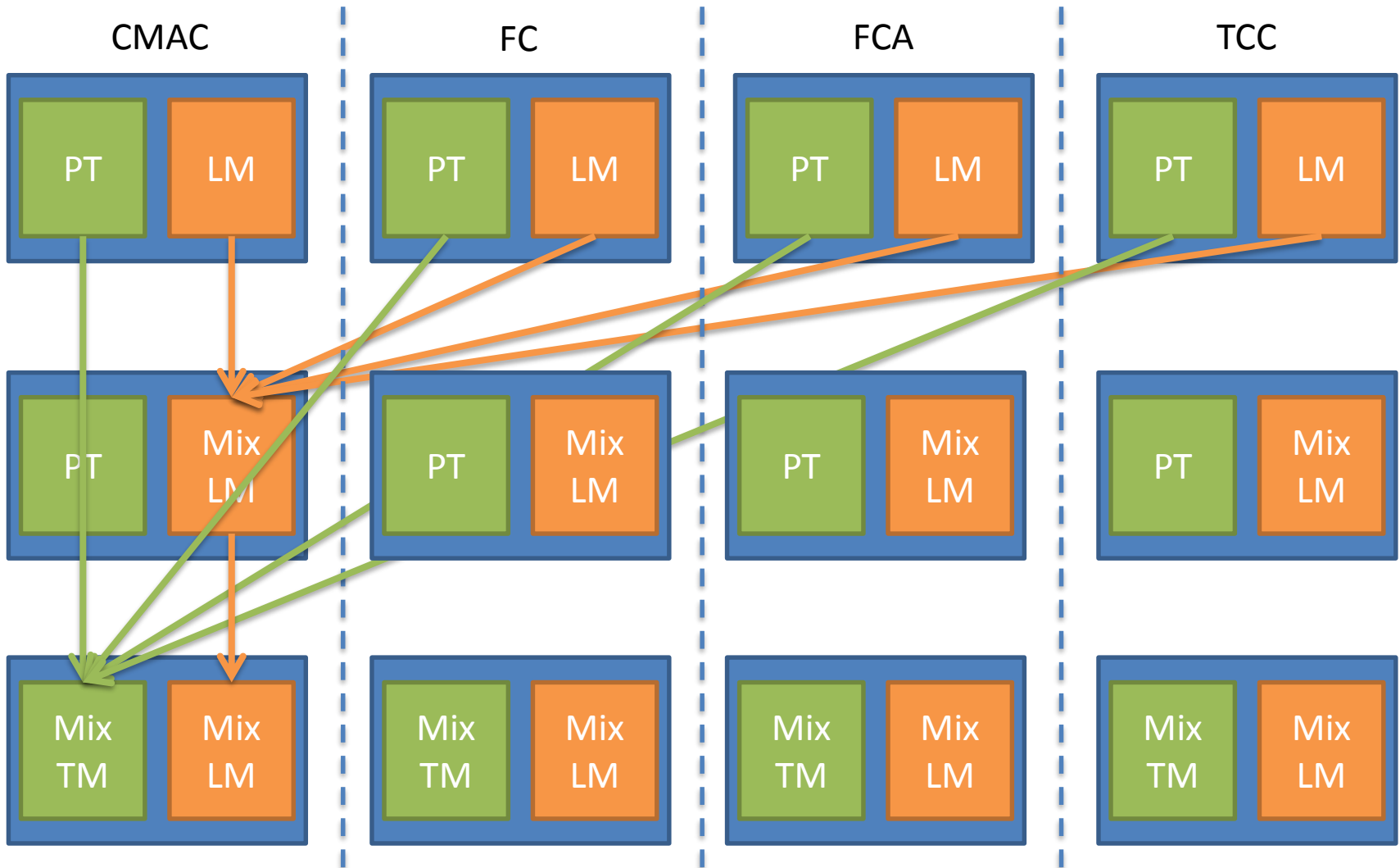
System Combinations

Court	Baseline	+ mixLM	+ mixLM + mixTM
CMAC	41.8	43.1	**45.4**
FC	**50.4**	50.0	50.3
FCA	48.9	49.5	**51.7**
TCC	47.6	47.6	**47.8**

- CMAC & FCA benefit the most from combinations
  - CMAC is small
  - FCA is very much related to matters in FC
- No clear benefit for FC and TCC
  - FC is much larger than other domains
  - TCC is probably distinct
- **Combinations never (significantly) hurt performance**

Final Systems

Court	EN→FR	FR→EN
CMAC	45.4	46.5
FC	50.3	52.6
FCA	51.7	54.4
TCC	47.8	52.0

- Gains relative to initial baseline systems range from **+3.8** (FC & TCC) to **+9.0** (FCA)

Integrating Portage in Matecat

- Needed a translation environment that allows MT to be integrated with translation memory
  - no TenT being used at CAS
- Matecat: a cloud-based CAT system
  - product of EU FP7 aimed at minimizing PE time
  - advantage for CAS: requires no local infrastructure or computer support; accessible everywhere
  - Matecat is free! Perfect for a pilot project
  - allows integration of different MT systems

Pilot Project at CAS

- To what extent can Portage help TRs increase productivity and decrease turnaround times?
- Two translation students hired for summer
  - pro: enthusiastic & open to technology
  - con: little experience in legal translation
- Translations carefully revised by professionals before publication
- Compare translation times with/without MT

Pilot Project Framework

- Focus on immigration decisions
- Statistics obtained from onsite coordinator
  - total no. texts/ words translated by each student
  - no. of texts with/without MT
  - productivity with/without MT
- Follow-up training provided to students
  - feedback obtained from two revisers
- Trial began on 11 May and ended on 25 August

The Results

Translator	No MT			With MT				
	total # texts	total # words	avg. # words/hr.	total # texts	total # words	avg. # words/hr.	diff. # words/hr	gain +MT vs. -MT
ADB	14	19,998	238	85	77685	373	135	57%
AL	19	20,538	291	109	86,918	390	99	34%

- Results compare very favorably with legal translators currently handling CAS decisions

Trial Results (cont'd)

- Caveats:
  - TRs didn't have access to a complete TM; only the one created as they translated. Some of gain attributed to MT would normally come from TM
  - We should have recorded revision times to ensure +TM texts didn't require more revision
- Still, no doubt that student TRs benefited substantially from Portage input
  - revisers report surprising errors in non-MT

Room for improvement

- Matecat had its problems:
  - handling intricate formatting; not always parallel in English & French
  - reintegrating results of spelling & grammar checking
  - lack of flexibility in revision mode
- Portage had its problems
  - handling named entities, i.e. knowing when and when not to translate these NPs
  - surprising number of errors of grammatical agreement, particularly in E > F direction

Discussion

- Recall: these students had no prior experience in legal translation
  - yet with the help of MT, in a few short months…
  - aided by Portage's acquisition of terms & phrases that are common in court decisions
- What makes these court decisions such a good application for MT
- Future plans

Thank you for your attention!

Any questions?