

# **A tool for automatic evaluation of human translation quality within a MOOC environment**



**Miguel Betanzos Atienza**

Supervisor: Lluís A. Belanche Muñoz

Co-supervisor: Marta R. Costa-jussà

*Master in Artificial Intelligence*

October 2015



## **Acknowledgements**

And I would like to thank Lluís A. Belanche Muñoz, Marta R. Costa-jussà and Manuel C. Feria García, without whom, this work would not exist.



## **Abstract**

The present work describes the process of development of a tool intended to evaluate translation quality within an Edx course. The document consists of two main parts, one related to creation and management of a course to obtain translation data, and one related to the analysis of the collected data, and the possibilities of using it as the building block for an evaluation tool.



# Contents

<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xi</b>
<b>1 Introduction and motivation</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Project motivation . . . . .	2
1.3 Project Overview . . . . .	3
1.3.1 Managing the MOOC platform and course . . . . .	3
1.3.2 Data analysis and the quality estimation tool . . . . .	4
<b>2 State of the art</b>	<b>5</b>
2.1 On MOOCs and open EdX . . . . .	5
2.2 On Translation quality estimation . . . . .	6
2.3 Automated essay scoring . . . . .	6
<b>3 Working with the open EdX platform</b>	<b>7</b>
3.1 Installing, Configuring, and Running the Open edX Platform . . . . .	7
3.1.1 Getting an open EdX instance to work. . . . .	8
3.2 Course Design . . . . .	9
3.2.1 Promotion . . . . .	9
3.2.2 Exercises . . . . .	11
3.2.3 Peer Review . . . . .	11
3.2.4 Course results . . . . .	12
<b>4 Translation evaluation processes</b>	<b>15</b>
4.1 Evaluation as a prediction / classification task . . . . .	15
4.2 The evaluation tool . . . . .	15
4.2.1 Results . . . . .	16

---

<b>5</b>	<b>Conclusions</b>	<b>21</b>
5.1	The course . . . . .	21
5.2	The translation quality assessment tool . . . . .	22
	<b>Bibliography</b>	<b>23</b>
	<b>Appendix A Translation texts</b>	<b>25</b>
	<b>Appendix B Dataset information</b>	<b>31</b>



# List of Figures

- 3.1 Peer review scheme . . . . . 13
- 4.1 Confusion matrix. Text 1 Multiclass . . . . . 18
- 4.2 Confusion matrix. Text 2 Multiclass . . . . . 18
- 4.3 Confusion matrix. Text 1 Binary . . . . . 19
- 4.4 Confusion matrix. Text 2 Binary . . . . . 19



# List of Tables

3.1	No. of visits to the pre-inscription page by country . . . . .	9
3.2	Submissions for the first exercise of each text and for each translation . . . .	11
4.1	Choice of 10 best features for the classification of text 1 translations into 4 classes (left columns), and into two classes (right columns). . . . .	20
4.2	Choice of 10 best features for the classification of text 2 translations into 4 classes (left columns), and into two classes (right columns). . . . .	20
B.1	Ordered list of the features extracted from the translations . . . . .	33



# Chapter 1

## Introduction and motivation

### 1.1 Introduction.

The explosion of the MOOC phenomenon, in the year 2012[9], with the arrival of the big names like Coursera, EdX or Udacity, that was soon followed by many other MOOC providers, might well be considered the biggest innovation in education of our time. The figures are indeed big. By 2013 the number of students who had enrolled in a MOOC was in the millions, thousands of courses had been offered, and hundreds of universities were offering their courses in this format[3]. But as of today, browsing through MOOC catalogs, we can see that very few of these courses are offered on the subject of natural languages. This is surprising, since this field could possibly generate a strong interest.

In this project, we describe the process of development of a MOOC in the topic of human translation for the Arabic - Spanish language pair, aiming to reach some conclusions about the possibility of evaluating the translations in an automated way, using the material submitted by the participants.

The course offers participants collaborative learning; they receive evaluations and suggestions from other participants, and they analyse the mistakes and successes in the translations of other participants, following a specific rubric. It is a practical way to learn and reflect on the mechanisms of translation. The texts are real and current, ranked from least to most difficult, and again divided into areas or themes. Additionally, the course contains some exercises and support materials, and there is the possibility of discussing related topics in the forum.

In the final step, we compile a corpus using the translations obtained during the course, with the intention of building a tool able to perform automatic evaluation of the quality of new translations. This tool relies on several linguistic features extracted from the translation corpus, and the evaluations provided from the participants.

## 1.2 Project motivation

One of the main issues with MOOCs and in general, with learning without a human teacher, is the difficulty of getting the feedback that allows us to know how well we are doing and thus to keep working in a direction or change it. Now, while it is difficult to provide feedback to MOOC students, this feedback is given instantly most of the time, which in turn can be seen as an advantage, since the “changes in direction” can be quickly made one after another; that is, a student can propose an answer to a certain question, obtain feedback about it in an immediate manner, and if the answer can be improved, the student would be able to straightforwardly work in a new answer and then submit it, starting the cycle again[15].

Regarding feedback on natural language related tasks, if we intended to give feedback for written and oral production, and for written and oral understanding, it would seem that human intervention is mandatory.

Understanding is typically checked with some questions, as seen in most school textbooks; after presenting the student with a subject, the next step is usually to ask her several questions addressing the main points of said subject. The answers to these questions could be computer evaluated, specially if the questions are asked in a way that keeps answers simple and direct.

But on the other hand, written and oral production are much harder to evaluate. It seems that without the ability to understand meaning, there is no way that we can go past a mere checking of grammar and spelling. Nevertheless, there are several automatic essay scoring tools that grade student essays with enough accuracy as to be helpful to teachers. There are also ongoing studies on automatic assessment of oral language proficiency, though they are more oriented to assess the speaker’s pronunciation than to assess other aspects of the speech.

Now, since acquiring a second language is a basic step in the education of every person nowadays, creating MOOCs devoted to teach a certain language to foreigners could prove very interesting.

The main idea behind the present work is to test whether a tool that provides automatic evaluation of translation quality could be successfully implemented within a MOOC. In the process of learning a new language, the ability to get immediate feedback on a given translation could be a basic block, much needed to develop courses on second language acquisition, or even courses on translation like the one set up for this work. It must be noted though, that in order to provide immediate feedback, computational overhead has to be kept under control, which is a restraint.

