Synthesizing Image Representations of Linguistic and Topological Features for Predicting Areas of Attention

Pascual Martínez-Gómez[1,2], Tadayoshi Hara[1], Chen Chen[1,2], Kyohei Tomita[1,2], Yoshinobu Kano[1,3], and Akiko Aizawa[1]

[1] National Institute of Informatics
{pascual,harasan,chen,kyohei,kano,aizawa}@nii.ac.jp
[2] The University of Tokyo
[3] PRESTO, Japan Science and Technology Agency

Abstract. Depending on the reading objective or task, text portions with certain linguistic features require more user attention to maximize the level of understanding. The goal is to build a predictor of these text areas. Our strategy consists in synthesizing image representations of linguistic features, that allows us to use natural language processing techniques while preserving the topology of the text. Eye-tracking technology allows us to precisely observe the identity of fixated words on a screen and their fixation duration. Then, we estimate the scaling factors of a linear combination of image representations of linguistic features that best explain certain gaze evidence, which leads us to a quantification of the influence of linguistic features in reading behavior. Finally, we can compute saliency maps that contain a prediction of the most interesting or cognitive demanding areas along the text. We achieve an important prediction accuracy of the text areas that require more attention for users to maximize their understanding in certain reading tasks, suggesting that linguistic features are good signals for prediction.

1 Introduction

Reading is an important method for humans to receive information. While skilled readers have powerful strategies to move fast and optimize their reading effort, average readers might be less efficient. When producing a text, the author may or may not be aware of the text areas that require more attention from users. Moreover, depending on the reading objective or strategy, there might be different areas that catch user's attention for longer periods of time. Ideally, readers would know *a priori* what are the pieces of text with the most interesting linguistic characteristics to attain his/her reading objectives, and proceed to the reading act consequently. However, due to the uncertainty on the distribution of time-demanding portions of text, users may incur in an inefficient use of cognitive effort when reading.

The goal is then, given a user and a text, provide a map with the most interesting text areas according to user's reading objective. We work under the assumption that on-line cognitive processing influences on eye-movements [10], and that people with different reading strategies and objectives fixate on words and phrases with different linguistic

P. Anthony, M. Ishizuka, and D. Lukose (Eds.): PRICAI 2012, LNAI 7458, pp. 312–323, 2012.

features. Thus, given the same text, different maps might be displayed when users have different objectives or reading strategies.

Traditionally, in the field of computational linguistics, features and models have been developed to explain observations of natural language, but not to explain the cognitive effort required to process those observations by humans. In the present work, the first step is to quantify the influence of linguistic features when users are performing different reading tasks. We will use an eye-tracker device to capture gaze samples when users read texts. Then, we will synthesize image representations of these gaze samples and image representations of several linguistic features. By assigning a relevance weight to the image representations of linguistic features, we can find the best configuration of the value of these scaling factors that best explain the image representation of the reference gaze samples. After obtaining the influence of each linguistic feature on every reading objective, maps showing the attention requirements of new texts could be automatically obtained and displayed to users before they start their reading act, or documents could be conveniently formatted according to these maps.

As far as we know, our approach is the first attempt to use natural language processing (NLP) techniques while preserving the topology of the words appearing on the screen. The necessity to develop this framework arises from the integrated use of gaze information that consists in spatial coordinates of gaze and the linguistic information that can be extracted using traditional NLP techniques [8]. We believe that by synthesizing image representations of linguistic features, image processing techniques become available to perform natural language processing that inherently incorporates the geometric information of the text and gaze.

This paper is organized as follows. Next section describes previous work related to the present investigation. Then, we briefly introduce a recent text-gaze alignment method using image registration techniques that we borrow from [9] for completeness. In section 4 we introduce the technique to build image representations of reference gaze data. We describe in detail in section 5 how image representations of linguistic features are synthesized and how their influence on reading behavior is estimated. A description of the reading objectives, experimental settings and empirical results are in section 6. Some conclusions and future directions are left for the final section.

