Linguistically Aware Information Retrieval: Providing Input Enrichment for Second Language Learners

Maria Chinkina **Detmar Meurers**
LEAD Graduate School
Department of Linguistics
Eberhard Karls Universität Tübingen
{maria.chinkina,detmar.meurers}@uni-tuebingen.de

Abstract

How can second language teachers retrieve texts that are rich in terms of the grammatical constructions to be taught, but also address the content of interest to the learners? We developed an Information Retrieval system that identifies the 87 grammatical constructions spelled out in the official English language curriculum of schools in Baden-Württemberg (Germany) and reranks the search results based on the selected (de)prioritization of grammatical forms. In combination with a visualization of the characteristics of the search results, the approach effectively supports teachers in prioritizing those texts that provide the targeted forms.

The approach facilitates systematic input enrichment for language learners as a complement to the established notion of input enhancement: while input enrichment aims at richly representing the selected forms and categories in a text, input enhancement targets their presentation to make them more salient and support noticing.

1 Introduction

Acquisition of a language directly depends on the learner's exposure to it. Hence, the importance of *input* in second language (L2) learning is systematically emphasized in Second Language Acquisition (SLA) research (Krashen, 1977; Swain, 1985; Gass and Varonis, 1994). Krashen even proposed the *input hypothesis* arguing that exposing learners to language input containing target structures is *the* single most important component of both first and second language learning. While later SLA approaches are more balanced in terms of considering input, output, and interaction (as well as implicit and explicit learning), they also further advanced our understanding of the role of input in terms of the frequency and perceptual salience of constructions needed for L2 learners to acquire a second language (e.g., Slobin, 1985; Schmidt, 1990).

We will refer to a method ensuring that a targeted structure is frequently represented in a text as *input enrichment*. While the isolated positive effect on L2 acquisition of the related notion of *input flooding* (Trahey and White, 1993) remains to be empirically substantiated (Reinders and Ellis, 2009; Loewen et al., 2009), input enrichment clearly is a meaningful component of the repertoire of language teachers. At the same time, manually searching for such reading material takes a lot of time and effort so that teachers often fall back on schoolbook texts designed to introduce the relevant constructions. This limits the choice of texts, and schoolbook texts typically are less up-to-date and in line with student interests than other authentic texts could be.

We therefore investigated how we can support teachers in selecting reading material that is (i) at the learner's level of language proficiency, (ii) in line with the teacher's pedagogical goal, and (iii) offers content of interest to the learner. The paper motivates input enrichment and presents *FLAIR*[1] (Form-focused Linguistically Aware Information Retrieval), a web search system striving to provide a balance of form and content in the search for appropriate reading material.

[1] http://purl.org/icall/flair

Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications, pages 188–198,
San Diego, California, June 16, 2016.

In terms of envisaged use cases, in the most straightforward case, FLAIR helps the teacher identify reading materials appropriate for a class or individual students in terms of form, content, and reading level. The system can also feed into platforms providing input enhancement such as *WERTi* (Meurers et al., 2010) or generating exercises from text such as *Language Muse*^{SM} (Burstein et al., 2012), ensuring that the form targeted by enhancement or exercise generation is as richly represented as possible given the text base used.

In scenarios putting more value on learner autonomy or data-driven learning, FLAIR makes it possible to distribute the specification of the form and content criteria between teacher and the learner: The teacher uses their pedagogical background in foreign language teaching and learning and their knowledge of the learner's abilities and needs to configure FLAIR in a way that prioritizes (i.e., reranks highly) the texts that best satisfy these form specifications. Using the teacher-configured FLAIR, the learner then takes control and enters search queries in line with their personal interests or information needs. The outcome is a collection of documents that was retrieved based on the learner's search query, with the results ranked according to the pedagogical language learning needs defined by the teacher.