But what is more, a MOOC using an automatic tool for translation quality evaluation could gather huge amounts of significant data in the field of translation studies. This means that a loop could be established, where new translations could help refine the tool, and also open paths to explore translation in unprecedented ways. Parallel corpora could be easily gathered, with a focus on the subjects chosen by the course creator. Corpora of translations made by both native and non native speakers could be gathered, and also corpora of translations made by learners in different stages of learning, which could help identify common mistakes and help create new language learning content, usable both in online and face to face environments.

## **1.3 Project Overview**

This project consists of two well differentiated parts:

### **1.3.1 Managing the MOOC platform and course**

#### **Setting up open EdX platform**

This step involves the installation, configuration and running of the open source EdX platform, able to host massive open online courses, on a server, providing access to participants.

#### **Setting up up the course**

This step includes the design, creation, promotion and management of the course throughout the time is open for participation. While the main purpose of the course is not teaching translation at the moment, several exercises are included for each translation task. This could be seen as a way to assure that participants get something back for their collaboration.

### **1.3.2 Data analysis and the quality estimation tool**

#### **Data collection**

From each proposed text in Arabic, we collect the translations into Spanish submitted by participants. In addition, each of these translations has three evaluations given by peers.

#### **Feature Extraction**

From each translation, a number of different linguistic measures is extracted into an array of numerical data linked together with the corresponding evaluations, which work as the label.

#### **Building the tool for automatic evaluation**

An automatic evaluator is trained, using the feature arrays and the evaluations provided. The model it creates is used to predict evaluations for new translations.



# Chapter 2

## State of the art

We describe in this section the current trends on MOOCs and translation evaluation.

### 2.1 On MOOCs and open EdX

Massive Open Online Courses are web-based courses that allow anyone with an internet connection to enroll, because they are open, and have no-cost, and they have no maximum enrollment limits. They contain all the content or references required for the course freely available ; and they have very low instructor involvement from a student perspective after the course begins[1].

These days MOOCs are a not a novelty anymore, they have become a staple resource for students all around the world. Many just enroll them to access a certain resource or out of curiosity. That is why many research is being conducted analysing their large dropping rates [8].

As of today, the trend in MOOCs is offering series of courses to provide a more structured formation that could be regarded as more valuable, and thus more worthy of paying for it. Universities are offering official credits for certified course completion, and while they are still open and free, it seems that the time to monetize the field has arrived.

On the other hand, the open EdX initiative is completely open source and can be adopted by any education institution willing to do so. for instance, Catalanian universities have created <http://ucatx.cat/>, but the tendency is to join bigger platforms, as EdX itself, and offer courses through a centralized big platform instead of deploying (and managing) a dedicated

platform.

## 2.2 On Translation quality estimation

Automatic evaluation of translation quality is mostly used to assess the output of Machine Translation systems. It is indeed very necessary to improve the work of such systems, which can produce enormous quantities of translations that could not possibly be all assessed by human experts. Or as put by M. Snover[13], "Machine Translation has proven a difficult task to evaluate. Human judgments of evaluation are expensive and noisy".

Translation evaluation could be categorized into two main branches, one that needs reference translations to produce the assessment, and another that only uses the source test. The one that uses only source test is necessarily bounded to a classification or prediction model, which could, in a way, be taken as reference translations.

Recent work on the topic involves mainly finding the most informative features to extract from the text but also, since there are many of those features, feature selection algorithms are being refined more and more[12], because different combination of features can yield different results. We will see that extraction and processing of features makes evaluation systems quite slow.

## 2.3 Automated essay scoring

Automated essay scoring is very related to translation assessment, since translations must also complain with many constraint very similar to essays. Automated essay scoring is at the heart of the hypothesis that motivated this project. Specifically, AES employs annotated data, and employs peer review assessment. But specially promising is the notion that automated essay scoring tools benefit from texts rated by multiple human raters and texts of significantly varying quality [1]. This is what we have attempted to throw into the mix with this project.

# Chapter 3

## Working with the open EdX platform

### 3.1 Installing, Configuring, and Running the Open edX Platform

There are two possible installations of the EdX open source[Edx], the DevStack and the FullStack. Both have the same system requirements, but Devstack simplifies certain production settings to make development more convenient. For example, the unicorn or nginx server options, able to give support to ten thousand simultaneous connections, are disabled in Devstack; which uses Django runserver instead.

In this project the FullStack was used, which is definitely overkill, taking into account that the number of users registered into the platform has been 130.

The services included in the open EdX full stack are:

1. edx-platform: The platform allows user registration, and is able to handle thousands of users and to provide hundreds of courses, from different course creators.
2. configuration: This is the tool to manage the platform, it controls registered users, passwords, course calendar setting, etc.
3. cs-comments-service: This is the discussion forum service that can be set up together with any course, in order to allow participant interaction.
4. notifier: This is to handle the bulk email, the registration verification via email, and allows to send reminders to enrolled participants.

5. `edx-certificates`: This creates a certificate of completion for a given course, that is granted to participants that fulfill some requisites of course participation or completion, or have passed specific exams to achieve it. Participants are usually motivated by it, and in this course they were actually asking for some kind of certificate, even if it was of no value at all. But no certificates were issued in this course.
6. `xqueue`: Allows the use of external grading services.
7. `edx-documentation`: Exhaustive information on how to set up the open EdX platform and how to build and run open EdX courses.
8. `edx-ora2`: This allows the implementation of different exercise types, and it gives capability for using the peer review exercise which has been central for this project
9. `XBlock`: The structure needed to create new modules for the EdX platform.

### **3.1.1 Getting an open EdX instance to work.**

There are several methods to serve open EdX to the public. It can be done on a virtual server, on a Ubuntu 12.04 dedicated server, or by means of images that are stored on the cloud. Only recently these images have become available, and they solve most of the problems of configuration that we can find during the process of installing open EdX.

The open EdX stable release used for this work has been Birch. As of October of 2015, the latest stable release is one after Birch, named Cypress. We chose a Birch image served by the Spanish based company Bitnami, through the Amazon Web Services. It is stored in a t2 medium layer which has a cost of approximately 45€ per month, and has been more than enough to deal with the 130 enrolled students. This means that it is relatively inexpensive to provide courses for that number of participants.

Problems encountered during the installation come mainly from updates to the stable release. Updates are applied from git repositories, and they can affect the functionality of the platform since they often need special dependencies that also need special dependencies and it is quite easy to end up with a non working instance.

