Predicting Source Gaze Fixation duration: a Machine Learning Approach

Tanik Saikh
CSE Department
Jadavpur University
Kolkata, India
tanik4u@gmail.com

Srinivas Bangalore
Intelligent Services Research Lab
AT&T Labs - Research
New Jersey, US
srini@research.att.com

Michael Carl
IBC
CBS
Copenhagen, Denmark
mc.ibc@cbs.dk

Sivaji Bandyopadhyay
CSE Department
Jadavpur University
Kolkata, India
sivaji_cse_ju@yahoo.com

***Abstract*— In this paper an attempt has been made to predict the gaze fixation duration at source text using supervised learning techniques. The machine learning models used in the present work make use of lexical, syntactic and semantic information for predicting the gaze fixation duration. Different features are extracted from the data and models are built by combining the features. Our best set up achieves close to 50% classification accuracy.**

Keywords—Eye Tracking; Gaze fixation duration; Machine Learning; Support Vector Machine

I. Introduction

Staring from the year of 2006 CRITT has developed data acquisition software called Translog [1][2] with which translator's keystroke and gaze activities can be recorded. This tool is now the most widely used tool of its kind [1]. Translog – II[1] [1] is the recent version of Translog. There are many machine translation research has been done. Now it's time to check the quality of machine translation output. So in recent time research on machine translation quality is the matter of focus. The root of the motivation has come from this notion. In this context, focus of research comes in the field of post editing, eye tracking etc. Gaze is the sum of fixations. Two fundamental components of eye behavior are - (a) Gaze-fixation or simply, Fixation i.e. long stay of the visual gaze on a single location, and second one is Saccade i.e. very rapid movement of the eyes between positions of rest [8]. So if it is possible to predict the gaze fixation duration for an unknown token based on some independent feature, it will be very much helpful for several kinds of research activities, and also will be able to how much time a particular translator spent on thinking or writing etc. for translating that particular token. So it is the main motive of this task. Gaze to word mapping data have been taken in order to predict the class level of dependant variable, train the model using independent variable which were the feature extracted from the datasets i.e. Parts of Speech of the token, probability of unigram and bigram, Lexical Entropy of human translator, and many others features are the very good predictor of gaze duration at source sentence.

[1] http://bridge.cbs.dk/platform/?q=Translog-II

This work was supported by EU's 7th Framework Program (FP7/2007-2013) under grant agreement 287576 (CASMACAT) and partly by project "Development of Indian Languages to Indian Languages Machine Translation System (IL-ILMT) Phase II" funded by Department of Electronics and Information Technology, Ministry of Communication and Information Technology, Govt. of India.

II. Related Work

Eye gaze has been extensively studied in psycholinguistic comprehension tasks. Some psycholinguistic studies have focused on the role of eye gaze in languages production task. Reference [6] studied eye movements in an object naming task. It was shown that people fixated objects prior to naming them. Objects that are difficult to name were fixated for a longer period of time than those that are easy to name [5]. Reference [4] presents a model for gaze prediction in egocentric video by leveraging the implicit cues that exist in camera wearer's behaviors.

III. Data

The data that has been used for this experiment is Translation Process Research (TPR) data. TPR- data was collected by the Translog tool[2] and released in 2012 as a Translation Process Research Database (TPR-DB) in CRITT, CBS. The dataset (KTHJ08 dataset) which has been used in this experiment is from the TPR-DB database. Only the source text files have been considered. The KTHJ08 dataset represents the translation of three short general news texts (~150 words) from English into Danish. All three texts have been translated by 24 translators.

IV. Feature Analysis

In general, the feature selection always plays an important role in any machine learning framework and depends upon the data set used for the experiments. Based on a preliminary investigation of the dataset, some of the following features have been identified. Different combinations of the features have also been used to get the best results from the classification task.

A. Parts of Speech

First, The Parts of Speech (POS) tag of a token plays an important role while predicting the Gazes fixation duration. Linear regression model with this feature, gives a Correlation coefficient of 0.32 to Gaze fixation duration, which is standard statistical measure. Regression model is used here to predict the result of an unknown dependant variable, given the values

[2] http://bridge.cbs.dk/platform/?q=resources

of the independent variables. Weka[3] tool has been used for this purpose.

B. Frequency of Unigram and Bigram

Unigram frequency and bigram frequency has been used as a feature for this experiment. It is well known that a words frequency has a strong effect on eye movements [13] less frequent words are more likely to be refixtated than more frequent words. Unigram frequency and bigram frequency have the correlation coefficient of 0.18 and 0.12 in linear regression model respectively to Gaze fixation duration.