2 FLAIR Architecture

The *FLAIR* functionality is realized using a pipeline architecture with four modules, the Web Searcher, the Text Extractor, the Parser and the Ranker:

1. *The Web Crawler* utilizes Microsoft Bing search engine[2] to retrieve the top *N* results given a query.

2. *The Text Extractor* integrates the HttpURLConnection Java library[3] to retrieve the full html code of each page and take care of redirects. The Boilerpipe library[4] then extracts plain text from it.

   The choice of Boilerpipe is motivated by the high performance of the library compared to other text extraction techniques (Kohlschütter et al., 2010). It provides several algorithms for the extraction of the main textual content from different types of web pages. We tested the DefaultExtractor, the ArticleExtractor and the LargestContentExtractor on a development collection of 50 documents, which established DefaultExtractor as the best choice for our task; the other two options extracted too little text in some cases when the main content was divided into several parts.

3. *The Parser* module employs Stanford CoreNLP[5] (Manning et al., 2014) to identify numerous linguistic forms using the syntactic category and dependency information obtained from it. We discuss this step further in the next section. Long sentences are quite frequent in web texts, so we employed the Stanford Shift-Reduce Parser, which is less sensitive to sentence length. The parser has also been reported to outperform the older Stanford constituency parsers.

4. *The Ranker* is responsible for reranking the top *N* results based on the statistical analysis of the data received from the previous modules. We chose the classical IR algorithm BM25 (Robertson and Walker, 1994) as the basis for our ranking model. An advantage of BM25 is the fact that it allows for any normalization unit and readily balances a multitude of query components. The final score of each document determining its place in the ranking is calculated as

$$G(q,d) = \sum_{t \in q \cap d} \frac{(k+1) \times \mathrm{tf}_{t,d}}{\mathrm{tf}_{t,d} + k \times (1 - b + b \times \frac{|d|}{avdl})} \times log \frac{N+1}{df_t}$$

where q is a *FLAIR query* containing one or more linguistic forms, t is a linguistic form, d is a document, $\mathrm{tf}_{t,d}$ is the number of occurrences of t in d, $|d|$ is document length, $avdl$ is the average document length in the collection, and b and k are free parameters we set to 0 and 1.7 respectively. The free parameter b specifies the importance of the document length. We used it to give the user control over the importance of document length (implemented in the interface using a slider that can take values from 0 to 1).

[2]http://www.bing.com

[3]http://docs.oracle.com/javase/7/docs/api/java/net/HttpURLConnection.html

[4]https://code.google.com/p/boilerpipe/

[5]http://nlp.stanford.edu/software/corenlp.shtml

3 Identification of Linguistic Forms

We based the identification of linguistic forms on the official school curriculum for English in the state of Baden-Württemberg (Germany).[6] The taxonomy of topics in the official curriculum defines the language skills and knowledge that the pupils are expected to acquire in the course of their studies at school; it is not tailored to one particular textbook or approach. Overall, we implemented the identification of 87 grammatical constructions integrating a broad range of morphological, lexical and syntactic properties – the full set of constructions is listed in Appendix A. As constructions motivated by language teaching and learning, they do not necessarily map directly to the standard categories that NLP tools typically identify and are evaluated on. How the two worlds were linked is discussed next.

3.1 Between shallow and deep analysis

NLP makes use of different approaches for characterizing language data, from shallow matching to deep grammar formalisms, which are equally well-motivated in language learning as application domain (Meurers, 2015, sec. 3.2). While string matching can work for some basic cases (e.g., identification of articles), the detection of other constructions requires analyses going well beyond the surface level, such as an analysis based on syntactic dependencies. Even for the seemingly simple case of distinguishing different types and cases of ***pronouns*** to retrieve subjective, objective, reflexive as well as possessive pronouns, a lexical look-up has to be supplemented with dependency parsing in order to distinguish the subjective from the objective *you* or the objective from the possessive *her*.