Common issues include configuring the mail verification, localization of the platform (for this project the South American Spanish localization was used, since the translation in peninsular Spanish is not finished yet. Bulk email sending is another issue, because the

Table 3.1 No. of visits to the pre-inscription page by country

Country	No. of visits
Spain	838
Egypt	273
Morocco	141
Tunis	40
Algeria	31
France	20
Switzerland	14
UK	13
USA	11
Oman	11
Jordan	11

instance can become listed as spam server. Finally, once the platform has been properly configured, courses can be created and provided to any registered user.

## 3.2 Course Design

### 3.2.1 Promotion

The first requisite for having a course running is that there are students taking it. Promotion of the course is indeed a matter to be taken into account, and in this case, having no students would have meant having no material to work with. Open EdX recommends using a short video for promotion, and also prior to the course release, it is possible to put up a reference page in the platform, that gives an overview of the course, who is giving it, the content, the prerequisites, duration, and other information that could be of interest for the participants.

The course we designed for this project was called **TRADARES** (short for “Traducción Árabe - Español”), and we started promoting it via a wordpress page: [tradares.wordpress.com](http://tradares.wordpress.com). This page was intended to allow pre-inscription of participants while the platform was still not running. The course was offered in <http://openedx.tradares.es/>. It can be accessed using the dummy account:

```
user: user@example.com // password: volare20
```

Pre-inscription promotion was mainly done via social networks, especially Facebook, and also Twitter. Some webpages dedicated to Arabic teaching resources were so kind to put up links for the course as well. During the month of august the site received 752 visits

from 34 different countries, the ten countries where most of the visit came from are shown in table 3.1.

Preinscription ran from **01/08/2015** to **16/08/2015**.

The course ran from **17/08/2015** until **08/09/2015**

The schedule was as follows:

#### 1. Week 1

- (a) Translation 1: On the topic of migration through Melilla
- (b) Translation 2: On the topic of a Moroccan personality, Abdelkrim El Khattabi
- (c) Translation 3, On the topic of Ada Colau becoming mayor of Barcelona and the reaction of the Arab community.

#### 2. Week 2

- (a) Translation 4. On marriage in Morocco.
- (b) Translation 5. On Machine Translation.
- (c) Translation 6. On comic books in Arabic. Not released.

#### 3. Week 3: Evaluations

The reasoning behind the selection of texts was to offer something that could be of interest to participants, since interest was the only way to get them to translate. For the Arabic into Spanish language pair, Moroccan texts are the most widely translated, due to the vicinity of the countries and due to the fact that the Moroccan community is by far the largest Arab community in Spain. So typically, a translator working with this two languages will encounter texts as the ones presented, with the exception of the last two, who were specifically designed in order to spice the translations up, after 4 texts that deal with more serious topics.

In any case, the ideal setting for this kind of work would be either have a truly massive participation or be carried out within formal education, so that participants work on the translations pursuing some kind of official recognition, maybe as part of university classes.

Table 3.2 Submissions for the first exercise of each text and for each translation

	Text 1	Text 2	Text 3	Text 4	Text 5
Exercise 1	75	57	55	36	37
Translation	32	23	12	12	15

### 3.2.2 Exercises

Each text was accompanied by exercises. These exercises were mostly extracted from an unpublished textbook on Arabic -Spanish translation that is being written by Dr. Manuel Feria from the University of Granada, to be used by students of the Grade in Translation and Interpreting. An implementation of several of the exercises present in the book was carried out, adapting them to the content of the proposed translations, and taking advantage of the possibilities offered by the open EdX platform, which allow for instance, to convert tables into drag and drop puzzles, or to transform questions into simpler multiple choice automatic graded exercises.

### 3.2.3 Peer Review

The whole idea for this project was based on the possibilities that peer review exercises give as annotation tools. For this project, participants were asked to provide one evaluation to three different translations submitted by three different participants. In turn they would receive three evaluations for their submitted work. That is the workflow for each of the five translations proposed, as seen in figure 3.1.

Peer review through open EdX has a time constraint. If the window of time given to provide reviews is too small, some participants may be left out. If it is too big, participants who submitted on a Monday may not be there on Friday to assess the work that one of their peers submitted in the last moment. We have only collected translations with three peer assessments.

#### The peer review rubric

In order to provide an assessment of the quality of other peers' translations, participants were instructed to give those translations a number between 1 and 4, roughly based on translation edit rate, as seen in Specia[Specia and Cristianini], The rubric they used was the following:

**Amount of edition necessary for the translation to be acceptable:** The point is to indicate how many edits should be needed roughly in order to turn the translation we are evaluating into a working translation that serves its purpose of channeling the meaning of the text from the source language to the target language.

1. (Label 1) Almost everything should be edited. It is necessary to translate the text again. This translation does not serve its purpose.
2. (Label 2) A significant amount of edition is needed, but it is not necessary to translate again. It is faster to edit than to start the translation again.
3. (Label 3) Little edition is needed. The translation we are evaluating serves its purpose, though some changes still need to be done.
4. (Label 4) There is no need to edit. The translation serves its purpose and there is no need to make any change or just some minimal change.

As an optional part in the rubric, the following questions were raised:

Could you point out the reasons for your evaluation? What kind of error did you find in the present translation? You could specify if they are spelling errors, grammar errors (verb tenses, gender and number concordances, etc), translation sense errors, omissions, unfinished translation.

The idea was that each translation would also be tagged with keywords extracted from the optional part, such as “grammar”, “spelling”, “omissions”, etc, aiming to establish a relationship between the translation features and these tags. This part did not work out because several other tags were used, which must have made sense for the evaluator but that turn out to be ambiguous or lack meaning, namely, we have found several comments such as “translation is a little bit literal” or “too literal”; or saying “good style”, and “great style”, or “natural translation”, and not enough references to whether the translation needs edition because poor spelling or is unfinished.

