As We May Study: The Potential of the The Web as the Language Textbook of the Future

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

ABSTRACT

UPDATED—September 10, 2017.

This paper describes the implementation and deployment of a system that supports learners of a foreign language in reading materials that are personally interesting by allowing them to read in an interactive reader news, blogs, and other materials from the internet.

The system provides translations for the unknown words at the touch of the screen while at the same time monitoring the current state of the knowledge of the learner in order to be able to estimate the texts of the appropriate difficulty.

This paper reports on the results of deploying the system for one month with sixty Dutch highschool students learning French. The students were very positive about the system, and their teacher has decided to redeploy the sytem for the next academic year.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; See http://acm.org/about/class/1998/ for the full list of ACM classifiers. This section is required.

Author Keywords

Authors' choice; of terms; separated; by semicolons; include commas, within terms only; required.

INTRODUCTION

It is known that, when learning a new language, free reading is one of the best ways of improving ones vocabulary. Reading something that's interesting to the learner will increase their desire to read ML: a prof in california was arguing this I think . And reading is important since it is a microcosm of all the other skills. [3]

However, to enjoy the benefits of free reading the learner must already be quite fluent in the target language: even knowing 95% of the vocabulary in a text means that a learner has to look up a word on every line of text. [1]

Before fluency, most of the times, learners use language text-books as reading material. Textbooks are an artefact of the last century which still proves to be useful today. They are designed by experts who make sure that the texts that the readers are reading are simple enough for the desired level and interesting enough for a broad audience. One of their advantages is that they have exercises which are based on the texts that the readers read. Over the years, the textbooks have become more colorful, they even come with complementary audio lessons, but in their essence they still are a collection of texts with associated exercises.

We believe that one of the main limitations of the textbook approach stems from the fact that they are designed for the average learner and they can not adapt to the individual. As the US Air Force learned the hard way sixty years ago, "there is no average pilot": when cockpits, jumpsuits, and instruction were designed for the *average pilot*, the actual pilots had a hard time maneuvering the planes; performance improved only when the cockpit was designed in such a way as to adjustable to the individual.

There are many other domains where the averages are being replaced with individualised attention: medicine[1], computer security ML: cite that paper from the lab notebook maybe , and education. In this paper we are limiting our attention to a subset of education, language learning, which in itself is a very broad issue with the potential to impact the lives of a very large number of people: the British Council estimates that by 2020 there will be 2 billion people learning only English as a foreign language.

---

[1] The nascent discipline of "personalized medicine" suggests that analysis of the genetic makeup of an individual may guide health care decisions far more precisely than big group studies do

We can see a few ways in which the design of the language education system for the average is hurting the individual in the domain of reading foreign language materials:

- Average Interest. The texts that one can read are limited to generic topics, which must appeal to the entire audience. Unfortunately a text that is good enough for everybody is likely not exciting for anybody. This limits the amount of reading that learners will do, and thus, limits their learning.

- Average Knowledge. Every learner has a different prior knowlege and a different mix of already known languages.

- Average Pace. ...

The fact that textbooks do not work that well is illustrated also by the fact that some of the teachers of foreign languages that we have spoken to, do not even use a textbook anymore for their students, but instead find articles online and share them with their students ML: to ask Wim for details. This clearly does not solve the problems that we listed before, but at least, it solves the problem for the teacher, who can now find a text that is interesting for him, and he hopes willbe interesting for the students.

However, the limitation of the textbooks do not apply to the vast amounts of information available on the Internet in all the possible languages that somebody would want to learn. Blogs, News, eBooks exist for all the major languages. A student passionate about sports, might read with pleasure 10 articles of sport rather than one about "Maria who is a babysitter in Spain". However, there are three problems that prevent readers to read materials on the Internet:

- The materials might be too difficult for them. The articles in the daily *Neue Züricher Zeitung* have a very high degree of variability in their textual difficulty. A learner that picks an article randomly might choose an unpleasantly difficulty article.