2 Related Work

Gaze data and natural languages are different modalities that suffer from an important ambiguity. In order to use these sources of information successfully, meaningful features and consistent patterns have to be extracted. Under this perspective, there exist two types of approaches.

In the first approach, there is a search for linguistic features that activate cognitive processes. A recent example following this idea can be found in the field of *active learning*. Supervised machine learning NLP approaches require an increasing amount of annotated corpus and active learning has proved to be an interesting strategy to optimize human and economical efforts. Active learning assumes the same cost to annotate different samples, which is not accurate. With the purpose to unveil the real cost of annotation, [15] propose to empirically extract features that influence cognitive effort by

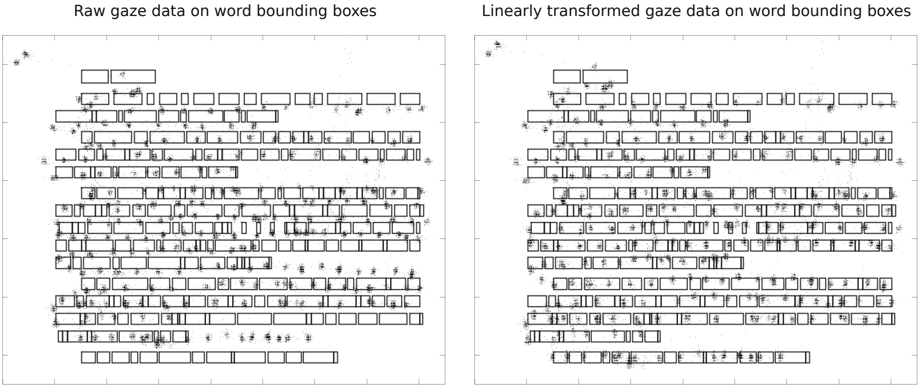

Fig. 1. On the left, raw gaze data superimposed on text bounding boxes. On the right, linearly transformed gaze data (translation $= -38$ pixels, scaling $= 1.13$) on text bounding boxes. Most gaze samples in the figure on the right are mapped onto a reasonable bounding box.

means of an eye-tracker. They affirm that their model built using these features explains better the real annotation efforts.

The objective of the second approach is to find reading behavior characteristics[1] that reflect certain linguistic features. Authors in [3] found that certain characteristics of reading behavior correlate well with traditional measures to evaluate machine translation quality. Thus, they believe that eye-tracking might be used to semi-automatically evaluate machine translation quality using the data obtained from users reading machine translation output.

In the present work, we make use of gaze data reflecting cognitive processing to extract the importance of linguistic features with the objective of predicting the attention that users will have when reading text. This will be implemented using techniques from image recognition for error-correction, function optimization and synthesis.

3 Text-Gaze Alignment

Eye-tracking technology is improving fast and modern devices are becoming more affordable for laboratories and companies. Despite of this rapid development, eye-trackers still introduce measurement errors that need to be corrected at a preprocessing stage. [4] describe two types of errors in a typical eye-tracking session. The first type is variable errors that can be easily corrected by aggregating the gaze samples into fixations. The second type is systematic errors. Systematic errors often come in the form of vertical drifts [5] and they are more difficult to correct.

Researchers and practitioners of eye-tracking have developed their own error correction methods, but they are either too task-specific or introduce constraints on reading behavior. [9] model the error-correction problem as an image registration task that is

[1] Under the assumption that reading behavior is an indicator of cognitive processing. This is also called the *eye-mind assumption*.