### 3.2.4 Course results

As is the norm in MOOCs[11], only a small percentage of the participants enrolled did actually complete all the proposed translations. Also, from the 130 accounts registered in the platform, 101 effectively registered for the course. The rate of completion of both exercises and translations drops from one translation to another and from one week to another; in the end we could say that 12 participants finished the course. Table 3.2 contains some examples



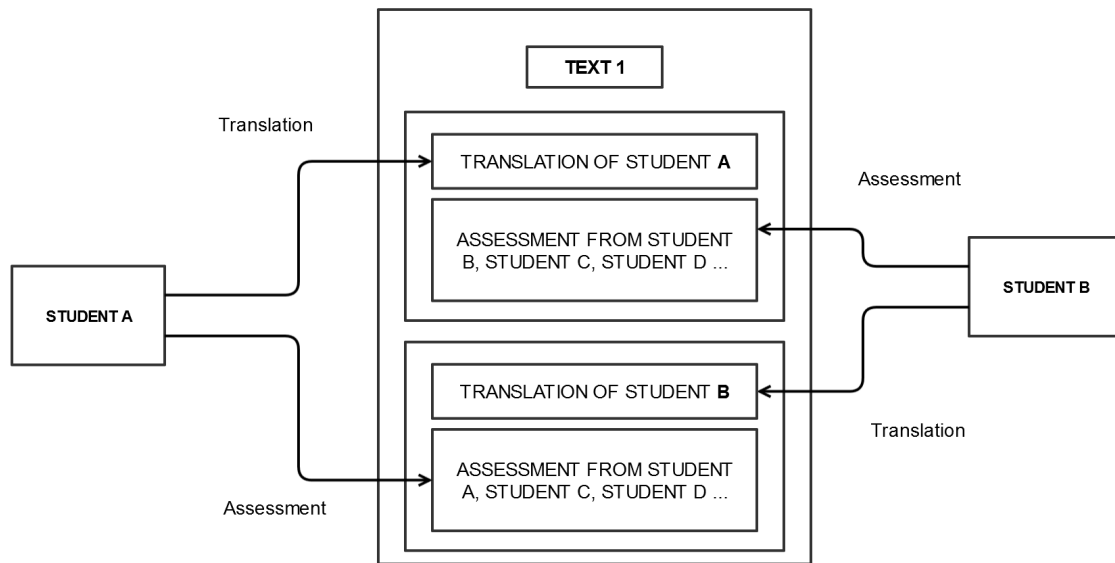


Figure 3.1 Peer review assessment scheme used for the translation evaluation

of the participation figures. The first exercise for each text was usually some easy question or drag and drop puzzle. Exercise 4 involved navigating to a page that provides optical character recognition and required more effort to complete, so that, coupled with the start on the second week, may explain the drop in participation.



# Chapter 4

## Translation evaluation processes

### 4.1 Evaluation as a prediction / classification task

### 4.2 The evaluation tool

As can be seen in figure 3.2 on page 11, the amount of translations collected during the course is small. This situation has called for the use of only two of the translation corpora gathered; that is, the corpus made from translations of the first text and the corpus made from translations of the second text.

The idea behind the evaluation tool is to rely on the abundance of data that can ideally be gathered through massive open online courses, to build the prediction model. Since we want to use the tool within a course, evaluation should not take too long, it should be as immediate as possible. But the most accurate evaluators are based on the data provided by features that are very costly to extract, like language models, parse trees or language pair information. Extracting these features is costly both in time and in expert knowledge.

#### **The two datasets**

The features extracted from the data set have been chosen to keep computational overhead to a minimum, and we have avoided completely the use of features that need to be extracted by other systems, like Moses[7] translation models, or language models, or the use of any other resource that would make the evaluation asynchronous, since we aim to provide quick feedback to the user.

The features extracted, then, are a combination of some of the baseline features mentioned by Specia (those that can be extracted straightforwardly), some features taken from the field of forensic linguistics and authorship attribution [5], TF-IDF counts of the translation in relation with the whole corpus of translations collected, and basic POS tagging.

The following are the labels (the evaluations collected from participants), for the dataset containing the corpus of translations for text 1:

[323, 222, 223, 233, 223, 112, 444, 334, 333, 322, 232, 221, 344, 444, 222, 111, 433, 333, 233, 122, 111, 443, 323, 232, 442, 443, 112, 444, 343, 233, 343, 122]

And for text 2:

[344, 322, 123, 112, 334, 112, 222, 344, 333, 343, 224, 322, 222, 111, 333, 423, 322, 223, 433, 334, 233, 133, 232]

The target label was created by simply calculating the mean between the three assessments. After trying Ridge regression algorithm recommended by Wisniewski by [16] using as target the average of the evaluations, the classification model chosen was built by a multi-class one vs all SVM classifier with linear kernel, as implemented by Scikit Learn [10]. Train - test split was 66% - 33% and different randomized samples were tested as well.

### 4.2.1 Results

The result for the 4 class evaluation does not grant very good results. The sparsity of data makes it easy for one class to be left out of the test data. But it seems that misclassifications are found mainly between close classes, as can be seen in 4.1 and 4.2. Lowest quality translation are the ones that the model identifies the better, as shown in 4.1 and 4.2. Also, if we take into account the labels, there are two things we can notice: one, there are less instances of the extreme classes than of the middle ones, so the model should reflect that; and two, human evaluations also differ slightly, which possibly means that the boundaries between the classes that are next to each other are blurry. Even for human assessment.

This naturally leads to the idea that classification in two classes, one for "good" translations, and one for "bad" translations might be more accurate, and still be of use in the context of an exercise that provides immediate assessment.

Confusion matrices for two class classification show that the main misclassification occurs when the classifier says a "good" translation is "bad", as can be observed in 4.2. This is possibly because otherwise good translations might contain translation errors, which make them differ very little from good ones, but still edition is needed for them to be valid. And here, features used in automatic essay scoring would also not work so well. For instance the issue we can see in the sentence "Abdelkrim El Khattabi luchó junto a los españoles" vs. the sentence "Abdelkrim El Khattabi luchó contra los españoles" is a difficult problem to solve by a computer, since meaning is the key.

Some feature selection methods were tried with Scikit Learn's ensemble of trees algorithm, but surprisingly for a small dataset like this one, where dimensionality almost doubles the observations, the method just recommended dropping 5 features. None of the features is specially salient. The 10 best proposed by Scikit Learn can be checked in 4.1 and 4.2.

A negative classification result could be given with a suggestion to improve, and the student, can work in the submission to try to obtain a "good" classification.