- The existing reading tools might not be appropriate. Given that people are less likely to buy a product if they have to click multiple times to have it delivered to their home[2] it is very likely that people will shy away from doing more than a click to obtain a translation for an unknown word.

- The texts available on the Internet do not come with exercises that would help a learner retain newly learned words, and improve their own vocabulary.

We believe that education could be made much more personalized, and more pleasurable, if there was a way to allow the learners to:

- Express their interest in materials that are interesting for them.

- Be presented with texts which are at the right difficulty level.

- Have access to a very convenient way of translating unknown words in those cases when they are incident.

On top of that, we could also design personalized exercises that would be generated automatically based on the past reading of the students,

Finally, textbooks are also designed with exercises which help settle the words in such a way that the words get repeated.

Moreover, textbooks usually come with exercises at the end of a text. However, if one allows the readers to read whatever they like: how are they going to do exercises? Can we generate exercises that are tailored based on the past experience of students free reading?

Contributions
The contributions of this paper are:

- A minimal architecture that enables such a system

- Results from deploying the system with seventy students for about one month

---

[2]Ergo the Amazon patent for 1-click buy, and the fact that Apple licensed it from them

THE SYSTEM

For the long term, we envision an open educational ecosystem in which different creators can integrate their applications by interacting with a core API that provides the basic contextual translations, user knowledge estimation, and recommendations for words to be studied and texts to be read [2].

However, ~~in order to bootstrap, experiment with, and~~ show the potential of such an ecosystem, we present here our basic implementation of each of the components. In particular we describe our own implementation of:

Text Recommender – which consists of a feed subscription mechanism, an article browser that presents the results of crawling the selected feeds that the user is interested in

Text Reader – which is a web-based interactive text reader which provides seamless in place translations

Vocabulary Trainer – which consists of an exercise platform which generates exercises based on a reader's past reading experience

The components that we present here are implemented using HTML5 and Javascript technologies for the front end and Python and Flask for the backend. For the frontend, we have experienced in the past with writing native applications, however, the interaction with the native elements was too clumsy. Moreover, maintaining multiple systems for multiple platforms is too expensive for an academic environment. The decision turned out to be practical since the users of our system come from multiple platforms.

The Text Recommender

ML: make sure that this flows well for the reader while also keeping it short The reading recommender is simple at the moment. It consists of a Feed Subscription and an Article Browser.

Feed Subscription
When the user indicates that they would like to subscribe to a new source, as explained above, the subscription dialog is displayed. Zeeguu categorizes feeds by their language, and thus we allow users to select any language available and retrieve a list of that language's sources. Languages are represented by flags, as their compact and iconic representation should be universally understood.
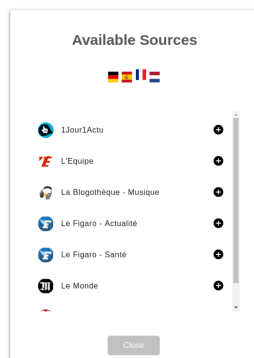

Figure 1. Different users subscribe to different sources

Article Browser
Article listing presents the source, a summary of the article, and an estimated difficulty level of the article.

In order to properly visualize the reading difficulty of an article in an intuitive manner, there are three levels of information displayed here. First we display a flag representing the language of the article. This is so because a learner could be actually registered to feeds in multiple languages. Second, we allow the user to rapidly judge difficulty on an intuitive level by color coding the difficulty from green to yellow to red. When a particular article has grasped the user's attention, we allow for a more cognitive judgment by scoring the article from 0 to 5 in difficulty.

