

Automatic Input Enrichment for Selecting Reading Material: An Online Study with English Teachers

Maria Chinkina Ankita Oswal Detmar Meurers

Universität Tübingen
Department of Linguistics, ICALL research group*
LEAD Graduate School & Research Network
{mchnkina, aoswal, dm}@sfs.uni-tuebingen.de

Abstract

Input material at the appropriate level is crucial for language acquisition. Automating the search for such material can systematically and efficiently support teachers in their pedagogical practice. This is the goal of the computational linguistic task of automatic input enrichment (Chinkina and Meurers, 2016): It analyzes and re-ranks a collection of texts in order to prioritize those containing target linguistic forms. In the online study described in the paper, we collected 240 responses from English teachers in order to investigate whether they preferred automatic input enrichment over web search when selecting reading material for class. Participants demonstrated a general preference for the material provided by an automatic input enrichment system. It was also rated significantly higher than the texts retrieved by a standard web search engine with regard to the representation of linguistic forms and equivalent with regard to the relevance of the content to the topic. We discuss the implications of the results for language teaching and consider the potential strands of future research.

1 Introduction

Input material at the appropriate level is important for language learners – whether it is a revision of the already acquired linguistic forms or an introduction of the structures to be acquired next, in line with the input hypothesis by Krashen (1977). Automating the search for such material can systematically and efficiently support teachers and is the goal of the computational linguistic task of automatic input enrichment (Chinkina and Meurers, 2016): It provides reading material containing target grammatical and lexical forms by analyzing and re-ranking a collection of texts. Automatic input enrichment systems rely on rigorous NLP analysis of texts

provided either by a search engine or by the user. As a result, the most linguistically appropriate texts are prioritized and presented to the user.

Automatic input enrichment is in essence closely related to the notion of input flood substantially motivated and discussed in second language acquisition research (Trahey and White, 1993) and is a necessary step in providing any type of text-based activities for language learning. It has been shown that a richer representation of target linguistic forms in the input leads to a better acquisition of these forms by the learner (Pigada and Schmitt, 2006). However, the benefits of input flood for language teachers have not been empirically tested so far.

In order to fill this gap, we developed an online study investigating whether English teachers preferred automatic input enrichment, or input flood, over web search when selecting reading material for class. The study implemented a repeated measures design: Participants read and rated 20 news articles on ten different topics. The articles were presented in pairs, with one of them being the top search result retrieved by a standard search engine and the other one provided by an automatic input enrichment system. A topic and a pair of target linguistic forms were kept constant for each pair of articles. The repeated measures design allowed us to collect a sufficient number of responses (n=240) discriminating different types of linguistic forms.

We start by reviewing the relevant research from the field of second language acquisition in Sec. 2 and dwell on the importance of automatic input enrichment for language teaching and its practical implementation in Sec. 3. We then describe the design of the current study and the obtained results in Sec. 4 and discuss the findings in Sec. 5. Finally, we conclude with the implications of the results and ideas for further research in Sec. 6.

2 Motivation and Related Work

Research on second language acquisition has provided insights on effective language teaching and learning techniques. The role of comprehensible input (Krashen, 1977) has been emphasized by many researchers, and extensive exposure to written input has shown positive effects on vocabulary (Krashen, 1989; Waring and Nation, 2004) and grammar acquisition (Pigada and Schmitt, 2006).

While stressing the importance of input, researchers agree that in order for the learner to acquire a linguistic form, it has to be frequent and salient enough in the input (Slobin, 1985). At the same time, the learners should be provided with pedagogical support to notice (Schmidt, 1990) and process the forms (VanPatten, 1990).

The effectiveness of activities targeting certain linguistic forms has been thoroughly investigated by second language acquisition researchers: According to Long (1991), focus on form instruction encourages learners to attend to form within a communicative classroom environment, which has proved to be superior to purely communicative instruction (Leeman et al., 1995). Pointing out the importance of systematic focus on target linguistic forms, VanPatten and Oikkenon (1996) found that contextualized practice activities were more effective than explicit explanations of rules for intermediate learners of Spanish. In a meta-review of research on reading and second language acquisition, Chio (2009) also emphasized the potential of supplementing reading with discussion or interactive activities targeting certain linguistic forms.

Either incidentally drawing learners' attention to certain vocabulary and grammar or providing exercises targeting those, all of the aforementioned approaches rely on the existence of appropriate reading material with a rich representation of linguistic forms for effective language acquisition. The following section provides information on how language teachers can efficiently search for such material.

