

Project adaptation over several days

Item Type	Presentation
Authors	Blain, Frederic;Hazem, Amir;Bougares, Fethi;Barrault, Loic;Schwenk, Holger
Citation	Blain, F., Hazem, A., Bougares, F., Barrault, L. and Schwenk, H. (2015) Project adaptation over several days, Translation in Transition 2015, 29th-30th January, 2015, Mainz University, Germany.
Publisher	Johannes Gutenberg University of Mainz
Download date	2026-06-18 12:15:09
License	https://creativecommons.org/licenses/by-nc-nd/4.0/
Link to Item	http://hdl.handle.net/2436/623578

Project Adaptation Over Several Days

Frédéric Blain, **Amir Hazem**, Fethi Bougares, Loic Barrault and
Holger Schwenk

LIUM, University of Le Mans, France

January 30, 2015



Outline

1 Introduction

Outline

- 1 Introduction
- 2 Related Work

Outline

- 1 Introduction
- 2 Related Work
- 3 Approach

Outline

- 1 Introduction
- 2 Related Work
- 3 Approach
- 4 Experiments and Results

Outline

- 1 Introduction
- 2 Related Work
- 3 Approach
- 4 Experiments and Results
- 5 Conclusion and Future Work

MateCat

MateCat is the result of a 3-year research project led by a consortium composed:

- The international research center FBK (Fondazione Bruno Kessler) (Italy)
- Translated srl (Italy)
- Université du Maine (LIUM) (France)
- University of Edinburgh (Scotland)



MateCat

MateCat

- an enterprise-level, web-based computer aided translation (CAT) tool

MateCat

- an enterprise-level, web-based computer aided translation (CAT) tool
- provides a complete set of features to manage and monitor translation projects

MateCat

- an enterprise-level, web-based computer aided translation (CAT) tool
- provides a complete set of features to manage and monitor translation projects
- supports and facilitates the translation process
 - ▶ boosts the productivity of professional translators
 - ▶ makes post-editing and outsourcing easy

MateCat

- an enterprise-level, web-based computer aided translation (CAT) tool
- provides a complete set of features to manage and monitor translation projects
- supports and facilitates the translation process
 - ▶ boosts the productivity of professional translators
 - ▶ makes post-editing and outsourcing easy
- improves the integration of machine translation and human translation within the CAT framework

CAT tools

- CAT tools are usually based on:

CAT tools

- CAT tools are usually based on:
 - ▶ translation memories (TM)
 - ▶ terminology dictionaries
 - ▶ concordancers and spell checkers
 - ▶ recently, Machine Translation (MT) systems

CAT tools

- CAT tools are usually based on:
 - ▶ translation memories (TM)
 - ▶ terminology dictionaries
 - ▶ concordancers and spell checkers
 - ▶ recently, Machine Translation (MT) systems
- If a segment to be translated is present in the TM:
 - ▶ display the corresponding translation through its editor —> translators can use it directly

CAT tools

- CAT tools are usually based on:
 - ▶ translation memories (TM)
 - ▶ terminology dictionaries
 - ▶ concordancers and spell checkers
 - ▶ recently, Machine Translation (MT) systems
- If a segment to be translated is present in the TM:
 - ▶ display the corresponding translation through its editor —> translators can use it directly
- if a segment to be translated is not present in the TM:
 - ▶ use the MT system to translate the segment

Motivations

- MT systems: not "yet" able to produce a suitable translation

Motivations

- MT systems: not "yet" able to produce a suitable translation
 - ▶ suitable \rightarrow comparable to a translation produced by a human translator

Motivations

- MT systems: not "yet" able to produce a suitable translation
 - ▶ suitable —> comparable to a translation produced by a human translator
- MT systems: can help translators to gain time

Motivations

- MT systems: not "yet" able to produce a suitable translation
 - ▶ suitable —> comparable to a translation produced by a human translator
- MT systems: can help translators to gain time
 - ▶ correct the MT output instead of translating the whole segment

Motivations

- MT systems: not "yet" able to produce a suitable translation
 - ▶ suitable —> comparable to a translation produced by a human translator
- MT systems: can help translators to gain time
 - ▶ correct the MT output instead of translating the whole segment
- MT systems: influence the translator