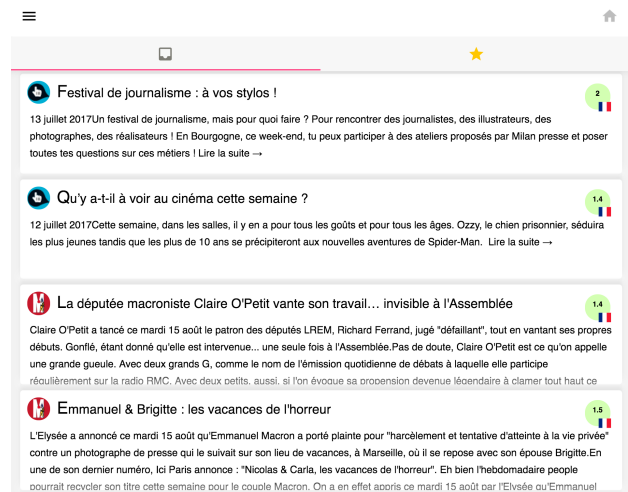ML: TODO: write about how is difficulty estimation done currently

Figure 2. Article listing presents the source, a summary of the article, and an estimated difficulty level of the article

The Text Reader

The goal of the reader is to make reading as facile as possible. To do this we optimized for the most frequent action that a reader might want to perform. Translating a word.

We explored several types of interactions and we settled on the following: a user clicks on a word, a translation is inserted right after the word, as Figure 3 illustrates:

Figure 3. A translated word is inserted after the tapped word.

Ohter alternatives that we explored and eventually dropped for each had disadvantages were:

- Showing a popup of the translation, and then hiding it again. This had the disadvantage of in the case of a more difficult sentence the reader forgetting the word at the begining of the sentence by the time he arrived to the end, and having to re-translate it.

- Piggybacking on the native selection mechanism. We experimented with allowing the learner to select a word in the same way this is normally done on the corresponding platform and to add the translation as an option in the corresponding popup menu. The problem with this is that native selection is slow, requiring about one second before it is activated.

Chaining Multiple Adjacent Translations
The user can chain a few consecutive words into a single translation by simply tapping adjacent words which are automatically merged in a translation bubble (Figure 4). This is useful for collocations and in cases where by expanding the translated set of words the precision of the translation increases.

La vicepresidenta del Vice president of Gobierno, Soraya Sáenz de Santamaría, ha advertido este martes ante la pretensión de los soberanistas catalanes de aprobar una ley que en 48 horas permita la declaración de la independencia, que "al Estado le bastan 24 horas para recurrirla y obtener su paralización".

La vicepresidenta del The Vice-President of the Gobierno, Soraya Sáenz de Santamaría, ha advertido este martes ante la pretensión de los soberanistas catalanes de aprobar una ley que en 48 horas permita la declaración de la independencia, que "al Estado le bastan 24 horas para recurrirla y obtener su paralización".

Figure 4. When adjacent words are tapped the translation bubble is extended accordingly

This minimalistic interaction model serves a double purpose - it enables and eases the translation of several chained words but it discourages users from translating entire sentences or phrases. This is good because it is in line with the recommendations of the literature (e.g. Renandya argues that extensive reading should discourage intensive use of translations[5]) but also because it reduces the amount of characters which are being translated by the learner (and thus the costs of the system, since some of the translation services have a per-character fee).

One of the limitations of this interaction is that it is not clear (at least at the moment) how to expand it for the situations in which expressions are present that are composed of words which are not adjacent (e.g. particle verbs in German and Dutch).

Alternate Translations
Due to the limitations of machine translations, it can be the case that multiple translations might seem appropriate in a context. In such a case the system will insert the most likely alternative but allow the reader to discover the others. With a click on the translation, a drop-down menu appears in which alternatives are presented. Besides the alternatives, there's also an input box in which the learner can provide his own version if none of the offered alternatives are correct.

Based on telemetry from the system it seems that ML: must check and limit this only to the french-learning kids the alternative translations menu is opened one in every eight translations.

Para ella, eso ha ocurrido Occurred cuando los catalanes han votado dentro... ...ción, cuando se ha respetado la neutra... ...y cuando cualquier voto de los catala... ...nstruir el marco constitucional.

    Occurred
    happened
    took place

Figure 5. A translated word is inserted after the tapped word.