3 Automatic Input Enrichment for Language Teaching

Automatic provision of reading material for language learners has been guided by text complexity (Vajjala and Meurers, 2012), lexical and grammatical properties (Brown and Eskenazi, 2004; Bennöhr, 2007), and the learner's language proficiency (Collins-Thompson et al., 2011).

We refer to automatic selection of lexically and grammatically appropriate texts as automatic input enrichment and approach it as a web search task (Chinkina and Meurers, 2016). We developed a linguistically aware web search system FLAIR¹ that provides automatic input enrichment of certain lexical and grammatical forms by detecting them in a collection of texts and reordering the texts accordingly. This process can be seen as vocabulary and grammar retrieval.

Vocabulary retrieval is indeed the core of any web search engine: One obtains an appropriate text containing target lexical items by including them in a search query. Grammar retrieval, on the other hand, requires an extension to web search as the user is unlikely to find appropriate texts by simply searching for, e.g., *texts containing present perfect*. Such an extension is implemented in FLAIR as an algorithm detecting linguistic forms relevant for English learners, such as regular and irregular verb forms. The heatmap at the top of Fig. 1 demonstrates that although these two linguistic forms are highly frequent, they are not equally represented across the top 60 search results retrieved by Microsoft Bing.² The heatmap at the bottom of the same figure shows the result of automatic input enrichment by FLAIR: a re-ordered list of the same search results with those containing the best representation of both regular and irregular verbs closer to the top (i.e., to the left in the figure).

FLAIR is built on top of a web search engine Microsoft Bing, relies on third-party tools for text extraction and parsing, detects 87 linguistic forms from the grammar section of the official curriculum of English, and uses a ranking algorithm for prioritizing texts containing the target linguistic forms specified by the user. Once the user has typed in a search query, specified the target linguistic forms and a number of search results to retrieve, they receive a list of web pages, with those that contain the best representation of the target forms at the top of the list. The user can then explore the retrieved texts with the highlighted target linguistic forms and select the texts of appropriate complexity and length (see Fig. 2).

We used FLAIR to find out whether teachers benefit from automatic input enrichment, as compared to a standard web search engine, when

¹www.purl.org/icall/flair

²www.bing.com

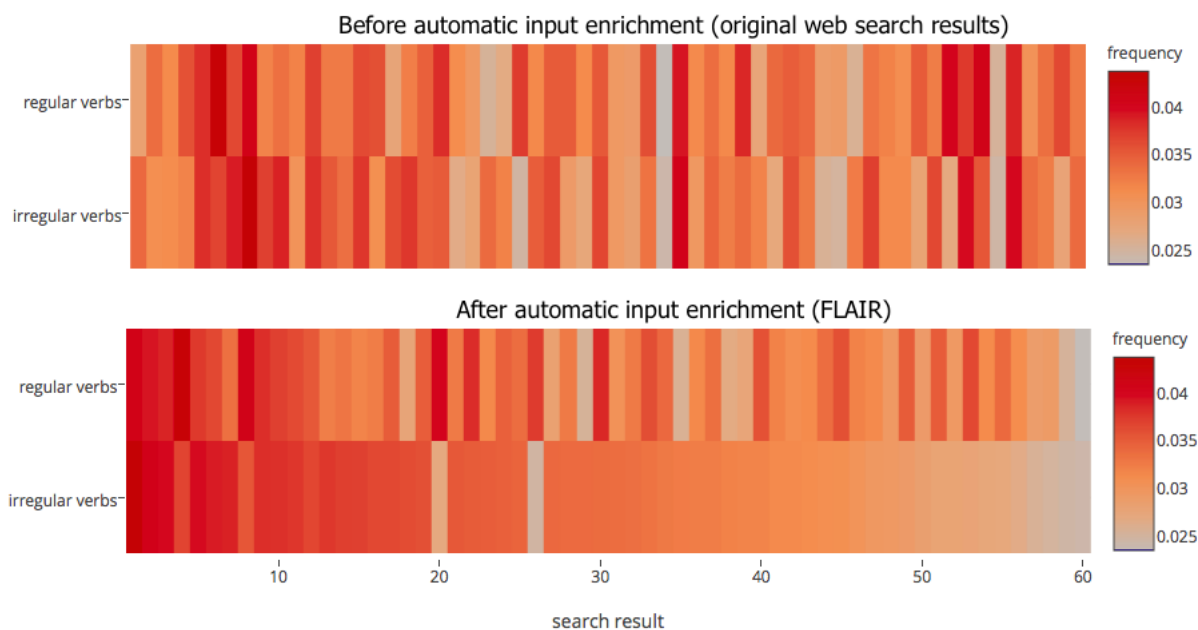


Figure 1: Comparison of the top results retrieved by a standard web search engine before and after automatic input enrichment. The 60 search results are plotted along the X axis, and the two target linguistic forms, regular and irregular verbs, are plotted on the Y axis.

searching for reading material for their students. The following section presents our research questions and hypotheses, the design of the online study, and the results.