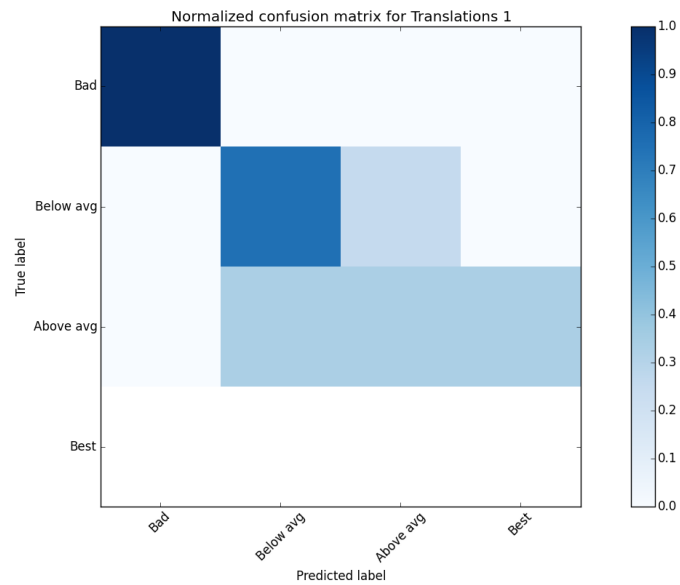


Figure 4.1 Confusion matrix of the 4 class classification of text 1 translations

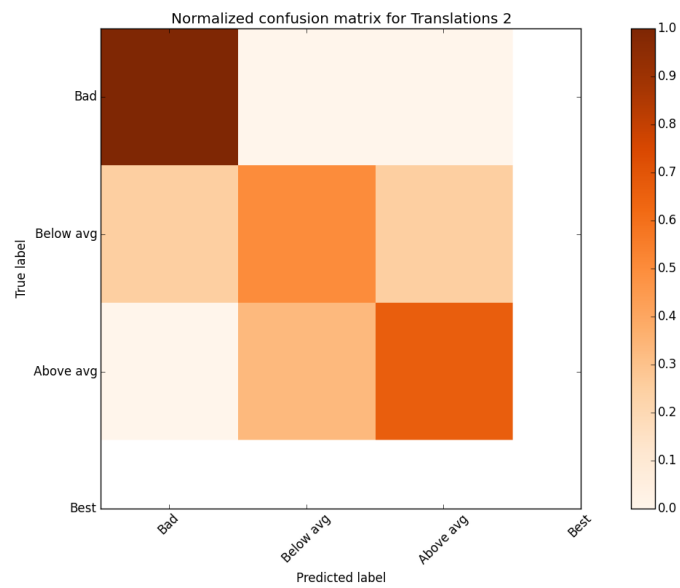


Figure 4.2 Confusion matrix of the 4 class classification of text 2 translations

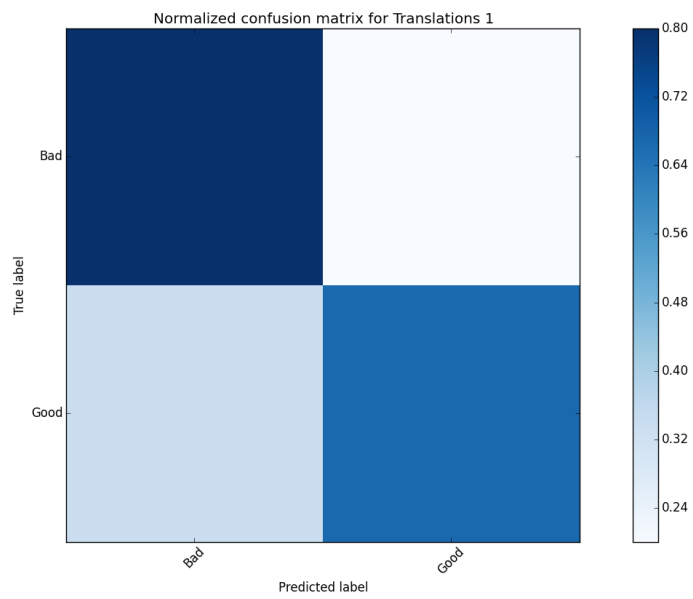


Figure 4.3 Confusion matrix of the binary classification of text 1 translations

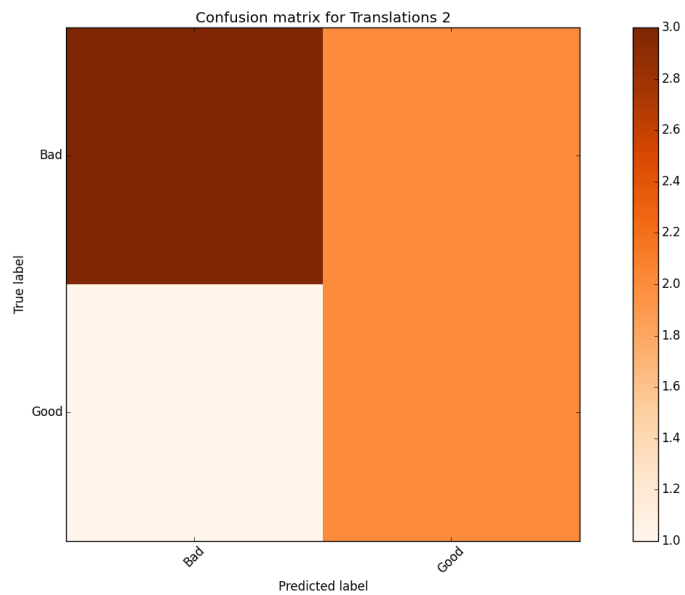


Figure 4.4 Confusion matrix of the binary classification of text 2 translations

Table 4.1 Choice of 10 best features for the classification of text 1 translations into 4 classes (left columns), and into two classes (right columns).

Rank 4	Feature no.	Rank 2	Feature no.
1.	29	1.	58
2.	72	2.	4
3.	58	3.	72
4.	24	4.	29
5.	4	5.	71
6.	62	6.	35
7.	35	7.	8
8.	1	8.	1
9.	71	9.	13
10.	64	10.	51

Table 4.2 Choice of 10 best features for the classification of text 2 translations into 4 classes (left columns), and into two classes (right columns).

Rank 4	Feature no.	Rank 2	Feature no.
1.	45	1.	41
2.	57	2.	57
3.	56	3.	69
4.	5	4.	54
5.	69	5.	5
6.	52	6.	40
7.	49	7.	52
8.	47	8.	10
9.	35	9.	42
10.	43	10	45



# Chapter 5

## Conclusions

### 5.1 The course

Regarding the course, putting open EdX online is quite challenging. It must be noted that this platform has been developed for truly massive participation. It is managed by teams of people, and not by a single person. It is been said to be too hard to deploy to be worthy for smaller projects[Glance]. But recently, the installing process has been smoothed, more documentation has been written about it, more issues have been reported and solved through mailing lists and discussion forums, and from the present experience, it can be set up on the cloud for a relatively small amount. Now, having the full data analytics, which records all kind of interactions between the student and the platform, such as mouse clicks, time spent on each component, time between logging in, and many other metrics, requires a higher Amazon tier S3, the use of Hadoop or Celery, in short, high level expertise.