Pronounciation
One of the important features, which was suggested by early beta-testers and added to the system after that, was the option of having pronounciation of the given word. Currently this is the action associated with tapping on the highlighted translated word. Although no user has yet complained about it, this means that a user can not pronounce a word without it being translated first. It might also be that for some languages this is more important than for others, and we just did not have users learning those languages (e.g. Danish is notoriously hard to pronounce). In the future we plan to expand the interaction modes to allow pronounciation to exist seperately from translation.

The Vocabulary Trainer
Given the list of words that a user does not know we can generate exercises for them based on their past readings.

The various interactive elements (IEs) that are present in this exercise (and in some of the other exericses are):

A hint button (IE1) will present the user with the correct answer.

Check the answer (IE2) will verify the correctness of the answer.

Word pronunciation (IE3) will sound the pronounciation.

Control (IE4) over the exercise card allows for reporting or deleting the exercise.

Input box (IE5) allows for entering a solution.

Words Good for Study
The words good for study are the ones that are either starred by the user, or are important and of quality based on a set of heuristics.
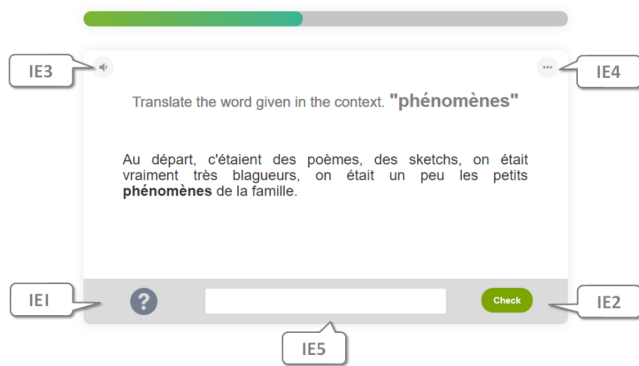
Figure 6. One of the exercise types with which the user is presented. Multiple exercise types are taken from the users past reading context

Important words are the ones which appear frequently in the language.

Quality words are most of the times single words or at most two adjacent words. They come with a context which is not too short but not too long.

Scheduling Exercises

The scheduling algorithm is based on an adaptive, response-time-based scheduling algorithm [was developed] to increase the efficiency of perceptual learning by Mettler et al. [4]. After evaluating several alternative scheduling strategies we settled on the Mettler one since it has been proven to have gains with both familiar, seen items as well as with new, unseen instances and the benefits of adaptive scheduling were present at an immediate test as well as at a delay [4].

One of the problems with this is that sometimes the context is too long and sometimes

Translation Service

The translations are provided by our server. The main advantage of this indirection is that this allows the server to track the words that are looked up and the context (sentence) in which they are being looked up. This information is then used for estimating learner knowledge and for generating later personalied exercises.

To avoid depending on a single service and to also increase the likelihood that at least one of the alternative translations is the correct one, the translation service dispatches in parallel requests to at least three third party translation APIs: Google Translate, Microsoft Translate, and Glosbe – a free translation API. The first two provide contextual translations and multi-word translations, while the third is a simple dictionary.

The dependency of the translation service on multiple third party APIs allows for a higher reliability and a chance to guarantee a low response time: when a service is down or too slow to respond, the results from it are ignored.

The Teacher Dashboard

Although not the focus of this paper, the system has also a dashboard for the teacher. The teacher can see the history of what his students have read.

THE TESTING ENVIRONMENT

The population that we tested our infrastructure with consists of sixty students from a public highschool in Groningen, The Netherlands. Their native language is Dutch but all speak very good English. They represent three classes of students that have the same language teacher.

At the begining of June 2017, we visited the school, and during one hour we introduced the way the tools and their usage in each of the three participating classes.

The system was used officially in class from the second week of June until the end of the month. With one or two exceptions all the students created an account and started using the system the latest on June 9th.