4 Automatic Input Enrichment vs. Web Search for Selecting Reading Material

The current study focuses on teachers as media between students and reading material. It assesses teachers' experience and satisfaction with the every-day task of searching for supplementary texts online and provides insights on this process.

The *research questions* of the study address the importance of content and linguistic form and teachers' attitude towards their optimal balance: Does automatic input enrichment succeed in giving teachers the material that:

- is enriched with target linguistic forms relevant in the context of language learning,
- is in line with the information need expressed via a search query, and
- is suitable as a reading assignment for their students?

The online study was designed to operationalize these research questions. In the study, news articles retrieved by the standard web search engine

Microsoft Bing were compared to those provided by the automatic input enrichment system FLAIR. As FLAIR relies on Bing for retrieving web pages, the study in fact evaluates the impact of the NLP-driven re-ranking provided by FLAIR. The following *hypotheses* guided the design and the contents of our study:

H1: Teachers prefer texts provided by FLAIR over those provided by Bing when choosing a reading assignment for their students.

H2: Texts provided by FLAIR are perceived to have a richer representation of target linguistic forms than those provided by Bing.

H3: Texts provided by FLAIR are perceived to be less relevant to the topic than those provided by Bing.

H4: The more infrequent the target linguistic forms are, the more teachers prefer texts provided by FLAIR over those provided by Bing.

4.1 Design of Online Study

In order to address the aforementioned hypotheses, we designed an online study where the participants were asked to rate and compare pairs of news articles: One was the top search result from a standard search engine and the other one was a search result prioritized by FLAIR after specifying the target linguistic forms. Each article had to be rated on two scales: (i) its relevance to a given

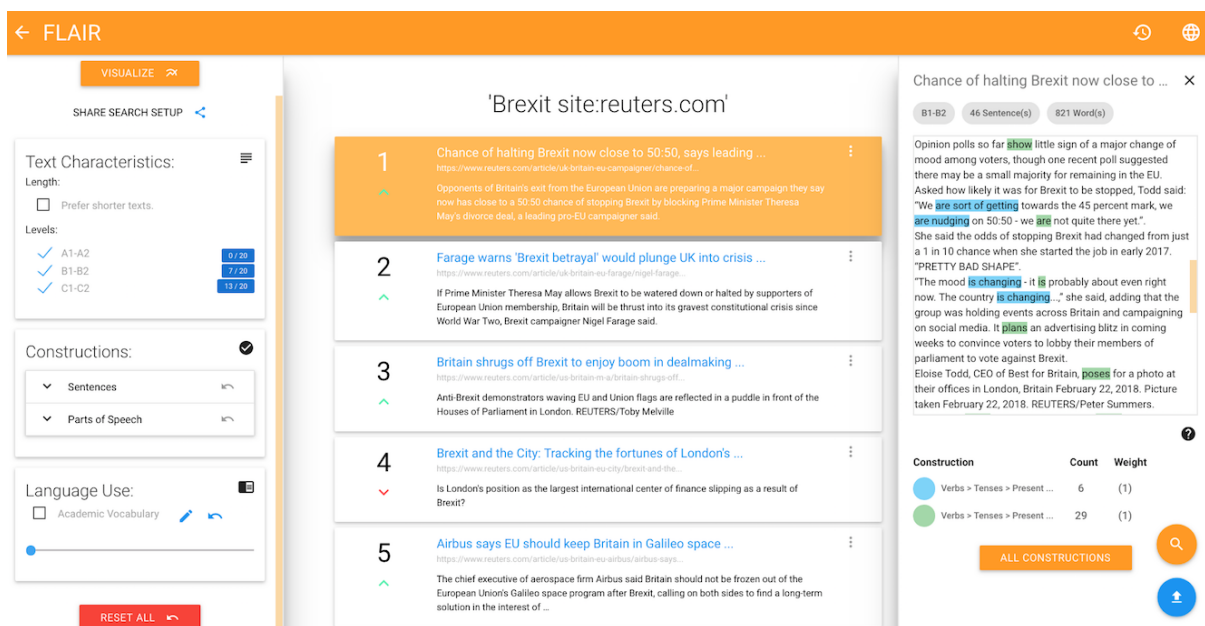


Figure 2: FLAIR interface.

topic and (ii) the representation of given linguistic forms in it. These two criteria are an integral part of language teachers' pedagogical practice: Teachers want to expose their students to language richly containing the structure to be taught or revised using a text that is on a topic that is relevant and motivating to the students.