Motivations

- MT systems: not "yet" able to produce a suitable translation
 - ▶ suitable —> comparable to a translation produced by a human translator
- MT systems: can help translators to gain time
 - ▶ correct the MT output instead of translating the whole segment
- MT systems: influence the translator
- CAT tools: not "yet" completely satisfactory
 - ▶ most MT systems are not fully integrated with the human translation workflow

Goals

- make the MT system more specific to the project
- make the MT system fully integrated with the human translation workflow
- minimize the MT output errors throughout the duration of the project
- reduce as much as possible human intervention

Domain adaptation

Domain adaptation

Translate documents of specific domains using statistical machine translation (SMT) systems:

Domain adaptation

Translate documents of specific domains using statistical machine translation (SMT) systems:

- Medical domain
- Legal domain
- IT domain

Domain adaptation

Translate documents of specific domains using statistical machine translation (SMT) systems:

- Medical domain
- Legal domain
- IT domain

Motivation

Domain adaptation

Translate documents of specific domains using statistical machine translation (SMT) systems:

- Medical domain
- Legal domain
- IT domain

Motivation

- more data we have, better is the performance of SMT systems

Domain adaptation

Translate documents of specific domains using statistical machine translation (SMT) systems:

- Medical domain
- Legal domain
- IT domain

Motivation

- more data we have, better is the performance of SMT systems
- if a significant out-of-domain data is added to the training corpus, translation quality can drop [Wang et al., 2014]
 - ▶ more data \longrightarrow better model, no longer valid when dealing with specific domains

Introduction

Solution \rightarrow Data Selection

Introduction

Solution \rightarrow Data Selection

- in-domain data can be enriched with suitable sub-parts of out-of-domain data

Introduction

Solution —> Data Selection

- in-domain data can be enriched with suitable sub-parts of out-of-domain data
- out-of-domain sub-parts are considered to be close enough to the specific domain

Related Work

Related Work

- [Eck et al., 2004] introduced the first language model (LM) adaptation for machine translation

Related Work

- [Eck et al., 2004] introduced the first language model (LM) adaptation for machine translation
 - ▶ translate each test document using the general domain model
 - ▶ select most similar documents using the resulting translations (IR techniques)
 - ▶ build the adapted language model according to the selected documents

Related Work

- [Eck et al., 2004] introduced the first language model (LM) adaptation for machine translation
 - ▶ translate each test document using the general domain model
 - ▶ select most similar documents using the resulting translations (IR techniques)
 - ▶ build the adapted language model according to the selected documents
- [Hildebrand et al., 2005] applied IR techniques for translation model (TM) adaptation

Related Work

- [Eck et al., 2004] introduced the first language model (LM) adaptation for machine translation
 - ▶ translate each test document using the general domain model
 - ▶ select most similar documents using the resulting translations (IR techniques)
 - ▶ build the adapted language model according to the selected documents
- [Hildebrand et al., 2005] applied IR techniques for translation model (TM) adaptation
- [Moore and Lewis, 2010] compared the cross-entropy according to in-domain and out-of-domain data (cross-entropy as a ranking function)

Related Work

- [Eck et al., 2004] introduced the first language model (LM) adaptation for machine translation
 - ▶ translate each test document using the general domain model
 - ▶ select most similar documents using the resulting translations (IR techniques)
 - ▶ build the adapted language model according to the selected documents
- [Hildebrand et al., 2005] applied IR techniques for translation model (TM) adaptation
- [Moore and Lewis, 2010] compared the cross-entropy according to in-domain and out-of-domain data (cross-entropy as a ranking function)
- [Axelrod et al., 2011] extended the cross-entropy difference to the bilingual level

Related Work

- [Eck et al., 2004] introduced the first language model (LM) adaptation for machine translation
 - ▶ translate each test document using the general domain model
 - ▶ select most similar documents using the resulting translations (IR techniques)
 - ▶ build the adapted language model according to the selected documents
- [Hildebrand et al., 2005] applied IR techniques for translation model (TM) adaptation
- [Moore and Lewis, 2010] compared the cross-entropy according to in-domain and out-of-domain data (cross-entropy as a ranking function)
- [Axelrod et al., 2011] extended the cross-entropy difference to the bilingual level
- [Wang et al., 2014] combined the cosine tf-idf approach with the perplexity-based and the edit distance-based approaches

Approach

- post-editing MT output —> increases the productivity of human translators [Guerberof, 2009, Plitt and Masselot, 2010, Federico et al., 2012, Green et al., 2013]