The teacher asked the students to use the infrastructure as much as they liked while preparing at home and to write reports on their activity. For evey half an hour of usage, the students would have to write a brief report that they would submit to the teacher about how they spent their time. The teacher could then decide to selectively test them on the basis of their self-reported activity cards.

We deployed the system with the translations from French to English instead of Dutch since, based on our observations, translations are of higher quality between those two languages and because the students and their teacher were confortable with the idea. We made it clear to the students that if for some of them this is not convenient, they can ask us, and we will modify their personal account in such a way as to receive translations in Dutch. However, none of the students requested this.

We also asked the students to feel invited to send us feedback at any time during the deployment of the system if they encounter problems, or alternatively have ideas for improvement. Several of them emailed us. Towards the end of the month, we also deployed several focused questions using a customer opinion elicitation service called HotJar.

The Students

Before creating accounts on our platform, the participants were directed to a survey form which asked them to provide personal information about their current level of knowledge, learning strategies, and interests.

The participants in our study are 54 female and 15 male with ages below 18 representing three different classes. Based on their own self characterization, 53 students are level B1 (i.e. can understand the main points of clear standard speech, can narrate an event, an experience or a dream) and 16 are level A2 (i.e. using simple words, can describe his or her surroundings and communicate immediate needs).

When asked whether they have favorite topics they would like to read about, half of the students mentioned such topics while the other half did not answer the question. From the topics that they mentioned as possible interests some of the more popular were: sports, music, travel, lifestyle, fashion, movies, and somebody mentioned as interest *no politics*.

We seeded the system with a variety of news french sources and blogs that cover these aspects: 1Jour1Actu, L'Equipe, La Blogoteque, Le Figaro, Le Monde.

Feed Subscriptions

Figure 10 represents an incidence matrix collected at the end of the study interval: the columns represent students, and the rows represent news feeds; if a student is registered to a given feed, at the intersection of the corresponding row and column we place ◇.

We would expect to see fully continuous horizontal rows of data-points if every user subscribed to the same feed, and fully continuous vertical rows if every user subscribed to all of the feeds available. The notion that these patterns are largely absent in Figure 4 supports our assumption that different individuals prefer to subscribe to different reading sources.

The figure illustrates that giving the students the freedom to choose the sources they wanted, allowed each one of them to express their interest.
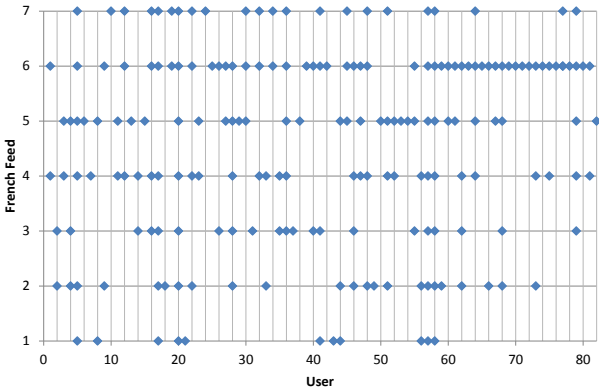

Figure 7. Different users subscribe to different sources

Of course some feeds are more popular than others. Projecting the data- points into the vertical axis and sorting the results leaves us with a histogram as can be seen in Figure 5.
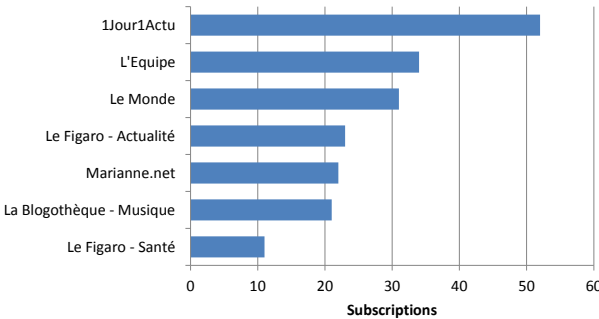