We opted for a repeated-measures within-subjects design and ensured a random order of news articles retrieved from Bing and FLAIR as well as a random combination of topics and pairs of linguistic forms in the main task. The study proceeded as follows.

Procedure Participants received a message with the link to the online study and were asked to carefully read the information for the participants and the consent form before registering. Upon registration, they filled out a short questionnaire asking for their age, gender, native language(s), English language proficiency, the highest degree in teaching, and the proficiency level(s) of their students. They were also asked whether they used web search to look for reading material for their classes. Once they submitted the answers to the questionnaire, they could read the detailed instructions, which were displayed on every login.

The flow of the main task is demonstrated in Fig. 3: Participants were presented with a topic and a pair of target linguistic forms. They read and rated each of the two provided news articles by an-

swering two questions and were asked to pick one article as a reading assignment for their students with a preference scale from *Definitely Text 1* to *Definitely Text 2*.

Once they have completed the ten topics, participants filled out a debriefing questionnaire, where they explained general strategies for answering each of the questions in the main task (e.g., *How did you decide on the relevance of an article to a given topic?*). Finally, they submitted their email address and received a 20 Euro voucher as reimbursement.

4.2 Implementation of Online Study

We implemented the online study as a Java J2EE web application. To ensure anonymity, the user personal information obtained from the questionnaire was stored separately from their responses. Upon registration, each user was assigned a list of ten topics in a random order. Each topic was randomly matched with one of the three types of linguistic forms (see Sec. 4.3 below), one news article provided by FLAIR and one news article retrieved by Bing. For each topic, the two articles were displayed in a random order, and participants could not change their rating of the first news article once the second one was displayed.

4.3 Data and Participants

The total of 60 news articles were used in the study. The texts were presented in pairs that shared

Topic 4/10

Topic: News

Linguistic forms: Comparative degree of short adjectives and adverbs (*smarter*)
 Comparative degree of long adjectives and adverbs (*more intelligent*)

Text 1

"Title of the news article 1..."

1. How relevant is the article to the topic?
 (irrelevant) 1 2 3 4 5 (relevant)

2. How rich is the representation of the two target linguistic forms in the article?
 (poor) 1 2 3 4 5 (good)

Text 2

"Title of the news article 2..."

1. How relevant is the article to the topic?
 (irrelevant) 1 2 3 4 5 (relevant)

2. How rich is the representation of the two target linguistic forms in the article?
 (poor) 1 2 3 4 5 (good)

Question

Which news article would you give as a reading assignment to your students?

Text 1: "Title of the news article 1..."

Text 2: "Title of the news article 2..."

Definitely Text 1 Likely Text 1 Doesn't matter Likely Text 2 Definitely Text 2

Figure 3: The main task in the online study included reading and rating two news articles and selecting one of them as a reading assignment for class.

the same topic (e.g., *Brexit*) and the same pair of target linguistic forms (e.g., *the present simple* and *the present continuous tenses*). One article in each pair was obtained by submitting a search query to the web search engine Microsoft Bing and selecting the top search result. The other article in each pair was obtained by submitting the same query to FLAIR, configuring the settings to prioritize texts with the two target linguistic forms and selecting the top search result from the re-ranked list. As FLAIR relies on Microsoft Bing for retrieving the original search results, the only variable that differed between the two conditions was the automatic input enrichment component implemented in FLAIR.

Linguistic forms For the current study, we selected three pairs of linguistic forms (frequent, mixed, and infrequent) based on their document co-occurrence frequency in a corpus of 2400 news articles. Table 1 provides the distribution of their mean relative term frequencies across the texts provided by Bing and FLAIR.

The *frequent* pair was represented by regular (e.g., *typed*) and irregular (e.g., *wrote* – *written*) verb forms. It had a high document co-occurrence frequency of 95%. This means that these two linguistic forms occur together in 95 out of 100 documents, on average. Both constructions are also highly frequent: in the texts chosen for our study, regular and irregular verbs both had an average relative term frequency of 0.016. We did not count those forms when they occurred in modifier positions (e.g., *is interested*, *coloured balloons*).