Approach

- post-editing MT output —> increases the productivity of human translators [Guerberof, 2009, Plitt and Masselot, 2010, Federico et al., 2012, Green et al., 2013]
- specializing the MT system on the documents to be translated is relatively new [Cettolo et al., 2014]

Approach

- post-editing MT output —> increases the productivity of human translators [Guerberof, 2009, Plitt and Masselot, 2010, Federico et al., 2012, Green et al., 2013]
- specializing the MT system on the documents to be translated is relatively new [Cettolo et al., 2014]
- domain adaptation (DA) only allows an MT system to be specific to a particular domain but not to a particular project

Approach

- post-editing MT output → increases the productivity of human translators [Guerberof, 2009, Plitt and Masselot, 2010, Federico et al., 2012, Green et al., 2013]
- specializing the MT system on the documents to be translated is relatively new [Cettolo et al., 2014]
- domain adaptation (DA) only allows an MT system to be specific to a particular domain but not to a particular project
 - ▶ purpose of this study → show that corrections performed by translators over one day of work are an important and a valuable resource to improve the MT system for the next day

Domain adapted system (Moses)

Domain adapted system (Moses)

- translation model (TM)

Domain adapted system (Moses)

- translation model (TM)
 - ▶ in-domain data (dgtna corpus)

Domain adapted system (Moses)

- translation model (TM)
 - ▶ in-domain data (dgtna corpus)
 - ▶ pseudo in-domain data —> extracted from the out-of-domain data

Domain adapted system (Moses)

- translation model (TM)
 - ▶ in-domain data (dgtna corpus)
 - ▶ pseudo in-domain data —> extracted from the out-of-domain data
 - ★ perform bilingual cross-entropy difference [Axelrod et al., 2011]
 - ★ select suitable sub-parts (EP7, NC7, Acquis, UN2000, IT, TM and KDE)

Domain adapted system (Moses)

- translation model (TM)
 - ▶ in-domain data (dgtna corpus)
 - ▶ pseudo in-domain data —> extracted from the out-of-domain data
 - ★ perform bilingual cross-entropy difference [Axelrod et al., 2011]
 - ★ select suitable sub-parts (EP7, NC7, Acquis, UN2000, IT, TM and KDE)
- language model (LM)

Domain adapted system (Moses)

- translation model (TM)
 - ▶ in-domain data (dgtna corpus)
 - ▶ pseudo in-domain data —> extracted from the out-of-domain data
 - ★ perform bilingual cross-entropy difference [Axelrod et al., 2011]
 - ★ select suitable sub-parts (EP7, NC7, Acquis, UN2000, IT, TM and KDE)
- language model (LM)
 - ▶ perform a monolingual cross-entropy difference [Moore and Lewis, 2010]

Domain adapted system (Moses)

- translation model (TM)
 - ▶ in-domain data (dgtna corpus)
 - ▶ pseudo in-domain data —> extracted from the out-of-domain data
 - ★ perform bilingual cross-entropy difference [Axelrod et al., 2011]
 - ★ select suitable sub-parts (EP7, NC7, Acquis, UN2000, IT, TM and KDE)
- language model (LM)
 - ▶ perform a monolingual cross-entropy difference [Moore and Lewis, 2010]
 - ▶ select suitable sub-parts (EP7, NC7, Acquis, UN2000, IT, TM and KDE)

Domain adapted system (Moses)

- translation model (TM)
 - ▶ in-domain data (dgtna corpus)
 - ▶ pseudo in-domain data → extracted from the out-of-domain data
 - ★ perform bilingual cross-entropy difference [Axelrod et al., 2011]
 - ★ select suitable sub-parts (EP7, NC7, Acquis, UN2000, IT, TM and KDE)
- language model (LM)
 - ▶ perform a monolingual cross-entropy difference [Moore and Lewis, 2010]
 - ▶ select suitable sub-parts (EP7, NC7, Acquis, UN2000, IT, TM and KDE)
 - ▶ add French newspapers data from 2007 to 2013 (news2007, news2008, news2009, news2010, news2011, news2012, news2013)

Domain adapted system (Moses)