C. FFDur

FFDur is First fixation duration. Linear regression with this feature shows it has the correlation coefficient of 0.32 to Gaze fixation duration.

D. TTNum

It is Target text number. It has a great impact on predicting Gaze fixation duration. Linear regression with this feature shows it has the correlation coefficient of 0.079 to Gaze fixation duration.

E. Lexical Entropy

Entropy H(s) is the sum of all observed word translation alternatives multiplied with their information content. It describes the different choices made by the translators in this dataset. It also represents the average amount of non non-redundant information provided by each new item. The word translation probabilities p (s→ti) of a ST (Source Text) word s and its possible translation ti...n are computed as the ratio of the number of alignments s→ti counted in TTs (Target Text) over the total number of observed TT tokens, while in language modeling, the entropy indicates how many possible continuations for a sentence exist at any time (Benjamin et. al). In this experiment it is found as a very good predictor of Gaze at source sentence. Adding entropy as an independent variables increases classification accuracy by nearly 10%, suggesting that the choices a translator has, have some effect on how long a particular word is gazed at during the whole task. A difference in word order, however, seems not to have any effect on GazeS. Linear regression shows it has the correlation coefficient of 0.17 to Gaze fixation duration.

F. Perplexity

Perplexity is related to entropy as an exponential function can be defined as $P = 2^{H}$. The perplexity of a model is a measure that indicates how many different, equally probable words can be produced, and thus how many choices are possible at a certain point in time. The higher the perplexity, the more similarly likely choices exist and hence the more difficult is a decision to make. (Benhamin et.al). Linear regression shows it has the correlation coefficient of -0.02 to Gaze fixation duration.

G. Syntactic Entropy

First, Syntactic entropy describes the entropy of the actual syntactic choices made by the translators in the dataset. Syntax was described using a shallow parse. Each segment was annotated in terms of three categories: valency (transitive, intransitive, ditransitive, impersonal), voice (active, passive) and clause (dependent, independent). The first letter of each of these tags was combined into one triplet per clause in each segment. The probabilities of these strings were computed and on the basis of these probabilities, the syntactic entropy was calculated. It was found that for higher syntactic entropies, behavioral measures were longer (GazeS/T and Kdur), suggesting that different syntactic structures were available to translators and choosing between these is effortful. The linear regression model shows a correlation coefficient of 0.1029 to Gaze fixation duration.

H. Supertag

In a lexicalized grammar such as the Lexicalized Tree Adjoining Grammar (LTAG), each lexical item is associated with at least one elementary structure (tree). The elementary structures of LTAG localize dependencies, including long-distance dependencies, by requiring that all and only the dependent elements be present within the same structure. As a result of this localization, a lexical item may be (and, in general almost always is) associated with more than one elementary structure. We will call these elementary structures supertags, in order to distinguish them from the standard parts of- speech tags. [15]. It has been seen that there is .0015 correlation coefficient of STAG alone to gaze fixation duration. If gives 82.1% classification accuracy along with prob1, prob2, PoS, ParalT, Cross, FFDur, Entropy, perplexity feature in decision tree J48. Linear regression with this feature shows the correlation coefficient of 0.0125 to Gaze fixation duration.

I. Translator Identity

The name of the translator plays an important role for predicting gazes. Linear regression shows it has the correlation coefficient of 0.2458 to Gaze fixation duration.

J. Degree of Polysemy

The degree of polysemy of a sentence is the sum of senses possessed by each word in the Wordnet[4] normalized by the sentence length.

$$DP_{Sentence} = \frac{\sum_{w \in W} Senses(w)}{lenght\ of\ sentence} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (1)$$

Here, Senses (w) retrieves the total number senses of a word P from the WordNet. W is the set of words appearing in the sentence. [9] So for getting the Polysemy of each token no of sense of a token is normalized by token length. So for accessing WordNet Rita.wordnet[5] package has been used, it provides support for accessing to the WordNet ontological database. It has been seen that polysemy has an impact on the prediction of gaze fixation duration.

V. EXPERIMENT AND RESULT ANALYSIS

Three sets of experiment have been performed. First on Gaze per character level and then on Gaze per word level and then the third experiment is on one of its component i.e. on drafting gaze have been carried out. These have been depicted in three subsequent sections.