The *mixed* pair of linguistic forms was represented by two grammatical tenses, present simple (e.g., *Kate plays guitar*.) and present continuous (e.g., *Kate is playing guitar now*.). Their respective relative term frequencies in the study were 0.012 and 0.003, with their document co-occurrence frequency being 50%. Predicates containing modal verbs were not counted as the present simple tense (e.g., *He can swim*.), with the exception of the verbs *have to*, *need*, and *want*. When a form constituted a part of a conditional sentence, it was not counted either (e.g., *I will not go out if it is still raining*.).

The *infrequent* pair was represented by the comparative degree of short adjectives and adverbs (e.g., *nicer*) and that of long adjectives and adverbs (e.g., *more beautiful*). In addition to only co-occurring in 4% of documents, these linguistic forms had low term frequencies of 0.002 and 0.001. When the comparative form *more* occurred as part of a longer form (e.g., *more intelligent*), the whole expression was counted as a long form, and *more* was not additionally counted as a short one.

Texts Using Microsoft Bing, we did a web search for Reuters³ news articles by expanding the search query with *site:reuters.com*. The following ten topics popular on Bing at the time served as search queries: Game of Thrones, healthcare, street artists, Roger's Cup 2017, SpaceX, electric cars, Bitcoin, Venezuela coup, Brexit, opioid epidemic. The top result for each topic was stored in our database as a Bing result, and the top 20 results

³www.reuters.com

		Bing	FLAIR
frequent	regular verbs	0.012	0.020
	irregular verbs	0.012	0.019
mixed	present simple	0.011	0.014
	present continuous	0.001	0.005
infrequent	comparative d. of short adj. and adv.	0.001	0.003
	comparative d. of long adj. and adv.	0	0.001

Table 1: Mean relative term frequencies of the linguistic forms used in the study across the top search results provided by Bing and FLAIR.

were used for further reordering.⁴ For each topic, we repeatedly configured the FLAIR settings to prioritize texts containing each of the three pairs of linguistic forms presented above and stored the three top hits as FLAIR results. In the end, we had three pairs of news articles per topic: One was the top web search result from Bing and the other one was the top one from FLAIR. The two texts for a given topic and a given pair of linguistic forms were of comparable length (the difference was at most 50% of the shortest article) and at the same or adjacent readability levels calculated using a simple Automated Readability Index (Senter and Smith, 1967).

Participants Twelve English teachers working with upper-intermediate and advanced learners of English in Germany were recruited through university and social media channels. Each participant was reimbursed with a 20 Euro voucher, and all 240 responses were anonymized.

The ages of the participants ranged from 25 to 59 years old, 91% of them being women. The first language of the majority of the participants was German (75%) followed by English (8%), French

⁴The number of texts to be retrieved can be configured in the interface. Fig. 1 presented the top 60 results for demonstration purposes. In practice, 20 results are quite heterogeneous and provide a good balance of sufficient variability and speed of analysis.

(8%), and Spanish (8%). All participants had an advanced level of English proficiency and a degree in teaching English. They worked at a secondary school (50%), a high school (42%), or a university (8%). The majority (75%) specified that they were using web search to look for reading material for their students, and 25% said they sometimes used web search for this purpose.

4.4 Results

All the analyses were conducted using R version 3.2.1 (R Core Team, 2009). Packages for individual tests and models are specified in the footnotes.

First, we compared the general preference for FLAIR to that for Bing. The option *Doesn't matter* was selected 25% of the time, and the corresponding responses were not included in the analysis. A chi-square test⁵ revealed a significant preference for FLAIR: Participants chose it over Bing 71% of the time; $\chi^2(1) = 16.04, p < .001$. They were also more confident in choosing FLAIR: The answer *Definitely* was selected three times more for FLAIR than for Bing; $\chi^2(1) = 12.60, p < .001$. Thus, our first hypothesis could be confirmed: Teachers indeed preferred the linguistically enriched texts provided by FLAIR over those provided by Bing when choosing a reading assignment for their students.

We conducted two logistic regression analyses⁶ to investigate how texts provided by FLAIR and Bing compared in terms of (i) representation of linguistic forms and (ii) relevance of the content to the topic. In line with the descriptive statistics in Tab. 1, logistic regression models showed that FLAIR ($M = 3.22, SD = 1.07$) was significantly more likely to be rated higher in terms of representation of linguistic forms than Bing ($M = 2.51, SD = 1.15$); $b = 1.89, SE = 0.51, p < .001$. Moreover, texts provided by FLAIR ($M = 3.67, SD = 1.08$) were perceived to be slightly more relevant to the topic than those provided by Bing ($M = 3.58, SD = 1.00$) although the difference failed to reach statistical significance; $b = 0.53, SE = 0.74, p = .470$.