- translation model (TM)
 - ▶ in-domain data (dgtna corpus)
 - ▶ pseudo in-domain data → extracted from the out-of-domain data
 - ★ perform bilingual cross-entropy difference [Axelrod et al., 2011]
 - ★ select suitable sub-parts (EP7, NC7, Acquis, UN2000, IT, TM and KDE)
- language model (LM)
 - ▶ perform a monolingual cross-entropy difference [Moore and Lewis, 2010]
 - ▶ select suitable sub-parts (EP7, NC7, Acquis, UN2000, IT, TM and KDE)
 - ▶ add French newspapers data from 2007 to 2013 (news2007, news2008, news2009, news2010, news2011, news2012, news2013)
 - ▶ interpolate the final language model

Domain adapted system (Moses)

- translation model (TM)
 - ▶ in-domain data (dgtna corpus)
 - ▶ pseudo in-domain data —> extracted from the out-of-domain data
 - ★ perform bilingual cross-entropy difference [Axelrod et al., 2011]
 - ★ select suitable sub-parts (EP7, NC7, Acquis, UN2000, IT, TM and KDE)
- language model (LM)
 - ▶ perform a monolingual cross-entropy difference [Moore and Lewis, 2010]
 - ▶ select suitable sub-parts (EP7, NC7, Acquis, UN2000, IT, TM and KDE)
 - ▶ add French newspapers data from 2007 to 2013 (news2007, news2008, news2009, news2010, news2011, news2012, news2013)
 - ▶ interpolate the final language model
- train the SMT system

Project adapted system (PA)

- a classical translation scenario -> translate a set of documents over several days assisted by a given CAT tool

Project adapted system (PA)

- a classical translation scenario -> translate a set of documents over several days assisted by a given CAT tool
- after the first working day -> knowledge about the new translated text is injected into the SMT system

Project adapted system (PA)

- a classical translation scenario -> translate a set of documents over several days assisted by a given CAT tool
- after the first working day -> knowledge about the new translated text is injected into the SMT system
 - ▶ add the newly translated text of the current day into the development set

Project adapted system (PA)

- a classical translation scenario -> translate a set of documents over several days assisted by a given CAT tool
- after the first working day -> knowledge about the new translated text is injected into the SMT system
 - ▶ add the newly translated text of the current day into the development set
 - ▶ perform a new data selection based on this new development set

Project adapted system (PA)

- a classical translation scenario -> translate a set of documents over several days assisted by a given CAT tool
- after the first working day -> knowledge about the new translated text is injected into the SMT system
 - ▶ add the newly translated text of the current day into the development set
 - ▶ perform a new data selection based on this new development set
 - ▶ retrain the system with the new selected data

Project adapted system (PA)

- a classical translation scenario -> translate a set of documents over several days assisted by a given CAT tool
- after the first working day -> knowledge about the new translated text is injected into the SMT system
 - ▶ add the newly translated text of the current day into the development set
 - ▶ perform a new data selection based on this new development set
 - ▶ retrain the system with the new selected data
- repeat the process throughout the duration of the translation project (5 days)

Project adapted system (PA)

- a classical translation scenario -> translate a set of documents over several days assisted by a given CAT tool
- after the first working day -> knowledge about the new translated text is injected into the SMT system
 - ▶ add the newly translated text of the current day into the development set
 - ▶ perform a new data selection based on this new development set
 - ▶ retrain the system with the new selected data
- repeat the process throughout the duration of the translation project (5 days)
 - ▶ allow the system to adapt and to learn from its errors

Experiments and Results

Experiments and Results

- English-French corpus from the Legal domain

Experiments and Results

- English-French corpus from the Legal domain
- 3 translators

Experiments and Results

- English-French corpus from the Legal domain
- 3 translators
- development set of 910 sentences

Experiments and Results

- English-French corpus from the Legal domain
- 3 translators
- development set of 910 sentences
- test set of about 150 sentences/day (3000 words/day)

Experiments and Results

Train	Document		
	segments	tokens(en)	tokens(fr)
Nc7	183K	4.65M	5.68M
EP7	2M	55.7M	61.8M
Giga	10.8M	291.8M	353.4M
Un2000	12.9M	361.9M	421.7M
Dgtna	728K	18.6M	20.6M
JRC acquis	2.7M	64.2M	70.3M
TM	150K	2.8M	3.2M
IT	77.5K	0.8M	0.9M
Kde	207K	1.9M	2.2M

Table : Data set size.

Table : BLEU scores for 5 days on the English-French data of the Legal domain for translators 1, 2 and 3. The BLEU score represented between brackets is calculated using the generic reference.