briefly described in this section. Image registration is a well studied technique to align two or more images captured by different sensors. In our context, it can be used as a general error correction method in unconstrained tasks where the objective is to align the image representation of gaze samples to the image representation of words appearing on the screen. This method works reasonably well under the assumption that users solely fixate on text. The key is to define a spatial transformation to map gaze coordinates into the space of words. Addressing the vertical drift reported in [5], a linear transformation is defined as $g_{a,b} : (x, y) \to (x, y \cdot b + a)$ where a (translation) and b (scaling) are the transformation parameters of the y-coordinates that have to be estimated by means of a non-linear optimization strategy. An easy objective function measures how well gaze samples are mapped *into* word bounding boxes. Let $\mathbf{G}_{a,b}$ be the image representation of the mapped gaze samples, where pixel $\mathbf{G}_{a,b}(i, j)$ has a high value if there is a gaze sample in coordinates (i,j). And let $\mathbf{W}$ the image representation of word bounding boxes in a text, where pixel $\mathbf{W}(i, j)$ has a high value if it falls inside a word bounding box. A measure of alignment between the two image representations can be defined as the sum of absolute differences of pixels (i,j):

$$f(\mathbf{G}_{a,b}, \mathbf{W}) = \sum_i \sum_j |\mathbf{G}_{a,b}(i, j) - \mathbf{W}(i, j)| \tag{1}$$

The intention is then to estimate the values $(\hat{a}, \hat{b})$ of the transformation parameters that minimize the objective function f:

$$(\hat{a}, \hat{b}) = \operatorname*{argmin}_{a,b} f(\mathbf{G}_{a,b}, \mathbf{W}) \tag{2}$$

$$= \operatorname*{argmin}_{a,b} \sum_i \sum_j |\mathbf{G}_{a,b}(i, j) - \mathbf{W}(i, j)| \tag{3}$$

Due to the non-convexity nature of the solution space, this optimization is iteratively performed using different levels of blurs in what is called multi-blur optimization and hill-climbing at every iteration. Typical results of the error correction of gaze samples using this method can be found in fig. 1. Then, by using the information on the structure of the text, gaze samples can be collapsed into fixations according to their closest word bounding box.

4 Gaze Evidence

Psycholinguistic studies have long noted that eye-movements reveal many interesting characteristics [14]. For example, fixation locations are usually strongly correlated with current focus of attention and, when users read text, they indicate the identity of the word or phrase that the subject is currently processing. Another variable, fixation durations, is useful in quantifying other hidden processes such as user's familiarity with the text or with specific terms, whether or not the text is written in the user's native language, etc. Among saccadic movements, length and direction of regressions are also interesting features of the reading act that occur in diverse situations as when the subject reads about a fact that gets in contradiction with prior knowledge. Although forward and backward saccadic movements provide relevant information on subjects and texts, they might be ambiguous and difficult to interpret by using automatic methods in unconstrained tasks. For this reason, in this work we will only focus on the interpretation of fixation locations and their duration.

Fixation locations and their durations depend on many variables that mainly come from two sources. The first source is from subject's personal characteristics, namely prior background knowledge, native language, cultural identity, interests or reading objectives. Examples of reading objectives are precise reading, question answering, writing a review or preparing a presentation. The second source of variables are related to the linguistic characteristics of the text.

As we have previously stated in the introduction, we are interested in finding the importance of individual linguistic features to explain certain reading behaviors when users are reading texts with different objectives in mind. When pursuing these objectives, different users may have different levels of success. There are many types of reading objectives that can be set to guide subject's reading strategy, where writing reviews or preparing presentations tasks are among them. The level of success of these reading tasks can be evaluated by assessing the performance of the actions that subjects have to carry out after reading, but it might be difficult since there are other variables that may influence subject's performance, such as personal (in)ability to prepare presentations or prior prejudices about the topic the subject writes the review about. For this reason, we limit ourselves to reading objectives whose attainment degree can be easily evaluated as a function of the level of understanding achieved, as measured by an interview with the subject to quantify the accuracy and completeness of his/her answers. Examples of this type of reading objectives are precise reading, question answering or obtaining general information in a very limited amount of time.