Taking things one step further, consider what is needed to detect the ***used to*** construction referring to a habitual action in the past. After making sure that the following word is a to-infinitive, and thus, excluding the option of misidentifying the different constructions *to be used to doing* and *to get used to doing*, one is still left with an ambiguous structure that can be either interpreted as the target construction, as in (1), or as a passive structure, as in (2):

(1) I *used to come* here every day.

(2) It is *used to build* rockets.

This ambiguity can be resolved by checking which POS tag was assigned to the verb *used*.

While some of the 87 grammatical constructions in the English language curriculum of Baden-Württemberg support relatively straightforward characterizations based on the syntactic analysis provided by the Stanford CoreNLP, others turned out to require more thought, so that we illustrate some of those in the next section.

3.2 Challenges and solutions

The identification of ***conditional sentences*** offers some interesting challenges. Narayanan et al. (2009) discuss a POS-based approach for identifying conditional types for the task of Sentiment Analysis. It mapped sequences of POS tags to tenses (VBD + VBN = Past Perfect) and further to conditional types (If + Past perfect, MD + Present Perfect = Third Conditional). However, two different types of conditionals can be used in the same sentence, producing a mixed conditional sentence, a common type not covered by this taxonomy. Puente and Olivas (2008) proposed a more granular classification of conditional sentences and an algorithm for detecting them. However, they point out that authentic texts containing conditionals pose a challenge since some retrieved sentences do not conform to their taxonomy. In order to be able to classify every conditional sentence, *FLAIR* limits itself to distinguishing two broad classes relevant for the curriculum, real and unreal conditionals.

Where two constructions are identical in form, additional analysis of the target form in context can be required. For instance, Meurers et al. (2010) employ about 100 Constraint Grammar rules to disambiguate ***gerunds*** and ***participles***, posing a challenge both for English language learners and parsers.

Real conditionals (3) and ***answers to indirect questions*** (4) are another example of ambiguity.

(3) I don't *come* if he is coming.

(4) (Do you know if he is coming?)
I don't *know* if he is coming.

In terms of the constituency and dependency structure provided by the parser, the two cannot be distin-

[6]The curricula for grades 2, 4, 6, 8, and 10 are accessible on the education portal website of the state of Baden-Württemberg: http://www.bildung-staerkt-menschen.de

guished. A simple solution based on a list of verbs followed by an *if*-clause (e.g., *know*) can help tackle this case but will not generalize to other ambiguous cases, such as different usages of ***Present Progressive*** demonstrated in (5) and (6).

(5) We are *waiting* for you.

(6) We are *leaving* next week.

Considering these two sentences, one may assume that a temporal phrase should be an indicator of the time, as in (6). There can be sentential time expressions (7), though a clause introduced by *when* will not always be a future marker (8).

(7) We are *leaving* when you are done.

(8) You are constantly *complaining* when things go poorly.

Richer NLP analyses are evidently needed to properly distinguish such cases. Either one targets relevant distinctions with specialized Constraint Grammars or supervised machine-learning approaches (e.g., Boyd et al., 2005), or one attempts a more global analysis using a linguistically rich grammar (HPSG, LFG, TAG, . . .).

In the education context, not differentiating between such ambiguous structures can mean exposing the learner to unfamiliar constructions far beyond their current level. According to the English curriculum we targeted, *Present Progressive* is introduced in the second grade, while it is only six years later, in the eighth grade, that school children are expected to use this linguistic form to express an arranged action in the future. The same applies to *real conditionals* as opposed to *answers to indirect questions* (Grades 6 and 8), *adjectives* and *quantifiers* (Grades 2 and 6), or different parts of speech ending in *-ing*, such as *gerunds* and *present participle* forms (Grades 2, 8 and 10). The *FLAIR* interface includes a reading view shown in Figure 1 that highlights and identifies the targeted constructions in a given text, so that at least for the teacher it is possible to judge on a case-by-case basis, which of the uses of an ambiguous form is part of a given text and whether it therefore requires additional explanation – or choice of another text for the envisaged audience.