[3] http://www.cs.waikato.ac.nz/ml/weka/

[4] http://www.wordnet.princeton.edu/

[5] www.rednoise.org/rita/reference/RiWordNet.html

A. Experiment on Gaze per Character

It is well known that word length has a strong effect on eye movements during reading e.g. long words are more likely to be re-fixated than short words.[14] However, the effect of the translation task on eye movements as compared to normal reading has not been studied in detail. One of the first studies to document that the task has an effect on late eye movement measures is the one by Jakobsen and Jensen (2008). These authors documented a tenfold increase in the number of fixations during translation as compared to reading for comprehension. In other words, total reading time on the source text (GazeS) can be seen as capturing task specific processes. In order to account for the effect of word length on GazeS, GazeS was normalized by the number of character of the source word. The normalized total reading time ranged from 0ms to 5144ms. The complete distribution of the gazes per character is given below in Fig. 1, R[6] and R studio[7] have been used for plotting histogram of the gaze at distribution. It shows the distribution of Gaze at source per character is clustered around 1000ms. It was therefore decided to discard normalized Gaze values above 1000ms. In this case maximum data loss 150 points happen, that may come as a noisy data or it may come as outliner.

Baseline: The dependent variable of this experiment is Gaze fixation duration. It represents the sum total of all fixations on a particular source text word during the whole task. It therefore summarize the visual attention of a particular source text word has received during the whole task. The independent variable is the combination of different features. In the context of predicting total gaze time on a source text word (GazeS) from the dataset the whole range of all gaze fixation duration values were taken into account. It normalized by corresponding word length, to obtain the Gaze per character. The values are bucketed into 4 buckets like G1 (0ms -38ms), G2 (39ms-100ms), G3 (101ms-207ms), G4 (208ms-1000ms) for each source token. In this experiment an attempt has been taken to predict a in which class a particular source token's gaze fixation duration belongs to. They are grouped in such a way that the distribution of data points in each group becomes uniform. In that case the baseline classification accuracy is 25%. The result is depicted in the following.

Methodology: The dependant variable of this system is Gaze fixation duration and different set of features are independent variable. The different features have been extracted from the data, combining different combination of feature different models are built, trains the model on supervised learning like support vector machine (SVM). SVM is widely used very popular supervised learning algorithm; it has proven to be the best performer on this dataset compared to other supervised learning algorithm available in weka tool. Weka ML platform [12] has been used to perform the experiments. Based on 10-fold cross validation different classification accuracies have been achieved.

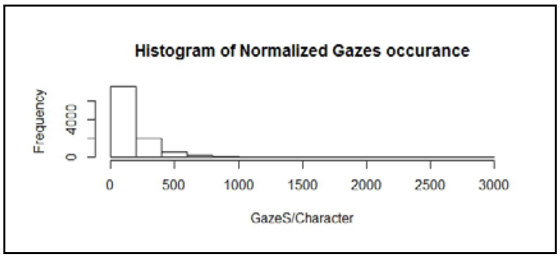

Fig.1. Distribution of the gaze/character value

Result: The best 10 classification accuracies and its corresponding model are depicted in Fig.2, and table 1 respectively. In this scenario the maximum of 42.6% classification accuracy has been achieved by model 10.

B. Experiment on Gaze per Word

Baseline: In this scenario the dependent variable is also Gaze fixation duration and different set of features are independent variable. In this context from the dataset the whole range of all Gaze fixation duration values were taken into consideration. The ranges of values are remain between 0ms to 27832 ms. It is seen from the distribution of Gaze value that there is nothing much more instances remains with gazes duration over 5000ms, which is shown in the Fig. 3, so instances with gaze value more than 5000ms has been discarded. The gaze values are bucketed into four buckets like G1 (0ms-159ms), G2 (160ms-440ms), G3 (441ms-970ms), G4 (971ms- 5000ms.) for each source token. As our aim is to predict a particular source token's gaze duration class. Each bucket contains uniform number of instances. In this case the classification accuracy is 25%.

Methodology: The dependant variable of this system is Gaze fixation duration the different independent variables are the independent variable. In this case different sets of feature have been extracted from the data and different models are built. Train the model on supervised learning. Support vector machine has been used for this purpose. Weka tool has been used for this purpose. The model was tested on 10 fold cross validation. The table below shows the effect of different sets independent variables on classification accuracy.