Our intention is to predict the text areas where subjects should fixate longer in order to maximize their level of understanding. Thus, we have to obtain gaze evidence that serves as a reference of effective reading behavior. One may be tempted to sample the population of subjects and select the most effective reader. There might be, however, other subjects that follow a different reading strategy and achieve other effectivity levels that we should take into consideration. In order to include this uncertainty in our system, we weight the gaze evidence obtained by all users according to their level of understanding. Let $\mathbf{G}_u$ be the image representation of gaze evidence obtained from user u when reading a certain text, and let λ_u his/her level of understanding on that text. By considering the image representation of gaze evidence as a matrix whose (i, j) positions denote pixels with a gray value between 0 and 1, we can obtain a reference gaze evidence by scaling the evidence of every subject with his/her level of understanding:

$$\mathbf{G}^{\mathcal{T}} = \sum_u \lambda_u \cdot \mathbf{G}_u \tag{4}$$

An schema representing the idea of scaling gaze evidence by user's level of understanding can be found in fig. 2. The linear combination of image representations of gaze evidence is carried out using scaling factors λ_u and it is essential to preserve the sense of uncertainty in our system.

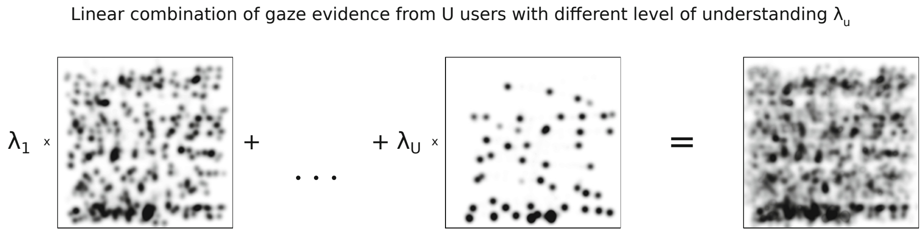
Linear combination of gaze evidence from U users with different level of understanding λu
λ1 ×
+
. . .
+ λU ×
=

Fig. 2. In order to obtain a reference gaze evidence $\mathbf{G}^{\tau}$ that also accounts for the uncertainty of different reading strategies, image representations of gaze samples are scaled by the level of understanding λ_u of every user.

5 Quantifying Influence of Linguistic Features

Eye-actions can be roughly classified into two important categories. The first one consists in fixations, where eyes remain still, gazing on a word or phrase for lexical processing. The second category is saccades, consisting in abrupt eye-movements that are used to place the gaze on different text locations. Syntactic and semantic integration of lexical information is believed to influence these eye-movements [10]. While both eye-actions provide much information about on-line cognitive processing, in this work we will only use information about fixation locations and their durations.

The identity of fixated words and their fixation duration depend on many factors. Some of these factors are related to personal characteristics, e.g. reading objective, reading skill, prior knowledge that serves as background, user's interests, etc. Other factors depend on the linguistic features of the text [14], e.g. lexical properties of words, syntactic or semantic features, etc. The study on the impact of each of these factors is interesting, but it will be limited in this work to linguistic features and reading objectives. The factor of reading objectives influences on data collection and it will be discussed in its own section.

Within the image recognition field, saliency maps [6] represent visual salience of a source image. Similarly, we can synthesize image representations of a text that describe it, while preserving the topological arrangement of words. There are multiple possibilities to describe text according to its linguistic features. Many interesting linguistic features can be numerically described. For instance, we can think of word unpredictability as a probability, as given by an N-gram language model. Other examples are word length, or semantic ambiguity, according to the number of senses in WordNet [11].

To synthesize an image representation of a certain linguistic feature, we filled the word bounding boxes with a gray level between 0 and 1, proportional to the numerical value of the linguistic feature. Fig. 3 shows two examples of image representations of linguistic features. In order to account for the uncertainty introduced by measurement and user errors, images are slightly blurred by convolving them with an isotropic gaussian filter with spread $\sigma = 10$ pixels. Finally, to normalize the intensity when comparing different image representations of linguistic features, the intensity of the images are adjusted so that only the upper and lower 1% of the pixels are saturated.