3.3 Pilot evaluation of target identification

Before evaluating the identification of the linguistic target forms, we inspected the performance of the Stanford Shift-Reduce Parser for the constructions our patterns depend on. Among the biggest challenges were *gerunds* that got annotated as either nouns (*NN*) or gerunds/present participles (*VBG*). *Phrasal verbs*, such as *settle in*, also turned out to be difficult for the parser.

Turning to the target form identification performed by FLAIR on the basis of the parsed output, we conducted a pilot study using news articles as a common type of data analyzed by *FLAIR*. We submitting three search queries and saved the top three results for each of them, obtaining nine news articles with an average length of 28 sentences. Table 1 shows the precision, recall, and F-measure for selected linguistic constructions identified by *FLAIR* and the medians and means across the 81 constructions, for which details are included in Appendix A.

Linguistic target	**Prec.**	**Rec.**	**F_1**
Yes/no questions	1.00	1.00	1.00
Irregular verbs	1.00	0.96	0.98
used to	0.83	1.00	0.91
Phrasal verbs	1.00	0.61	0.76
Tenses (Present Simple, ...)	0.95	0.84	0.88
Conditionals (real, unreal)	0.65	0.83	0.73
Mean (81 targets)	0.94	0.90	0.91
Median (81 targets)	1.00	0.97	0.95

Table 1: Evaluating identification of targets by FLAIR

As the numbers show, some constructions are easily detectable (yes/no questions) while others are less reliably identified by the parser (phrasal verbs). There are different reasons for lower performance: the ambiguity of the construction (*real conditionals*) and problems of the Stanford Parser (*-ing verb forms*) discussed above, as well as problematic output of the text extractor module and some limitations of the *FLAIR* patterns used for identification (*unreal conditionals*). *Conditionals* were identified with an average low F score of 0.73 due to the difficulty of their disambiguation partially discussed in section 3.2 and a particular choice we made: In order to avoid exposing learners to an unknown grammatical construction, we disambiguated all unclear cases of conditionals as the one appearing later in

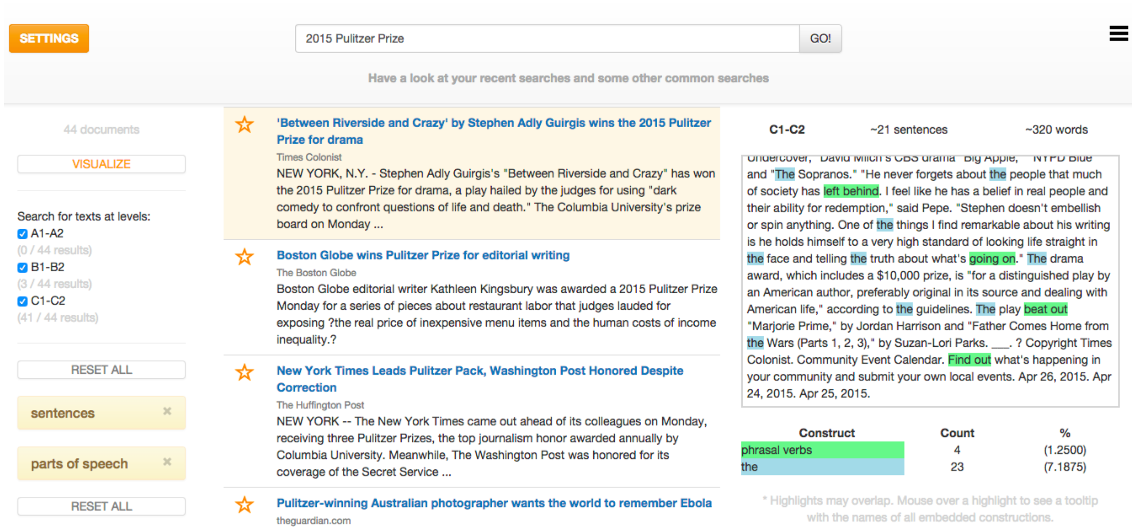

Figure 1: *FLAIR* interface: the settings panel, the list of results and the reading interface.

the curriculum, *unreal conditionals* (Grade 8). This way, any potential instances of this construction in texts at a lower level can be avoided (e.g., in Grade 6, when *real conditionals* are introduced).