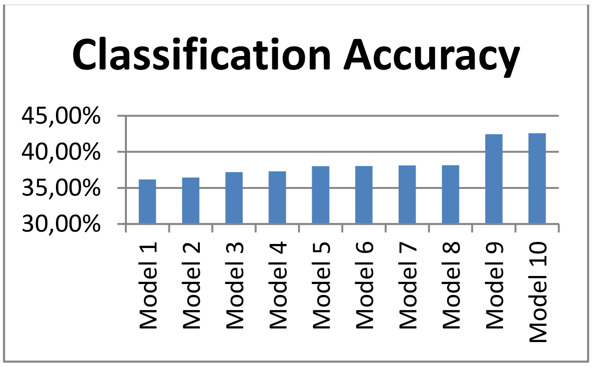

Fig.2. 10 best Classification accuracies on 10 different models (Gazes/character)

[6] http://www.r-project.org/

[7] http://www.rstudio.com/products/rstudio/download/

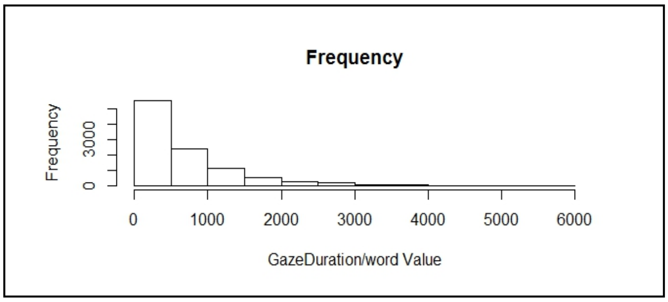

Fig.3. Distribution of the gaze/word value

Result: The classification accuracy against different model on 10-fold cross validation is shown below in Fig. 4, and its corresponding feature analysis is given in table2. Here 48.7% classification accuracy has been achieved by model10.

C. Experiment on Drafting Gaze

In this section experiment has been done on more granular level of gaze at source (gaze at source is the sum of drafting gaze, revision gaze and orientation gaze). Here experiment on drafting gaze has been carried out. Keeping the same baseline and same sets of feature train the model on supervised learning using weka machine leaning tool. In this case classification accuracies improve as compared to previous one.

Baseline: In this case also the instances with 5000ms Gaze values have been discriminated following the previous reason as distribution shows in Fig. 5, the remaining instances are bucketed into four bucket of equally distributed instances so in this case also the classification accuracy is 25%.

Methodology: The dependant in this scenario is drafting gaze value and different sets of feature are independent variable. Selecting different feature from the data different models are built, train the model on support vector machine using weka tool. The model was tested on 10 fold cross validation. The table below shows the effect of different sets independent variables on classification accuracy.

Result: Some of the best results based on 10-fold cross validation have been shown in Fig. 6, and its corresponding feature set has been shown in table 3 below. Here the maximum classification accuracy is 49.1%.

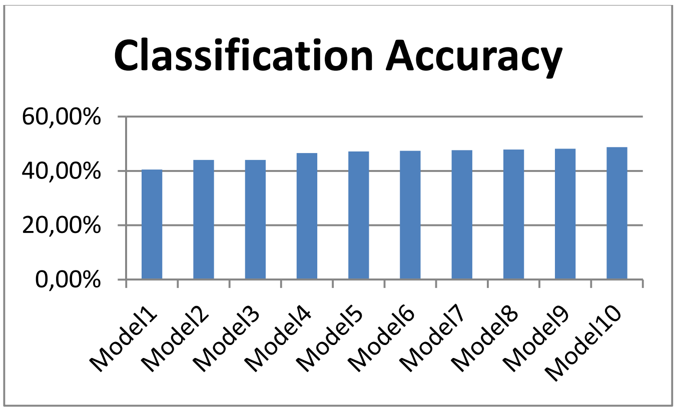

Fig.4. 10 best classification accuracy on 10 different models (Gazes/word)