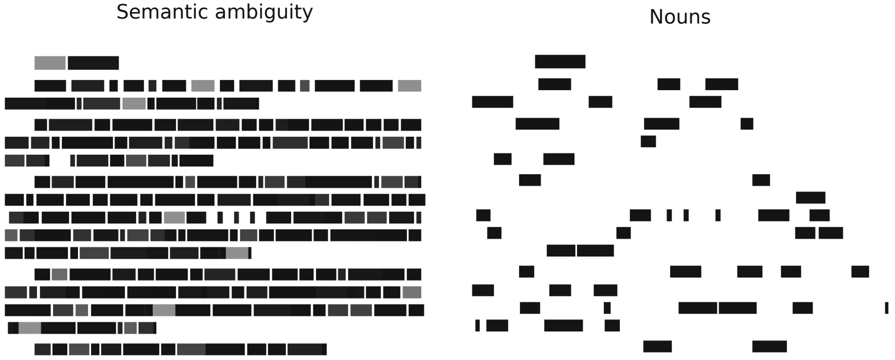

Fig. 3. Two examples of image representations of linguistic features. On the left, image representation of semantic ambiguity as given by the number of senses of every word in WordNet. On the right, image representation of Nouns, where bounding boxes of words that are nouns are filled with high pixel values. Note that the pixel values of these images are complemented for clarity.

The final step is to find the weight ω_f of the image representations of every linguistic feature. Let's consider again images as matrices whose pixels are elements with a value between 0 and 1 and denote by $\mathbf{G}^\tau$ the image representation of the error-corrected gaze evidence. We denote by $\mathbf{W}_1^F = \{\mathbf{W}_1, \ldots, \mathbf{W}_f, \ldots, \mathbf{W}_F\}$ the list of image representations of F linguistic features. A dissimilarity function between the gaze evidence and the linguistic features can be defined as the absolute pixel-wise (i, j) difference between the images:

$$g(\mathbf{G}^\tau, \mathbf{W}_1^F) = \sum_i \sum_j |\mathbf{G}^\tau(i,j) - \sum_{f=1}^{F} \omega_f \cdot \mathbf{W}_f(i,j)| \tag{5}$$

where the image representations of the linguistic features are linearly combined by scaling factors ω_f. Then, a standard algorithm can be used to perform a non-linear optimization to minimize the dissimilarity. Formally,

$$\hat{\boldsymbol{\omega}} = \operatorname*{argmin}_{\boldsymbol{\omega}} g(\mathbf{G}^\tau, \mathbf{W}_1^F) \tag{6}$$

$$= \operatorname*{argmin}_{\boldsymbol{\omega}} \sum_i \sum_j |\mathbf{G}^\tau(i,j) - \sum_{f=1}^{F} \omega_f \cdot \mathbf{W}_f(i,j)| \tag{7}$$

A graphical schema of the combination process can be found in fig. 4. The objective of the optimization is to estimate the importance of different linguistic features so that the linear combination best explains the gaze evidence used as a reference.

6 Experimental Results

6.1 Experimental Settings

We believe that the importance of different linguistic features to explain certain reading behavior depends on the reading objective. Following this hypothesis, we designed three

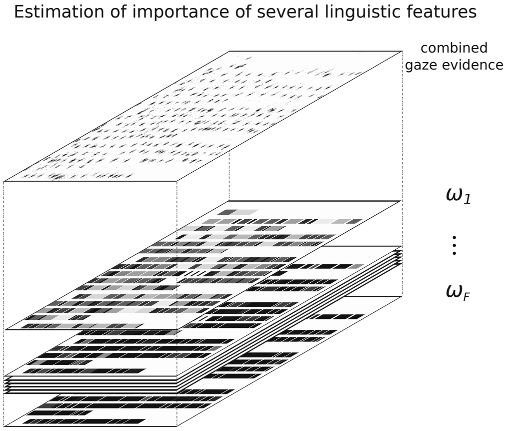

Fig. 4. Schema representing the estimation of several linguistic features (on the bottom) that best explain certain gaze evidence (on the top). Image representations of different linguistic features are linearly combined with scaling factors ω_f and their values are estimated by means of a standard non-linear optimization method. For clarity, the value of the pixels in the images are complemented and the image representations have not been blurred nor their intensity adjusted.