Figure 8. Some feeds are more popular than others

Feed *1Jour1Actu* is the most popular French feed, and feed Le Figaro - Santé is the least popular French feed. In order to see

whether or not this might be related to how they are presented in the dialog window of our system (see Figure 1), we can compare the order of popularity with the order in which they are displayed.
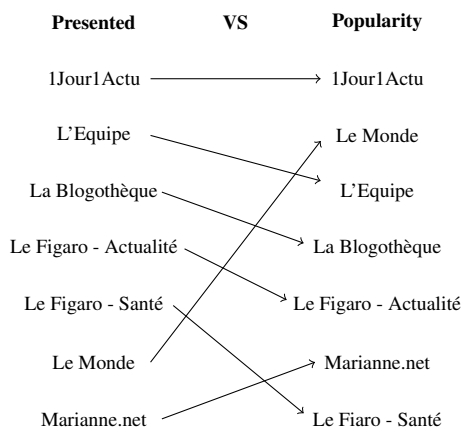

Figure 9. The popularity of the feeds vs. their ranking in the UI

One can see how the second-to-last presented feed, Le Monde, is the second most popular feed by measure of subscriptions. Conversely, the feed listed above Le Monde is actually the least subscribed-to feed in our listing.

Article Interactions

If we investigate the articles that the users interact with, we see the same pattern: each user explores their own interest, and there is no one article that is interesting for all ~~of them.~~
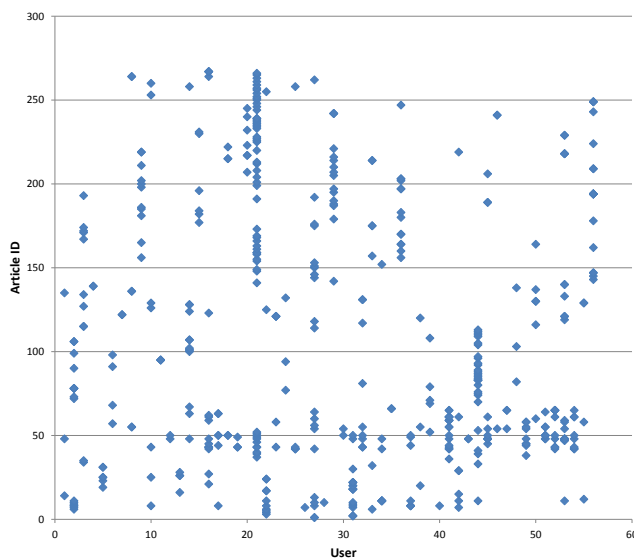

Figure 10. Every student has their own article reading preferences

WHICH OF THE INTERACTIVE FEATURES ARE MOST IMPORTANT WHEN READING?

We tracked the usage of the various features in the interactive reader. What we see is that in average, a students requests an alternate translation for one in every eight translations.

The third most used feature is the text to speech feature. This was a feature that was recommended by one of our expert beta-testers.
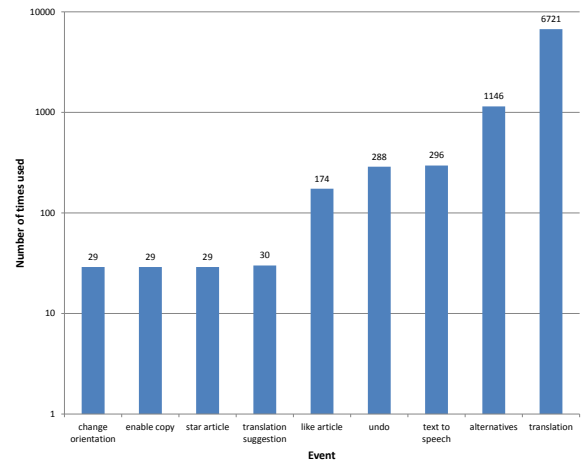

Figure 11. Popularity of features by their recorded usage-events

To see how widespread the various features are among our users, we also looked at the number of distinct users for each category of events. A larger number of distinct users indicates that the feauture is useful to the broadest number of students.