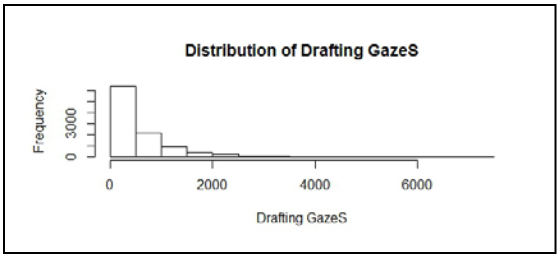

Fig.5. Distribution of the drafting gaze value

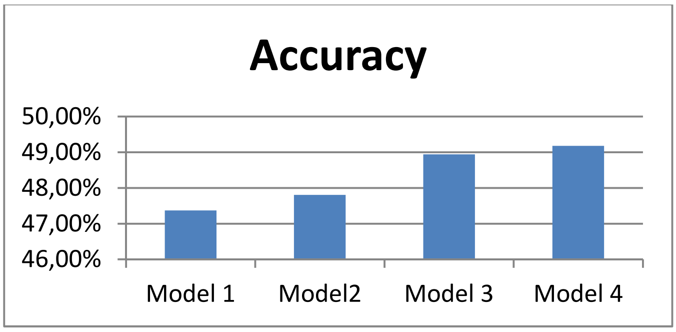

Fig.6. 4 best classification accuracies on 4 different models on drafting gaze

VI. COMPARISION OF THREE EXPERIMENT

The three sets of experiments best result has been compared in Fig. 7, It shows drafting gaze produce better result than Gaze per word then Gaze per character.

VII. CONCLUSION

It can be says that several independent features play important role to predict the gaze fixation duration at source text. The findings of this experiments shows that probability of unigram and bigram, parts of speech, entropy of human translation, TTNum, cross, syntactic entropy, perplexity, Super Tag, Translator Identity all are good predictor. If the gaze fixation duration of a source text can be predict it can be say how much people takes time while text reading comprehension and it will open many more research scope and how much time a translator spent on comprehending the source text.

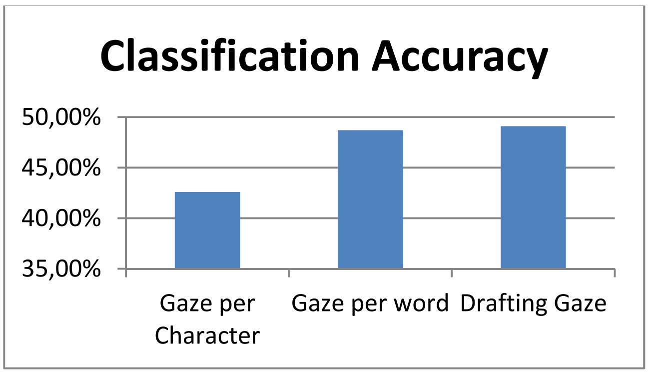

Fig.7. Showing the best accuracies of three sets of experiments

TABLE I. DIFFERENT MODEL AND ITS CORRESPONDING FEATURE SETS FOR EXPERIMENT A.

Feature	Model
Prob1, Prob2, PoS, TTnum, ParalS1,ParalT1,Power of Cross, FFDur, H, Perplexity	Model1
Prob1, Prob2 , PoS, TTnum, H, Perplexity	Model2
Prob1,prob2,Pos,TTNum,ParalT1,ParalS2, ParalT2,Powcross,FFDur,Ins,Del,Entropy,Perplexity, previous class	Model3
Prob1, Prob2 , PoS, TTnum, ParalT1, Cross, FFDur, H, Perplexity, Previous Class, STAGS	Model4
Prob1, prob2, PoS, TTNum, ParalT1, ParalS2, ParalT2, Powcross, FFDur, Entropy, Perplexity	Model5
Prob1,prob2,Pos, TTNum,ParalT1,ParalT2,Powcross,FFDur,Entropy,Perplexity	Model6
Prob1,prob2,PoS, TTNum,ParalT1,ParalS2, ParalT2,Powcross,FFDur,Ins,Del, Entropy, Perplexity	Model7
Prob1, Prob2, PoS, Ttnum, ParalT1,Power of Cross, FFDur, H, Perplexity	Model8
Prob1, prob2, PoS, ParalT1,Cross,FFDur,Entropy, perplexity, previous class	Model9
Prob1, Prob2, PoS, TTnum, ParalT1, PowerOfCount, FFDur, H, Perplexity, Previous Class	Model10

TABLE II. DIFFERENT MODEL AND ITS CORRESPONDING FEATURE SETS FOR EXPERIMENT B.

Feature	Model
Prob1,Prob2,PoS,TTnum,H,perplexity,polysemy	Model1
Prob1,Prob2,PoS,TTnum,len,Cross,FFDur,H,perplexity,SynH,Translator's Identity	Model2
Prob1,Prob2,PoS,TTnum,len,Cross,FFDur,H,perplexity,SynH,Translator's Identity, Polysemy	Model3
Prob1,Prob2,PoS,TTnum,Cross,FFDur,H,perplexity,Translator's Identity	Model4
Prob1,Prob2,PoS,TTnum,len,Cross,FFDur,H,perplexity,SynH	Model5
Prob1,Prob2,PoS,TTnum,len,Cross,FFDur,H,perplexity,SynH,Polysemy	Model6
Prob1,Prob2,PoS,TTnum,Cross,FFDur,len,H,perplexity,STAGS,polysemy	Model7
Prob1,Prob2,PoS,TTnum,len,Cross,FFDur,H,perplexity,Translator's Identity	Model8
Prob1,Prob2,PoS,TTnum,Cross,FFDur,len,H,perplexity,STAGS	Model9
Prob1,Prob2,PoS,TTnum,Cross,FFDur,len,H,perplexity,polysemy	Model10

TABLE II. DIFFERENT MODEL AND ITS CORRESPONDING FEATURE SETS FOR EXPERIMENT C.