4 Exploring FLAIR in use

Let us start with an example for the kind of distribution of grammatical patterns detected by *FLAIR* when analyzing the top 55 web search results returned for the query term "2016 US presidential elections". Figure 2 shows a heat map with selected constructions sorted in the ascending order by variance in their frequencies across the top 55 web search results. The figure showcases the high variability with which many of the grammatical constructions occur, which is in line with the result reported in Vajjala and Meurers Vajjala and Meurers (2013) that top web search results also differ significantly in terms of readability. This confirms that it is meaningful to rerank the top web search results in order to ensure a rich representation of specific constructions or prioritize a particular reading level.

For a more systematic exploration of the distribution of linguistic forms in web documents, we retrieved the top 60 documents for each of 40 queries using the Bing interface. In total, 2400 documents were retrieved and run through FLAIR. Among the most frequent constructions were *prepositions*, *regular* and *irregular verbs*, and *the simple verb aspect*, all of which appeared in more than 98% of the doc-

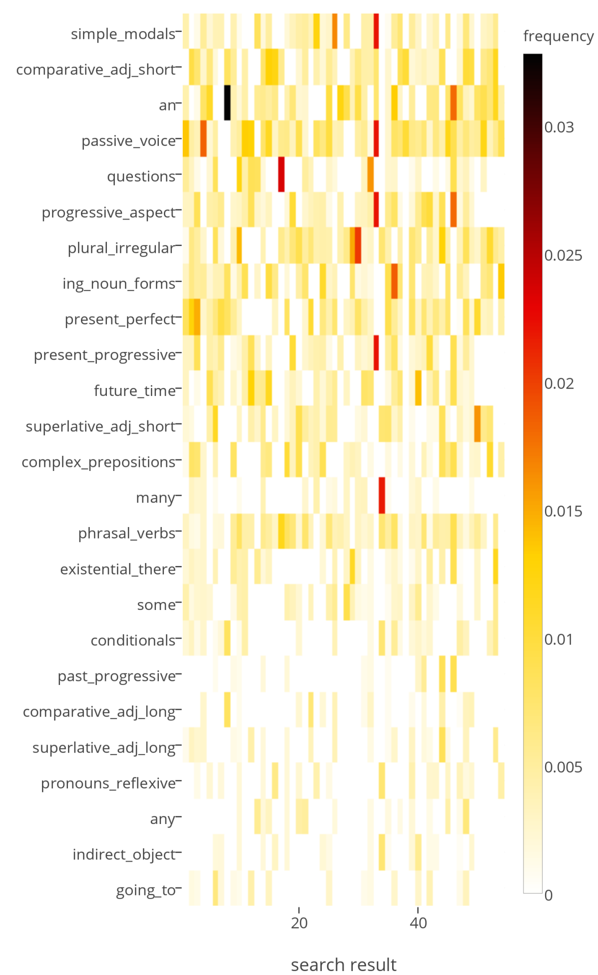

Figure 2: A heat map showing the distribution of grammatical construction across the top 55 results for the query "2016 US presidential elections" (normalization unit: document length)