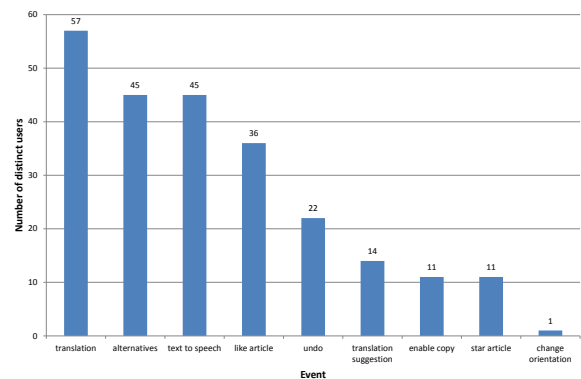> We should sort this graph similarly to the graph below. - Luc

Figure 12. The usage of the various reader features by the various users

In addition, we looked at the number of times the same word or phrase was pronounced by the same user. This data ranges from one single pronunciation to 14 pronunciations for the same word (phrase). The size of this interval is mostly due to the users' different proficiency in a certain language and the difficulty in pronunciation of the word (phrase) itself. Nevertheless, on average, the number approaches 1.66 pronunciation requests for the same piece of text, suggesting that users are generally sufficiently content with a pronunciation after hearing it the first time.

ARE THE PERSONALIZED EXERCISES USEFUL?

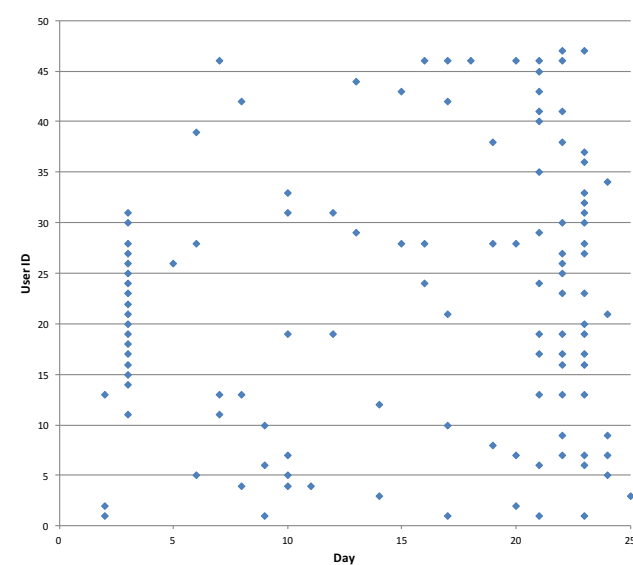The personalized exercises are a complement to the reading. To show that users actually use them ...

Figure 13. The students are doing exercises at their own pace throughout the one month interval

we could also show that they benefit from them...

USER FEEDBACK

Besides the analysis that we did based on the observed user data, we also asked the students a series of questions, among which whether they preferred the reading platform and why. Some of the qnswers can be seen in the screenshot below. It becomes clear that the students appreciate the possibility of reading what is interesting for them.

THREATS TO VALIDITY

Threat to external validity: we believe that the students we worked with are representative for the Dutch highschool student population.

We presented a system, and we showed that it has the potential to generate user involvement. However the study we performed is not sufficient to reach a strong conclusion about the impact of the system we present...

The feedback from the users was positive. However, they might have been influenced by our enthusiastic presentation of the system at the begining of the testing month.

We showed that the users are using the system extensively. However, this might be because the students had to use the system as part of their assignment in the class. We showed that the majority of the students used the system constantly throughout the one month period. If they only used it for a grade, we would have expected a more focused cramming at the end of the period (which we actually saw with a few of the students, but not with the majority).

The majority of the students who answered our post-usage survey said that they prefer our system to a textbook. However,

we still think this is not very conclusive since the number of students who answered our survey was quite limited: 12 of the 60 students represent about 20% of the participants.