	Translator 1	Translator 2	Translator 3
DA1	49.72 [24.89]	48.84 [24.89]	30.23 [24.89]

Table : BLEU scores for 5 days on the English-French data of the Legal domain for translators 1, 2 and 3. The BLEU score represented between brackets is calculated using the generic reference.

	Translator 1	Translator 2	Translator 3
DA1	49.72 [24.89]	48.84 [24.89]	30.23 [24.89]
DA2	48.35 [22.78]	44.07 [22.78]	30.68 [22.78]
PA2	49.04 [23.66]	47.23 [23.39]	32.29 [23.56]

Table : BLEU scores for 5 days on the English-French data of the Legal domain for translators 1, 2 and 3. The BLEU score represented between brackets is calculated using the generic reference.

	Translator 1	Translator 2	Translator 3
DA1	49.72 [24.89]	48.84 [24.89]	30.23 [24.89]
DA2	48.35 [22.78]	44.07 [22.78]	30.68 [22.78]
PA2	49.04 [23.66]	47.23 [23.39]	32.29 [23.56]
DA3	53.75 [24.16]	46.88 [24.16]	44.16 [24.16]
PA3	59.18 [26.35]	54.46 [27.17]	51.80 [27.15]

Table : BLEU scores for 5 days on the English-French data of the Legal domain for translators 1, 2 and 3. The BLEU score represented between brackets is calculated using the generic reference.

	Translator 1	Translator 2	Translator 3
DA1	49.72 [24.89]	48.84 [24.89]	30.23 [24.89]
DA2	48.35 [22.78]	44.07 [22.78]	30.68 [22.78]
PA2	49.04 [23.66]	47.23 [23.39]	32.29 [23.56]
DA3	53.75 [24.16]	46.88 [24.16]	44.16 [24.16]
PA3	59.18 [26.35]	54.46 [27.17]	51.80 [27.15]
DA4	51.48 [25.01]	43.22 [25.01]	39.77 [25.01]
PA4	52.48 [26.87]	49.59 [25.90]	47.50 [27.22]

Table : BLEU scores for 5 days on the English-French data of the Legal domain for translators 1, 2 and 3. The BLEU score represented between brackets is calculated using the generic reference.

	Translator 1	Translator 2	Translator 3
DA1	49.72 [24.89]	48.84 [24.89]	30.23 [24.89]
DA2	48.35 [22.78]	44.07 [22.78]	30.68 [22.78]
PA2	49.04 [23.66]	47.23 [23.39]	32.29 [23.56]
DA3	53.75 [24.16]	46.88 [24.16]	44.16 [24.16]
PA3	59.18 [26.35]	54.46 [27.17]	51.80 [27.15]
DA4	51.48 [25.01]	43.22 [25.01]	39.77 [25.01]
PA4	52.48 [26.87]	49.59 [25.90]	47.50 [27.22]
DA5	53.30 [24.78]	47.77 [24.78]	42.18 [24.78]
PA5	54.41 [27.14]	56.13 [26.96]	50.18 [27.15]

Table : TER scores for 5 days on the English-French data of the Legal domain for translators 1, 2 and 3. The TER between brackets is calculated using the generic reference.

	Translator 1	Translator 2	Translator 3
DA1	33.34 [54.59]	32.99 [54.59]	48.62 [54.59]

Table : TER scores for 5 days on the English-French data of the Legal domain for translators 1, 2 and 3. The TER between brackets is calculated using the generic reference.

	Translator 1	Translator 2	Translator 3
DA1	33.34 [54.59]	32.99 [54.59]	48.62 [54.59]
DA2	35.33 [56.63]	37.44 [56.63]	49.03 [56.63]
PA2	34.59 [55.97]	34.42 [56.34]	48.70 [56.26]

Table : TER scores for 5 days on the English-French data of the Legal domain for translators 1, 2 and 3. The TER between brackets is calculated using the generic reference.

	Translator 1	Translator 2	Translator 3
DA1	33.34 [54.59]	32.99 [54.59]	48.62 [54.59]
DA2	35.33 [56.63]	37.44 [56.63]	49.03 [56.63]
PA2	34.59 [55.97]	34.42 [56.34]	48.70 [56.26]
DA3	30.76 [55.49]	35.09 [55.49]	38.05 [55.49]
PA3	27.24 [53.54]	30.28 [53.37]	33.09 [53.41]

Table : TER scores for 5 days on the English-French data of the Legal domain for translators 1, 2 and 3. The TER between brackets is calculated using the generic reference.