In order to test whether the absence of statistical significance was due to chance or texts provided by FLAIR and Bing were indeed comparable with regard to content, we conducted two one-sided tests of equivalence (Schuirmann, 1987).⁷ The

⁵R native stats package, method *chisq.test()*

⁶R native stats package, method *glm()*

⁷R package *TOSTER*, method *TOSTtwo()*

results were statistically significant ($t_1 = 4.55$, $t_2 = -3.19$, $p_1 < .001$, $p_2 < .001$, 90% CI [-0.13; 0.31]), so we could confirm that the samples were equivalent with a medium effect size of 0.5 and an alpha level of .05.

Finally, we used a two-way repeated-measures analysis of variance⁸ to test whether the preference for FLAIR depended on the type of linguistic forms. We hypothesized that the more infrequent the target linguistic forms were, the more teachers would prefer texts provided by FLAIR. The first factor was the preference for FLAIR (a five-point scale), and the second factor was the type of linguistic forms (frequent, mixed, or infrequent). ANOVA did not show the tendency that we expected; $F(2, 90) = 0.87$, $p = .419$; so we inspected the means of all three groups and performed paired samples t-tests.

The biggest mean preference for FLAIR was found for the mixed pair of linguistic forms (present simple and present continuous; $M = 3.92$, $SD = 1.99$), followed by the infrequent group (comparative degree of short adjectives and adverbs; $M = 3.69$, $SD = 1.30$) and the frequent one (regular and irregular verbs; $M = 3.46$, $SD = 1.39$). When we turned the five-point scale into a binary outcome variable (i.e., either selected FLAIR as a reading assignment or not) and calculated the percentage of responses, we found 76% of responses favoring FLAIR in the infrequent group, 75% in the mixed group, and 65% in the frequent one.

As the data in the three groups were not normally distributed (Shapiro-Wilk's normality test⁹ yielded significant differences from a normal distribution), we opted for paired two-samples Wilcoxon tests.¹⁰ The paired tests revealed that there was no significant difference between the groups with regard to preference for FLAIR: infrequent and mixed groups, $Z = 128$, $p = .352$; mixed and frequent groups, $Z = 157$, $p = .643$; infrequent and frequent groups, $Z = 217$, $p = .727$.

5 Discussion

English teachers demonstrated an overall preference for FLAIR over a standard web search engine when choosing a reading assignment for their

students. This is in line with our first hypothesis and a strong argument in support of the automatic input enrichment approach.

Feedback from teachers suggested that the relevance of the article to the topic and the content of the article were the decisive factors in choosing one article over the other as a reading assignment. We were, therefore, particularly interested whether there was a trade-off between the content and the representation of linguistic forms in the articles because a large number of the news articles retrieved by FLAIR (40%) were not among the top ten original search results. Thus, we hypothesized that the texts retrieved by FLAIR would have a richer representation of linguistic forms while being less relevant to the topic.

As the number of occurrences of the given linguistic forms in the texts retrieved by FLAIR was higher (see Tab. 1), this indeed resulted in significantly higher teachers' ratings for the representation of linguistic forms. However, counter to our expectations, the texts provided by FLAIR were neither inferior nor superior to those originally retrieved by Bing in terms of content: They were rated slightly, but not significantly, more relevant to the given topic. This suggests that the most appropriate texts for language learners may not appear within the top web search results, and those texts that are not ranked high by standard web search engines can have a higher linguistic and pedagogical potential than the top hits.

As the study showed, automatic input enrichment is particularly beneficial for retrieving texts containing target linguistic forms of lower frequency levels, although the differences were non-significant. This can be explained by document and term frequencies: The high term and document frequencies of frequent linguistic forms make it likely for every retrieved text to contain at least several instances of each form. In this case, the texts prioritized by an automatic input enrichment system may not differ from the original top hits with regard to their linguistic characteristics. Other frequently co-occurring pairs of linguistic forms relevant for language teaching are, for example: adjectives and adverbs (co-occur in 97% of documents), the definite and the indefinite articles (96%), present simple and past simple (93%), *to* infinitives and *ing* verb forms (90%). We propose a way to improve the functionality of automatic input enrichment systems targeting frequent

⁸R native stats package, method *aov()*

⁹R package *dplyr*, method *shapiro.test()*

¹⁰R native stats package, method *wilcox.test()*

linguistic forms in the next section.

Infrequent linguistic forms, on the contrary, appear in few texts together, with a small number of occurrences within each text. The advantage of automatic input enrichment in this case is that it can detect those few texts containing the target infrequent linguistic forms. Other pairs of linguistic forms with a low document co-occurrence frequencies as well as low term frequencies are, for example: the modal verbs *can* and *may* (14%), past perfect and past progressive (12%), future simple and *going to* (9%), *wh-* questions and *yes/no* questions (7%), real and unreal conditionals (4%).