But with fewer resources, and in smaller environments, an approach like the one described in this work could very well be used in University departments or Official School of Languages, allowing them to easily build all kinds of language corpora, and with proper training of evaluation rubrics, participants would provide annotations for that corpora, which in turn could help gain valuable insight on language acquisition or translation.

But it is very difficult for the linguist to exploit the possibilities of computer analysis of textual data. Tools are written in different programming languages, including but not limited to Java, Python, Perl, C++, which is stopping many people from reaching to the very interesting conclusions that are waiting to be unveiled in the field of natural language processing. There is no doubt that serious programming foundations should be provided in language studies.

It is worth noting how textual data collections compiled years ago are still being used.[12]. Again, producing new textual data for study is difficult. Especially producing material of quality. Annotated data is rare, and research on what kind of annotations could be better is also needed[16]. Maybe such data could be obtained with platforms like EdX and their powerful analytics systems, but especially, with the collaborative effort of many people.

Thus said, the fact is that Edx has built an efficient anonymizing feature into their platform. Only because we are at the same time platform administrators and course creators can we cross the anonymized data from the course with the data from the platform account, but it is still a cumbersome task, that implies matching long string of random alphanumeric.

## **5.2 The translation quality assessment tool**

It has not been possible for us to build the tool back into an open EdX XBlock, since it still depends heavily in external libraries such as Scikit Learn and NLTK[2] that are not thought to be accessed from within the course. Still, there exist workarounds, we would like to see more work directed into building an evaluation Xblock. Also, setting up a course in a better moment would be interesting, as the present course was offered in the middle of the summer, and many professors contacted did not answer until late, or answered that they could not bother alumnus with emails.

# Bibliography

- [1] Balfour, S. P. (2013). Assessing writing in moocs: Automated essay scoring and calibrated peer review. *Research & Practice in Assessment*, 8(1):40–48.
- [2] Bird, S. (2006). Nltk: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 69–72. Association for Computational Linguistics.
- [3] Christensen, G., Steinmetz, A., Alcorn, B., Bennett, A., Woods, D., and Emanuel, E. J. (2013). The mooc phenomenon: who takes massive open online courses and why? *Available at SSRN 2350964*.
- [Edx] Edx. Open Edx documentation. <https://edx.readthedocs.org/en/latest/> year = 2015, note =.
- [5] García Barrero, D. (2012). La atribución forense de autoría de textos en árabe estándar moderno. estudio preliminar sobre el potencial discriminatorio de palabras y elementos funcionales.
- [Glance] Glance, D. The end of openEdx. <https://openedxdev.wordpress.com/2014/04/29/the-end-of-openedx/> year = 2014, note =.
- [7] Koehn, P. (2009). Moses—statistical machine translation system.
- [8] Onah, D. F. and Sinclair, J. (2015). Learners expectations and motivations using content analysis in a mooc. In *EdMedia 2015-World Conference on Educational Media and Technology*, volume 2015, pages 185–194. Association for the Advancement of Computing in Education (AACE).
- [9] Pappano, L. (2012). The year of the mooc. *The New York Times*, 2(12):2012.
- [10] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830.
- [11] Reich, J. (2014). Mooc completion and retention in the context of student intent. *EDUCAUSE Review Online*.
- [12] Shah, K., Cohn, T., and Specia, L. (2015). A bayesian non-linear method for feature selection in machine translation quality estimation. *Machine Translation*, pages 1–25.

- [13] Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of association for machine translation in the Americas*, pages 223–231.
- [Specia and Cristianini] Specia, L. and Cristianini, N. Estimating the sentence-level quality of machine translation systems.
- [15] Wilson, J., Olinghouse, N. G., and Andrada, G. N. (2014). Does automated feedback improve writing quality? *Learning Disabilities—A Contemporary Journal*, 12(1).
- [16] Wisniewski, G., Singh, A. K., and Yvon, F. (2013). Quality estimation for machine translation: Some lessons learned. *Machine translation*, 27(3-4):213–238.

# Appendix A

## Translation texts

### Text 1

قدمت وكالة الأنباء الإسبانية "إيفي" أرقاما جديدة توضح من خلالها أن السياح الحدودي الفاصل بين مدينة مليلية المحتلة وباقي التراب الوطني عاش خلال الثلاثة أشهر الأخيرة حالة من الهدوء لم يسبق أن يسجل منذ بداية تدفق، سنة 2012، أعداد كبيرة من مهاجري إفريقيا جنوب الصحراء على المناطق الشمالية المغربية، بغاية العبور إلى إسبانيا، بحثا عن معانقة الفردوس الأوربي، حيث إنه ما بين فاتح يناير وفاتح مايو من هذه السنة، تم تسجيل دخول 105 مهاجر غير شرعي فقط إلى مدينة مليلية المحتلة، مقابل 1260 مهاجرا في نفس الفترة سنة 2014، في حين لم تسجل الشهور الثلاثة الأخيرة، إلى غاية فاتح غشت الجاري، أي دخول وحسب الأرقام التي قدمتها ذات الوكالة، فإن آخر عملية اقتحام للسياح الحدودي لمدينة مليلية المحتلة سجلت في فاتح مايو الماضي، عندما حاول 400 مهاجر من دول جنوب الصحراء دخولوا المدينة بطريقة غير شرعية، وهي العملية التي لم يتمكن من خلالها أي واحد منهم من الولوج إلى داخل المدينة

### T1 Sample translations

#### T1. Translation evaluated as 4 (No edit needed)

"La agencia de noticias española EFE ha presentado nuevos datos que muestran que la valla fronteriza que separa la ciudad ocupada de Melilla del resto del territorio nacional ha vivido durante los últimos tres meses una situación de calma sin precedentes desde que en 2012 comenzara a afluir un gran número de inmigrantes subsaharianos a las regiones del norte de Marruecos con intención de acceder a España en su busca del paraíso europeo. Desde principios de enero a principios de mayo de este año se han registrado 105 entradas irregulares solo en la ciudad ocupada de Melilla, en comparación con las 1.260 durante el mismo periodo del pasado año. En los últimos tres meses, hasta principios de este agosto, no se ha registrado ninguna entrada.

Según los datos presentados por la agencia, el último asalto a la valla fronteriza de la ciudad ocupada de Melilla se registró a principios de mayo del año pasado, cuando 400 inmigrantes de países subsaharianos intentaron acceder a la ciudad de manera irregular, sin que ninguno de ellos tuviera éxito.”