The algorithms for scheduling are the state of the art in spaced repetition. However, we did not have a control group to see whether this approach works better than others. Moreover, note that other approaches for using spaced repetition already exist; what is unique in our approach is that the students learn based on personalized exercises generated based on the context of their past readings.

AVAILABILITY

The system described in this paper is deployed online and available at `https://zeeguu.unibe.ch`. If the reader of this article would want to test it they are invited to use the "CHI" code word while following the "Become a Betatester" link from the homepage.

Figure 14. The students appreciate the freedom of reading what is interesting to them

CHALLENGES

The lack of editorial work.

This is all fine and dandy, but how are we going to ensure the lack of false positives? The advantage of a textbook is the fact that the quality control is guaranteed. How are we going to ensure the quality of the texts, and the quality of the generated exercises?

How do we automatically verify the "learnability " of an example in the context? It is a great responsibility automatically selecting a word to study. False positives are really bad. Thus, we currently have a set of filters, but

Possibilty: crowdsourcing, with more advanced users being presented with exercises which also solve the problems of beginners.

For beginners, this is still not an option. So we can only do this for students who are already quite advanced.

> We should see whether there's a difference between the ones that were A2 vs. B1

Others

- Do the learners choose the right translation?

- In this article we showed that there is interest from the learners. The question that still remains to be answered is: if we personalize also for the

- How to generate exercises automatically? It is indeed desirable to find good examples of practice exercises from past readings. Sometimes, the context in which the learner looks up a word is too long and sometimes it is too short. How estimate the quality of an exercise. One measure that we are considering is: ensuring that all the words in the context

Do the students learn different words? Do they learn at different speeds?

We are publishing the data, anonymized. This is four weeks worth of reading by sixty students.

FUTURE WORK

Evolving the System

There are several directions that we plan to investigate in the future:

- See whether allowing people to register to topics rather than sources (i.e. newspapers) would make more sense for them.

- Students are asking for a better browser – with image preview, etc.

A Bigger Study

The study with students that we report here has been done with sixty students for less than a month. Based on the experience, the teacher of the class decided that they want to introduce the system in the next academic year with a slightly larger group of students – about one hundred and thirty for the entire academic year. This will give us a wealth of data that we plan to record and analyze the usage of the system and hope to learn even more about the possibilities and limitations of such a system.

The Teacher Perspective

The system we presented here has a very limited teacher dashboard. It shows the words that the student has looked up in the text. Much more can be done here because this is something else that is unique with such a system with regards to books. It can provide very individualized statistics about the student knowledge and activity to the teacher.

It also remains to be seen whether such an approach can be combined with a method that will ensure that the teacher does not have more work, or at least not considerably more work than when they had been using a textbook.

REFERENCES

1. David Hirsh and Paul Nation. 1992. What Vocabulary Size is Needed to Read Unsimplified Texts for Pleasure? *Reading in a Foreign Language* 8, 2 (1992), 689 – 696.

2. Mircea F. Lungu. 2016. Bootstrapping an Ubiquitous Monitoring Ecosystem for Accelerating Vocabulary Acquisition. In *Proccedings of the 10th European Conference on Software Architecture Workshops (ECSAW '16)*. ACM, New York, NY, USA, Article 28, 4 pages. DOI:`http://dx.doi.org/10.1145/2993412.3003389`

3. C McCarthy. 1999. Reading theory as a microcosm of the four skills. *The Internet TESL Journal* 5, 5 (1999), 1–6.

4. Everett Mettler and Philip J. Kellman. 2014. Adaptive response-time-based category sequencing in perceptual learning. *Vision Research* 99 (2014), 111 – 123. DOI:`http://dx.doi.org/10.1016/j.visres.2013.12.009` Perceptual Learning âĂŞ Recent advances.

5. Willy A. Renandya. 2007. The Power of Extensive Reading. *RELC Journal* 38, 2 (2007), 133–149. DOI:`http://dx.doi.org/10.1177/0033688207079578`