In case of mixed pairs of linguistic forms (i.e., the ones consisting of one frequent and one infrequent form), the reordering algorithm pushes the few texts containing the infrequent form to the top. Those texts are at the same time likely to also contain several occurrences of the frequent form because of its high term and document frequencies. Other mixed pairs of linguistic forms relevant for teaching English are: past simple and present perfect (63%), positive and comparative degrees of short adjectives (58%) and adverbs (45%), present simple and future simple (40%), past simple and past continuous (30%). The full list of pairs of linguistic forms with their co-occurrence document frequency was compiled by Chinkina (2015).

The aforementioned results show that, while relying on a standard web search engine for retrieving the results, automatic input enrichment succeeds in providing the texts that are a) enriched with respect to the linguistic forms, b) in line with the information need, and c) suitable as a reading assignment.

6 Conclusion and Outlook

In this paper, we described an online study investigating the effects of automatic input enrichment on English teachers selecting reading material for class. The results of the study show that participants preferred the texts provided by automatic input enrichment over those originally retrieved by a standard web search engine both in terms of representation of linguistic forms and content. The study also provides insights about which linguistic forms benefit the most from automatic input enrichment.

It is important to note that our goal was not to compare automatic input enrichment to web

search but to show that the linguistically motivated re-ranking of texts leverages the content and form aspects of the retrieved material. With the abundance of authentic texts available on the web, such reordering does not prioritize texts of low quality but selects the most linguistically appropriate ones in the pool of relevant texts. This means that such systems as FLAIR can rely on standard web search engines for retrieving texts of sound content. Whether automatic input enrichment systems also provide an effective learning environment for language learners should be tested in further end-to-end empirical studies.

Another interesting empirical question would be: For which kind of queries will an input enrichment system find enough texts? Our assumption is that the topics covered in a language classroom are current, prominent, and widely discussed: This is why we selected the texts on popular topics for our online study. However, when searching for texts on more specific topics – or in other less represented languages – fewer relevant texts may be retrieved and the balance of content and form may be skewed. This could be the case for courses targeting English for specific purposes, though for such courses it is likely that special repositories of sample texts from that specific domain would be used. Thus, the automatic analysis and re-ranking can be done on the provided corpus, which is also a capability of the FLAIR system.

Therefore, FLAIR provides an ecologically valid, real-life setting for an empirical evaluation of a number of phenomena discussed in second language acquisition research, such as input flood, input enhancement, structured input activities, and extensive reading. For instance, one could conduct a randomized controlled field study and compare the learning outcomes of two groups of students: one reading and working with the results re-ranked by FLAIR and the other one working with the standard Bing results. In fact, such an experimental yet real-world evaluation in essence only becomes possible thanks to a technology-enabled input enrichment approach such as FLAIR.

Finally, based on the feedback from the English teachers who took part in our study, we identified two strands for potential improvement of automatic input enrichment systems:

1. Providing a variety of contexts in which linguistic forms are used. This challenge can be addressed by the tasks of word and tense

sense disambiguation (Stevenson and Wilks, 2003; Reichart and Rappoport, 2010) that could be expanded to the disambiguation of other linguistic forms. The insights from the task of finding good dictionary examples (Kilgarriff et al., 2008) can help make sure that the contexts in which target linguistic form occur are informative, typical, and intelligible for the learner (Atkins and Rundell, 2008). This could be particularly advantageous for frequent linguistic forms that currently benefit the least from automatic input enrichment as they are richly represented across texts.

- Integration of a component that automatically generates exercises targeting the selected linguistic forms. The task of automatic question generation has explored generating factual wh- questions (Heilman, 2011), gap sentences (Becker et al., 2012), a combination of those, and grammar-concept questions asking for the meaning of linguistic forms (Chinkina and Meurers, 2017). In line with the idea of providing a variety of contexts, one could generate different types of questions targeting not only different linguistic forms but also different contexts in which those forms occur.

Acknowledgments

This research was funded by the LEAD Graduate School & Research Network [GSC1028], a project of the Excellence Initiative of the German federal and state governments. Maria Chinkina is a doctoral student at the LEAD Graduate School & Research Network.