### **T1. Translation evaluated as 3 (Needs little edit)**

”La agencia de noticias españolas EFE publicó un nuevo número para explicar que la valla fronteriza que separa Melilla del resto del territorio nacional vivió, durante los últimos tres meses, una situación de calma que no se había registrado desde el comienzo del flujo migratorio, en 2012. Un gran número de inmigrantes del África Subsahariana estaban en regiones del norte marroquí con el objetivo de pasar a España, buscando abrazar el paraíso europeo entre enero y febrero de este año. Se ha registrado la entrada de 105 inmigrantes irregulares solo en Melilla frente a los 1260 en el mismo período del 2014 donde en los últimos tres meses no se registró ninguna entrada hasta este mes de agosto.

Según los números que presenta esta agencia, en la última irrupción a la valla de Melilla se registró el pasado mayo cuando entraron alrededor de 400 inmigrantes de los estados del sur del Sahara de forma irregular. Sin embargo, no fue posible para ninguno de ellos acceder dentro de la ciudad.”

### **T1. Translation evaluated as 2 (Needs considerable edit)**

”La agencia española de noticias (EFE) ha ofrecido nuevos datos através los cuales esclarece que la valla que separa la ciudad ocupada de Melilla del resto del terretorio nacional ha vivido, durante los últimos tres meses, un estado de calma no ha registrado desde el comienzo de la afluincia de gran cantidad de inmigrantes al norte de Marruecos, procedentes de África subsahariana, en 2012, con el objetivo de cruzar a España en busca de abrazar el paraíso europeo;ya que entre el primer de enero y el 1 de mayo se ha registrado la entrada de sólo 105 inmigrantes a la ciudad ocupada de Melilla, frente a 1260 inmigrates en la misma etapa de 2014, en cuanto no se ha registrado ninguna entrada durante los últimos tres meses, hasta el 1 de agosto.

Según los datos que ha ofrecido la dicha agencia, el último intento de irrumpir la valla fronteriza se registró en el 1 del pasado mayo, cuando 400 inmigrantes de paises subsaharianos intentaban entrar, de forma ilegal, a la ciudad, y es el intento através el cual nadie de ellos pudo acceder dentro la ciudad.”

**T1. Translation evaluated as 1 (Must be rewritten)**

La Agencia de Noticias IFI ha presentado nuevos números que a través de ello se aclara que la valla que separa entre Melilla y resto del Territorio Nacional ha vivido una situación de relativa calma durante los últimos tres meses. Lo cual nunca se registró, desde el comienzo del derramamiento de grandes números de emigrantes subsaharianos, en el año 2012, con objetivo de pasar hacia España en búsqueda de abrazar el paraíso europeo, ya que durante el 1er enero y el 1er de este año se ha registrado solo 105 emigrantes ilegales con intento de penetración a Melilla, frente a 1260 emigrantes en el mismo período de 2014. Mientras en los últimos meses no se registró ningún intento de penetración.

Según los números que presentó la Agencia, la última penetración hacia la valla de Melilla se ha registrado en el 1er Mayo del mes pasado; cuando intentó 400 emigrantes de los países del sur del Sahara penetrar ilegalmente a Melilla; cuya ningún emigrante alcanzó acceder adentro de la ciudad

## Text 2

عبد الكريم الخطابي  
 ولد في بلدة أجدير في الريف المغربي بين مليلية وتطوان سنة 1301هـ/ 1883م، ودرس القرآن والعربية، وذهب لإكمال دراسته إلى "مليلية" وجامعة القرويين بفاس، وعاد منها ليُعيّن نائبا للقاضي في مليلية، ثم قاضيا، ثم صار أفضى القضاة (قاضي القضاة) وعمره آنذاك لم يتجاوز الثالثة والثلاثين، وهذا دليل على نبوغ مبكر، وكتب في الصحف، ودرس في بعض المدارس، وكان أبوه أميرا على البربر الذين في الريف المغربي، وجاهد مع أبيه في الحرب العالمية الأولى مع الدولة العثمانية، وذلك سنة 1334هـ/ 1915م  
 وفي ذلك الوقت كانت سبتة ومليلية بأيدي الإسبان (وهي إلى الآن بأيديهم) الذين اعتقلوا عبد الكريم لمد 4 أشهر ليضغطوا على أبيه حتى يكف عن الجهاد، وهذه مصيبة من المصائب التي لا يعرفها أكثر المسلمين. وذلك أن الإسبان كانوا يريدون أن يتوسعوا، ويخرجوا من سبتة ومليلية ليحتلوا باقي مناطق المغرب الأقصى الشمالية؛ لكنهم لما حققوا مع الابن فاجأهم بألوان من العزة والثبات، وأخبرهم أنه لا مناص له، ولا لأبيه إلا أن يقاتلوا مع الدولة العثمانية، فاضطروا لسجنه؛ لكنه تدلى بجبل من السجن ليفر؛ إلا أن الحبل كان قصيرا فتأرجح في الهواء فرمى بنفسه، فانكسرت ساقه، وأغمي عليه من الألم، فعثر عليه الإسبان فأعادوه إلى السجن، حيث مكث أربعة أشهر، ثم أطلقوا سراحه

## T2 Sample translations

### T2. Translation evaluated as 4 (No edit needed)

"Abdelkrim El Jattabi.

Nació en 1883 en Ajdir (el Rif), localidad situada entre Melilla y Tetuán. Estudió el Corán y la lengua árabe. Completó sus estudios en Melilla y en la Universidad Qarawiyyen de Fez. A su regreso fue designado representante del Cadí de Melilla, al que más tarde sustituyó. Finalmente alcanzó el puesto de Cadí de Cadíes cuando apenas contaba con treinta y tres años, lo que prueba que era un joven muy brillante. Escribió en diferentes periódicos e impartió clases.

Abdelkrim luchó junto a su padre, que fue Emir de los bereberes del Rif, durante la I Guerra Mundial (1915) en las filas de los otomanos. Los españoles ocupaban por entonces la ciudad de Melilla (y continúan haciéndolo). Allí encarcelaron a Abdelkrim durante cuatro meses para obligar a su padre a no continuar haciendo la yihad. Pocos musulmanes conocen estos padecimientos. Los españoles pretendían expandir su zona de influencia desde Ceuta y Melilla para ocupar el resto del Marruecos septentrional. Interrogado por lo españoles respondió que ni él ni su padre tenían otra alternativa que luchar junto a los otomanos, así que los españoles tuvieron que encarcelarlo. Se colgó con una cuerda desde la prisión para escapar, pero la cuerda era demasiado corta y se quedó colgando; se lanzó al vacío, se rompió una pierna al caer y el grito de dolor alertó a los españoles, que lo volvieron a encarcelar. Tras



---

cuatro meses de calabozo, lo pusieron en libertad.”