tasks with different reading objectives. There were two documents with a different topic for every task. In the first task, subjects were asked to carefully read a text and that some questions about the text will be asked after. Subjects were also told that they could spend as much time as they need to maximize their understanding about the text, but such a text would not be available during the evaluation of their understanding. In order to check the level of understanding, open questions, true-false and select-correct-answer questions were asked. In the second task, subjects were given questions before the reading act and asked them to find the answers in the text. We asked one and two short questions (for each document, respectively) for users to easily remember them and not to cause extra cognitive load. In the third task, we asked subjects to obtain as much information as possible from a text in only 10 seconds. We told them that they would be required to speak out as much information as they obtained after the reading and that their reading did not have to be necessarily sequential.

For every task, there were two documents in English presented to the subjects. These documents consisted in short stories extracted from interviews, fiction and news. The average number of sentences per document was 13.50 and the average number of words per sentence was 20.44.

While subjects were reading the documents on the screen, Tobii TX300 was used to capture gaze samples at a rate of 300 Hz in a 23" screen at a resolution of 1920 x 1080, resulting in more than 800MBs of gaze coordinates. Text was justified in the center of the screen, from the top-most left position located at pixels (600px, 10px) to the bottom-most right position located at (1300px, 960px). Words were displayed in a web browser using Text2.0 framework [1] allowing to easily format the text using CSS style-sheets

Table 1. List of linguistic features divided in three categories: lexical, syntactic and semantic. Instances of `$tag` are "Noun", "Adjective", "Verb", etc. Word unpredictability is computed as the perplexity by a 5-gram language model. Heads and parse trees are extracted using Enju [12], an English HPSG parser.

Category	Linguistic feature	type
Lexical	word length	Integer
	contains digit	Binary
	word unpredictability	Real
	contains uppercase	Binary
	is all uppercase	Binary
Syntactic	is head	Binary
	is POS `$tag` (23 features)	Binary
	height of parse tree of its sentence	Integer
	depth of the word in the parse tree	Integer
	word position in sentence	Integer
Semantic	is named entity	Binary
	ambiguity: number of senses from WordNet	Integer

and to recover the geometric boundaries of the words. The font family was Courier New of size 16px, text indentation of 50px and line height 30px, in black on a light gray background. A chin rest was used to reduce errors introduced by readers. Text on the screen was short enough not to require any action to entirely visualize it. There were 10 subjects participating in three tasks consisting in two documents each. The subjects were undergraduate, master, Ph.D. and post-doc students from China, Indonesia, Japan, Spain, Sweden and Vietnam with a background in Computer Science.

The eye-tracker was calibrated once per subject and document. Then, an unsupervised text-gaze alignment using the image registration technique [9] was used to automatically correct vertical measurement errors. There were two sessions[2] (out of 60) that needed manual correction of horizontal errors and another session had to be corrected using better vertical scaling and translation than the one automatically obtained by the unsupervised method.

There is a huge amount of linguistic features that could be considered to explain certain reading behaviors. Intuitively, some features might be more relevant than others, but ideally all of them should be included in the model with different scaling factors according to the gaze evidence. We divided the type of linguistic features into three classes: lexical, syntactic and semantic features. Although the list could be bigger, Table 1 contains the linguistic features that were used in this work. Examples of these features are the part-of-speech (POS) tag, or the word unpredictability as measured by the perplexity computed using a 5-gram language model estimated using the EuroParl corpus [7] and smoothed using modified Kneser-Ney smoothing [2]. In order to find the best estimates $\hat{\omega}$ of the scaling factors ω_f in eq. 7, we used Powell's dogleg trust region algorithm [13] as a standard non-linear optimization method.

[2] A session is defined as a subject reading a document.