References

- BT Sue Atkins and Michael Rundell. 2008. *The Oxford guide to practical lexicography*. Oxford University Press.
- Lee Becker, Sumit Basu, and Lucy Vanderwende. 2012. [Mind the gap: Learning to choose gaps for question generation](#). In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT '12*, pages 742–751, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Jasmine Bennöhr. 2007. [A web-based personalised textfinder for language learners](#). In *Data Structures for Linguistic Resources and Applications*, Tübingen. Gunter Narr Verlag.
- Jonathan Brown and Maxine Eskenazi. 2004. [Retrieval of authentic documents for reader-specific lexical practice](#). In *InSTIL/ICALL 2004 Symposium on Computer Assisted Learning, NLP and speech technologies in advanced language learning systems*, Venice, Italy. International Speech Communication Association (ISCA).
- Maria Chinkina. 2015. Form-focused language-aware information retrieval. M.a. thesis in computational linguistics, Seminar für Sprachwissenschaft, University of Tübingen.
- Maria Chinkina and Detmar Meurers. 2016. [Linguistically-aware information retrieval: Providing input enrichment for second language learners](#). In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 188–198, San Diego, CA.
- Maria Chinkina and Detmar Meurers. 2017. [Question generation for language learning: From ensuring texts are read to supporting learning](#). In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, page 334.
- Kit U Chio. 2009. [Reading and second language acquisition](#). *HKBU Papers in Applied Language Studies*, 13:153–174.
- K. Collins-Thompson, P. N. Bennett, R. W. White, S. de la Chica, and D. Sontag. 2011. [Personalizing web search results by reading level](#). In *Proceedings of the Twentieth ACM International Conference on Information and Knowledge Management (CIKM 2011)*.
- Michael Heilman. 2011. [Automatic factual question generation from text](#). Ph.D. thesis, Carnegie Mellon University.
- Adam Kilgarriff, Miloš Husák, Katy McAdam, Michael Rundell, and Pavel Rychlý. 2008. [Gdex: Automatically finding good dictionary examples in a corpus](#). In *Proceedings of EURALEX-08*.
- Stephen Krashen. 1977. Some issues relating to the monitor model. *On Tesol*, 77(144-158).
- Stephen Krashen. 1989. [We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis](#). *The modern language journal*, 73(4):440–464.
- Jennifer Leeman, Igone Arteagoitia, Boris Fridman, and Catherine Doughty. 1995. Integrating attention to form with meaning: Focus on form in content-based spanish instruction. *Attention and awareness in foreign language learning*, pages 217–258.

- Michael H. Long. 1991. Focus on form: A design feature in language teaching methodology. In K. De Bot, C. Kramsch, and R. Ginsberg, editors, *Foreign language research in cross-cultural perspective*, pages 39–52. John Benjamins, Amsterdam.
- Maria Pigada and Norbert Schmitt. 2006. [Vocabulary acquisition from extensive reading: A case study](#). *Reading in a foreign language*, 18(1):1.
- Development R Core Team. 2009. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Roi Reichart and Ari Rappoport. 2010. [Tense sense disambiguation: a new syntactic polysemy task](#). In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 325–334. Association for Computational Linguistics.
- R. Schmidt. 1990. [The role of consciousness in second language learning](#). *Applied Linguistics*, 11:206–226.
- Donald J Schuirmann. 1987. [A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability](#). *Journal of pharmacokinetics and biopharmaceutics*, 15(6):657–680.
- RJ Senter and Edgar A Smith. 1967. Automated readability index. Technical report, Cincinnati University Ohio.
- D. I. Slobin, editor. 1985. *The crosslinguistic study of language acquisition*. L. Erlbaum Associates, Hillsdale, NJ.
- Mark Stevenson and Yorick Wilks. 2003. [Word-sense disambiguation](#). In Ruslan Mitkov, editor, *The Oxford Handbook of Computational Linguistics*, chapter 13, pages 249–265. Oxford University Press.
- Martha Trahey and Lydia White. 1993. [Positive evidence and preemption in the second language classroom](#). *Studies in second language acquisition*, 15(02):181–204.
- Sowmya Vajjala and Detmar Meurers. 2012. [On improving the accuracy of readability classification using insights from second language acquisition](#). In *In Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 163–173, Montreal, Canada. Association for Computational Linguistics.
- Bill VanPatten. 1990. [Attending to form and content in the input: An experiment in consciousness](#). *Studies in Second Language Acquisition*, 12(3):287–301.
- Bill VanPatten and Soile Oikkenon. 1996. [Explanation versus structured input in processing instruction](#). *Studies in Second Language Acquisition*, 18(04):495–510.
- Rob Waring and I.S.P. Nation. 2004. [Second language reading and incidental vocabulary learning](#). *Angles on the English speaking world*, 4:97–110.