## **T2. Translation evaluated as 3 (Needs little edit)**

”Nació en Agadir, en la zona rural entre Melilla y Tetuán, en el año 1883 DC - 1301 AH. Estudió la lengua árabe y el Corán, después, por motivos académicos se trasladó a Melilla y a Fez para estudiar en la universidad de Karawein. Después de haber terminado la carrera, regresó a Melilla para ocupar el puesto de vice juez en y, un poco más tarde, el de juez. A los 33 años fue nombrado juez supremo y eso muestra lo predilecto que era: escribiendo en algunos periódicos e impartiendo clases en unos colegios. Su padre era el príncipe jefe de los bereberes de la zona rural de Marruecos y junto a él luchó en la primera guerra mundial con los otomanos en el año 1915 DC- 1334 AH.

En aquella época, Ceuta y Melilla estaban y siguen bajo la soberanía española que había detenido a Abdikarim durante 4 meses en un intento de presionar al padre para que no siguiese luchando. Es probable que la mayoría de los musulmanes no sepan que los españoles tenían la intención de ampliar su territorio y de ocupar más zonas del norte de Marruecos aparte de Ceuta y Melilla. Durante los interrogatorios con Abdikarim, los españoles fueron sorprendidos por la persistencia y el orgullo que había mostrado, dejándoles muy claro que no habría manera para que él y su padre pusieran fin a su lucha con el Estado otomano; por lo tanto lo ingresaron en la cárcel. Sin embargo, Abdelkarim intentó escaparse agarrándose a una cuerda que desafortunadamente era demasiado corta, y fue entonces cuando empezó a perder el equilibrio y decidió lanzarse. Se rompió la pierna y se desmayó de tanto dolor. Los españoles lo encontraron y lo llevaron de nuevo a la cárcel donde permaneció 4 meses, y finalmente le dejaron en libertad.”

## **T2. Translation evaluated as 2 (Needs considerable edit)**

”Abd Alkarim Alkhatabi

Nacido en la localidad de Axdir en el Rif marroquí entre Melilla y Tetuán en el año 1883, estudió El Corán y la lengua árabe y se trasladó a Melilla para finalizar sus estudios y a la universidad de Qarawiyyin en Fez. Volvió para ser designado representante cadí de Melilla, luego juez, luego juez de jurisdicción cuando no superaba los 33 años de edad, lo cual es un indicador de su talento temprano. Además, escribió en periódicos, estudió en diferentes escuelas y su padre fue príncipe de los bereberes que se encontraban en el campo marroquí. Abd Alkarim Alkhatabi luchó con su padre en la primera guerra mundial con el estado otomano en el año 1915.

En aquel momento, Ceuta y Melilla estaban en manos de los españoles (y aún lo están) que

detuvieron a Abd Alkarim durante 4 meses para forzar a su padre a abstenerse a luchar, y esta es una de las desgracias que no conocen muchos musulmanes. Así que los españoles quisieron expandirse y salir de Ceuta y Melilla para ocupar el resto de las zonas del norte de Marruecos; sin embargo, cuando identificaron al hijo los sorprendió con los colores de la gloria y de la solidez, y les informó de que no tenía escapatoria, ni él ni su padre, mas que combatir contra el estado otomano, pues necesitaban su detención; sin embargo, descolgó una cuerda de la cárcel para escapar. Aunque la cuerda era corta, pues oscilaba en el aire, se tiró y se partió la pierna y perdió el conocimiento del dolor. Finalmente, los españoles lo encontraron y lo devolvieron a la cárcel, donde residió cuatro meses, y finalmente le liberaron.”

## **T2. Translation evaluated as 1 (Must be rewritten)**

”Nació en la ciudad de ayadir en el Rif marroquí, entre Melilla y Tetuán 1301-1883, estudió El Corán y la lengua Árabe, y fue a completar sus estudios en Melilla y a la universidad en Fes. Regreso de ella nombrado Teniente alcalde, luego Juez, y se convirtió en Presidente del Tribunal supremo de Justicia. Ya a la edad de 33 años todavía no se había casado, y esta es la prueba de un Genio Precoz. Escribió en periódicos, y dio clases en algunas escuelas. Su padre era Príncipe de los BERÉBERES en el Rif Magrebi. Lucho con su padre en la Guerra Mundial, primero contra el imperio Otomano en el año 1915-1334. En aquel tiempo estaba Ceuta y Melilla en manos Españolas, y ellas ahora en sus manos, quienes arrestaron a Abdelkrim 4 meses para presionar a su padre para que detuviese la guerra, y esto es un desastre de los desastres que no reconocerían muchos musulmanes. Para aquellos Españoles querían ampliar y que saliesen de Ceuta y Melilla para ocupar el resto de las Regiones del extremo norte de Marruecos, sin embargo les ha sorprendido su hijo con su orgullo y fortaleza, y les dijo que no era inevitable para el y para su padre que muriesen con el Imperio Otomano, así que tuvieron que encarcerarle”

# Appendix B

## Dataset information

Feature no.	Description
1	el
2	la
3	los
4	las
5	un
6	una
7	unos
8	unas
9	a_el
10	de_el
11	que
12	quien
13	quienes
14	cual
15	cuales
16	a
17	ante
18	bajo
19	con
20	contra
21	de_el
22	desde
23	en

---

24	entre
25	hacia
26	hasta
27	para
28	por
29	según
30	sin
31	sobre
32	tras
33	y
34	o
35	pero
36	no
37	dot
38	comma
39	semicolon
40	utu
41	uttu
42	utb
43	uttb
44	btu
45	bttu
46	btb
47	bttb
48	tokens
49	types
50	numSentences
51	hapax 1
52	hapax 2
53	bigram
54	trigram
55	4gram
56	5gram
57	6gram

---

58	7gram
59	8gram
60	9gram
61	10gram
62	tfidf-3gram
63	tfidf-4gram
64	tfidf-5gram
65	tfidf-6gram
66	tfidf-7gram
67	tfidf-8gram
68	tfidf-9gram
69	tfidf-10gram
70	posV
71	posN
72	posA
73	posRG
74	posNone

Table B.1 Ordered list of the features extracted from the translations