	Translator 1	Translator 2	Translator 3
DA1	33.34 [54.59]	32.99 [54.59]	48.62 [54.59]
DA2	35.33 [56.63]	37.44 [56.63]	49.03 [56.63]
PA2	34.59 [55.97]	34.42 [56.34]	48.70 [56.26]
DA3	30.76 [55.49]	35.09 [55.49]	38.05 [55.49]
PA3	27.24 [53.54]	30.28 [53.37]	33.09 [53.41]
DA4	33.01 [55.90]	38.31 [55.90]	41.96 [55.90]
PA4	31.98 [54.27]	33.85 [55.41]	36.49 [55.07]

Table : TER scores for 5 days on the English-French data of the Legal domain for translators 1, 2 and 3. The TER between brackets is calculated using the generic reference.

	Translator 1	Translator 2	Translator 3
DA1	33.34 [54.59]	32.99 [54.59]	48.62 [54.59]
DA2	35.33 [56.63]	37.44 [56.63]	49.03 [56.63]
PA2	34.59 [55.97]	34.42 [56.34]	48.70 [56.26]
DA3	30.76 [55.49]	35.09 [55.49]	38.05 [55.49]
PA3	27.24 [53.54]	30.28 [53.37]	33.09 [53.41]
DA4	33.01 [55.90]	38.31 [55.90]	41.96 [55.90]
PA4	31.98 [54.27]	33.85 [55.41]	36.49 [55.07]
DA5	31.34 [54.78]	34.38 [54.78]	39.41 [54.78]
PA5	31.24 [52.46]	29.03 [53.37]	34.67 [53.86]

Conclusion and Future Work

Conclusion and Future Work

- Proposition of a new project adaptation method over several days

Conclusion and Future Work

- Proposition of a new project adaptation method over several days
- Evaluation with 3 translators
 - ▶ PA translations —> different corrections according to each translator

Conclusion and Future Work

- Proposition of a new project adaptation method over several days
- Evaluation with 3 translators
 - ▶ PA translations —> different corrections according to each translator
- Encouraging results

Conclusion and Future Work

- Proposition of a new project adaptation method over several days
- Evaluation with 3 translators
 - ▶ PA translations —> different corrections according to each translator
- Encouraging results
- In the future —> New projects and collaborations with Linguists and Translators?

Thank you for your attention

Thank you for your attention

 Axelrod, A., He, X., and Gao, J. (2011).

Domain adaptation via pseudo in-domain data selection.

In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, 27-31 July 2011, John McIntyre Conference Centre, Edinburgh, UK, A meeting of SIGDAT, a Special Interest Group of the ACL, pages 355–362.

 Cettolo, M., Bertoldi, N., Federico, M., Schwenk, H., Barrault, L., and Servan, C. (2014).

Translation project adaptation for mt-enhanced computer assisted translation.

Machine Translation, (28):127–150.

 Eck, M., Vogel, S., and Waibel, A. (2004).

Language model adaptation for statistical machine translation based on information retrieval.

In In Proc. of LREC.

 Federico, M., Cattelan, A., and Marco, T. (2012).

Measuring user productivity in machine translation enhanced computer assisted translation.

In Proceedings of the Tenth Conference of the Association for Machine Translation in the Americas (AMTA).



Green, S., Jeffrey, H., and Christopher D, M. (2013).
The efficacy of human post-editing for language translation.
In ACM Human Factors in Computing Systems (CHI).



Guerberof, A. A. (2009).
Productivity and quality in the post-editing of outputs from translation memories and machine translation.



Hildebrand, A. S., Eck, M., Vogel, S., and Waibel, A. (2005).
Adaptation of the translation model for statistical machine translation based on information retrieval.
In Proceedings of EAMT, volume 2005, pages 133–142.



Moore, R. C. and Lewis, W. D. (2010).
Intelligent selection of language model training data.

In *ACL 2010, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden, Short Papers*, pages 220–224.



Plitt, M. and Masselot, F. (2010).

A productivity test of statistical machine translation post-editing in a typical localisation context.

Prague Bull. Math. Linguistics, 93:7–16.



Wang, L., Wong, D. F., Chao, L. S., Lu, Y., and Xing, J. (2014).

A systematic comparison of data selection criteria for smt domain adaptation.

The Scientific World Journal, 2014:10.