Table 2. Average values and standard deviation of dissimilarity between the image representation of the test gaze evidence and the linear combination of weighted image representations of linguistic features in a cross-validation. The scale of the values can be found in the last column.

	precise reading		question answering		10-second reading		
	doc. 1	doc. 2	doc. 3	doc. 4	doc. 5	doc. 6	scale
baseline	403 ± 4.1	456 ± 5.9	463 ± 9.7	444 ± 4.5	431 ± 0.2	$\mathbf{415 \pm 0.1}$	$\times 10^3$
scaled feats.	$\mathbf{288 \pm 3.1}$	$\mathbf{325 \pm 2.6}$	$\mathbf{349 \pm 8.5}$	$\mathbf{371 \pm 3.6}$	$\mathbf{419 \pm 4.1}$	562 ± 5.7	$\times 10^3$

Table 3. Average correlation between the vectors of scaling factors obtained from the cross-validation. A high intra-document correlation can be appreciated between the scaling factors within the same document. A low inter-document correlation can be appreciated between the scaling factors estimated for different documents within the same task.

	precise reading		question answering		10-second reading	
	doc. 1	doc. 2	doc. 3	doc. 4	doc. 5	doc. 6
doc. 1	**0.94**	0.34	–	–	–	–
doc. 2	0.34	**0.96**	–	–	–	–
doc. 3	–	–	**0.93**	0.37	–	–
doc. 4	–	–	0.37	**0.96**	–	–
doc. 5	–	–	–	–	**0.96**	−0.19
doc. 6	–	–	–	–	−0.19	**0.94**

6.2 Results

In order to evaluate the predictive power of the linear combination of image representations of linguistic features, leaving-one-out cross-validation was used. Cross-validation was carried out among all subjects within the same document, and every observation consisted in a single session (per document) containing gaze evidence of a subject. Using the training set, scaling factors of the linguistic features were estimated and an image representation of the weighted linguistic features was synthesized and compared to the gaze evidence in the test data. The average results of such comparison can be found in Table 2. As a baseline, we used uniform weights to scale the image representations of the linguistic features in Table 1.

It can be observed that in the precise reading and question answering tasks, there is a consistent and significant reduction in the dissimilarity of the image representations of the test gaze observations and the linear combination of image representations of linguistic features, when compared to the baseline. However, in the 10-second task, the model fails at predicting the distribution of the gaze evidence since the dissimilarity is not consistently reduced. We have two hypotheses to explain such a fact. The first one is that subject's personal characteristics (e.g. background knowledge, native language, etc.) are essential features to explain the reading behavior in the latter task. The second hypothesis is that we have left out important linguistic features.

As we have seen, for the precise reading and the question answering task, the linear combination of image representations of linguistic features helps to explain the gaze evidence of readers within the same document. The remaining question is whether the scaling factors of the linguistic features are good predictors for different documents that are being read using the same reading strategy. To answer this question, we have computed the average correlation between the value of the estimated scaling factors for observations between different documents of the same reading task, together with the correlations within the same document. The results can be observed in Table 3. It can be appreciated that the inter-document correlation is low, suggesting that the estimated weights of the linear combination from one document are not good predictors for other documents. This contrasts with the high intra-document correlation, reinforcing the consistency of the estimations to explain gaze evidence from different subjects within the same document. Since the intra-document precision is considerable, the most plausible explanation is that more documents of the same reading objective are needed to robustly estimate the scaling factors of such amount of linguistic features.

7 Conclusions and Future Work

In the first stage, we have collected gaze evidence from subjects reading documents using three different reading objectives and measured their level of understanding. Well-known systematic errors were corrected using image-registration techniques. Then, a reference gaze evidence has been obtained by linearly combining the image representations of the gaze evidence of every subject scaled by their level of understanding. In the second stage, image representations of several linguistic features have been synthesized and the importance of every linguistic feature has been estimated to explain the reference gaze evidence. The predictive power of the linear combination of image representations of linguistic features have been assessed on held-out data. Our model obtains higher recognition accuracy than a non-informed system in the precision reading and question answering task. However, our model fails at predicting reading behavior in the 10-second reading task.

In order to evidence the generalization power of our model using the estimated scaling factors, we computed the correlation between the scaling factors trained on different documents of the same reading objective task. We found a high intra-document correlation but a low inter-document correlation within the same task.