Feature	Model
Prob1,Prob2,PoS,TTnum,len,Cross,FFDur,H,perplexity,fileName,polysemy	Model 1
Prob1, Prob2, PoS,TTnum,len,Cross,FFDur,H,perplexity,filename	Model2
Prob1,Prob2,PoS,TTnum,len,Cross,FFDur,H,perplexity	Model 3
Prob1, Prob2,PoS,TTnum,len, Cross, FFDur, H	Model 4

VIII. FUTURE WORK

As a further research scope it will be very much interesting if it can be predict the Gaze fixation duration target text also. Plan to do experiment of revision and orientation gaze. Explore different model with different feature. Clean up noisy data. Try improving the performance and to analysis the error. Gaze may be predictor of cognitive activity like comprehension, translation etc.

IX. ACKNOWLEDGE

Many credit goes to CRITT, CBS group for their organizing such a wonderful Translation Data Analytics (TDA) course without such course this experiment can't be happen, special thanks to CRITT for preparing such TPR-DB Data.

REFERENCES

[1] M. Carl, "Translog-II: A Program for Recording User Activity Data for Empirical Reading and Writing" Research in Proceedings of the Eight International Conference on Language Resources and Evaluation, European Language Resources Association (ELRA), 2012

[2] Jakobsen, Arnt Lykke and Schou, Lasse, "Translog documentation" in Probing the process in translation: methods and results. Copenhagen Studies in Language Series, vol 24, pp. 1-36, 1999

[3] M. Carl, M. J. Schaeffer, "The CRITT Translation Process Research Database V1.4" in The Bridge: Research Platform. Department of Department of International Business Communication (IBC). Copenhagen Business School, 2014

[4] M. Carl, "The CRITT TPR-DB 1.0: A Database for Empirical Human Translation Process Research." AMTA 2012 Workshop on Post-Editing Technology and Practice (WPTP-2012), 2012

[5] Li, Yin et.al., "Learning to predict gaze in egocentric video." Computer Vision (ICCV), IEEE International Conference on, 2013, pp 3216-3223

[6] Prasov, Zahar and Chai, Joyce Y, "What's in a gaze?: the role of eye-gaze in reference resolution in multimodal conversational interfaces" in Proceedings of the 13th international conference on Intelligent user interfaces, 2008, pp 20-29

[7] Meyer et. al., "Viewing and naming objects: Eye movements during noun phrase production" in Cognition, vol 66, pp. B25-B33, 1998

[8] Rayner. et.al, "Toward a model of eye movement control in reading." in Psychological review, vol 105, pp 125, 1998

[9] A. Mishra, P. Bhattacharya, M. Carl: "Automatically Predicting Sentence Translation Difficulty" in ACL, 2013.

[10] K. Rayner, & G. W. McConkie, (1976). "What Guides a Reader's Eye Movements?" Vision Research, 1976, 16(8), 829–837.

[11] A. L. Jakobsen, & K. T.H. Jensen "Eye Movement Behaviour Across Four Different Types of Reading Task. In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), Looking at Eyes. Eye-Tracking Studies of Reading and Translation Processing, 2008 (Vol. 36, pp. 103–124). Copenhagen: Samfundslitteratur.

[12] Witten, Ian H and Frank, Eibe, "Data Mining: Practical machine learning tools and techniques" in Morgan Kaufmann, 2005

[13] A. W. Inhoff, & K. Rayner, "Parafoveal word processing during eye fixations in reading: effects of word frequency." Perception & Psychophysics, 1986, 40(6), 431–9.

[14] Rayner, Keith and McConkie, George W, "What guides a reader's eye movements?" in Vision research, vol 16, pp. 829-837, 1976

[15] S. Bangalore, and A.K.Joshi, "Supertagging: An approach to almost parsing", in Computational linguistics, vol 25, pp. 237-265, 1999