The results of this work find an immediate application to collaborative filtering and recommendation using gaze data as implicit feedback, since using gaze data from different users within the same document proves to be useful for prediction. For user personalization, however, systems may need gaze data captured from a larger amount of documents.

For the future work, it might be interesting to include linguistic features that are related to the discourse of the document and that we believe may play a significant role in academic learning. Another interesting research direction is related to the study of the personal characteristics that help to explain certain gaze evidence beyond the linguistic features of the text. We acknowledge, however, the intrinsic difficulty of obtaining an accurate description of user's personal characteristics in a large scale real application. In such scenario where we are constrained to a low intrusion into personal characteristics, user's personal features can be included into the model as latent variables and they can be estimated, together with the patent variables (e.g. linguistic features), formulating the optimization as an incomplete data problem.

Acknowledgments. Work partially supported by Kakenhi, MEXT Japan [23650076] and JST PRESTO.

References

1. Biedert, R., Buscher, G., Schwarz, S., Möller, M., Dengel, A., Lottermann, T.: The Text 2.0 framework - writing web-based gaze-controlled realtime applications quickly and easily. In: Proc. of the International Workshop on Eye Gaze in Intelligent Human Machine Interaction, EGIHMI (2010)
2. Chen, S.F., Goodman, J.: An empirical study of smoothing techniques for language modeling. Computer Speech and Language 4(13), 359–393 (1999)
3. Doherty, S., O'Brien, S., Carl, M.: Eye tracking as an MT evaluation technique. Machine Translation 24, 1–13 (2010), `http://dx.doi.org/10.1007/s10590-010-9070-9`, doi:10.1007/s10590-010-9070-9
4. Hornof, A., Halverson, T.: Cleaning up systematic error in eye-tracking data by using required fixation locations. Behavior Research Methods 34, 592–604 (2002), `http://dx.doi.org/10.3758/BF03195487`, doi:10.3758/BF03195487
5. Hyrskykari, A.: Utilizing eye movements: Overcoming inaccuracy while tracking the focus of attention during reading. Computers in Human Behavior 22, 657–671 (2005), `http://dx.doi.org/10.1016/j.chb.2005.12.013`
6. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(11), 1254–1259 (1998)
7. Koehn, P.: Europarl: A parallel corpus for statistical machine translation. In: Proc. of the 10th Machine Translation Summit, September 12-15, pp. 79–86 (2005)
8. Martinez-Gomez, P.: Quantitative analysis and inference on gaze data using natural language processing techniques. In: Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces, IUI 2012, pp. 389–392. ACM, New York (2012), `http://doi.acm.org/10.1145/2166966.2167055`
9. Martinez-Gomez, P., Chen, C., Hara, T., Kano, Y., Aizawa, A.: Image registration for text-gaze alignment. In: Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces, IUI 2012, pp. 257–260. ACM, New York (2012), `http://doi.acm.org/10.1145/2166966.2167012`
10. McDonald, S.A., Shillcock, R.C.: Eye movements reveal the on-line computation of lexical probabilities during reading. Psychological Science 14(6), 648–652 (2003)
11. Miller, G.A.: Wordnet: A lexical database for English. Communications of the ACM 38, 39–41 (1995)
12. Miyao, Y., Tsujii, J.: Feature forest models for probabilistic HPSG parsing. Computational Linguistics 34, 35–80 (2008), `http://dx.doi.org/10.1162/coli.2008.34.1.35`
13. Powell, M.: A new algorithm for unconstrained optimization. Nonlinear Programming, 31–65 (1970)
14. Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychological Bulletin 124, 372–422 (1998), `http://dx.doi.org/10.1037/0033-2909.124.3.372`
15. Tomanek, K., Hahn, U., Lohmann, S., Ziegler, J.: A cognitive cost model of annotations based on eye-tracking data. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, ACL 2010, pp. 1158–1167. Association for Computational Linguistics, Stroudsburg (2010), `http://dl.acm.org/citation.cfm?id=1858681.1858799`