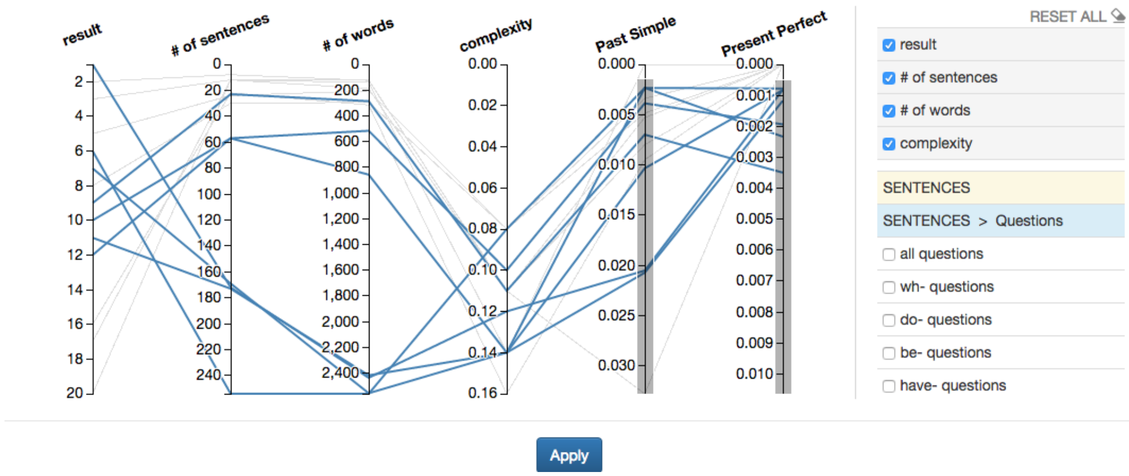

Figure 3: Interactive visualization in *FLAIR* with each line representing a document and vertical axes showing characteristics

uments. The least frequent linguistic constructions were *tag questions* (0.8%), *Past Perfect Progressive* (2.6%), and *imperatives* (3.2%).

Turning to a particular use case, it is a common teaching practice to not only expose the learners to one linguistic form but to contrast it with another one in the same context, e.g., *regular* vs. *irregular verbs*. We therefore selected 70 pairs of grammatical constructions from the book *English in Use* (Murphy, 2012) that are known to be challenging for English learners. We then calculated the document frequency for their pairwise co-occurrence in the collection of 2400 web texts. Among the most frequent construction pairs (i.e., where both forms occurred in many documents) were the following ones: *adjectives* vs. *adverbs* (96.7%), *the definite article* vs. *the indefinite article* (95.7%), *irregular verbs* vs. *regular verbs* (95.2%), and *Present Simple* vs. *Past Simple* (93.2%). Some construction pairs are not so easily found within the top retrieved results – either because of the low frequency of at least one of them or due to the fact that they occur in different documents. Among such construction pairs that had a document frequency of less than 10% were *degrees of comparison of adverbs*, *real conditionals* vs. *unreal conditionals*, and *wh- questions* vs. *yes/no questions*. The highest scoring pair of modal verbs, *can* vs. *could*, appeared in 20% of documents, with other modal pairs scoring significantly lower.[7]

[7] A more detailed analysis going beyond the space available here could use *odds ratios* to quantify how strongly the presence and absence of constructions are associated.

4.1 Interactive visualization

The *FLAIR* tool includes an interactive visual component that makes it possible to inspect and further select documents based on the multi-faceted nature of the retrieved documents. The interface illustrated in Figure 3 is based on the visualization technique of parallel coordinates used for visualizing multivariate data. Vertical axes represent parameters: any linguistic forms selected by the user, the number of sentences, the number of words and a global readability score. Each polyline stands for a document and records its linguistic characteristics by going through different points on the parameter axes. The interface supports mouse interaction allowing the user to restrict the range of values permitted for particular parameters, with other documents becoming greyed out in the interface and removed from the search results. In the figure, only documents with a non-zero frequency for both *Past Simple* and *Present Perfect* are selected. The numbers on the vertical axes for the grammatical constructions correspond to their relative frequencies in documents. Once the Apply button is selected, the search result list is restricted to those documents satisfying the constraints specified in the visualization module.

The visualization makes it possible to get an overview of the distribution of linguistic characteristics in the set of documents to be reranked. The interface also supports interaction with the visualization, providing fine-grained control over a user-selected set of linguistic characteristics. Users can

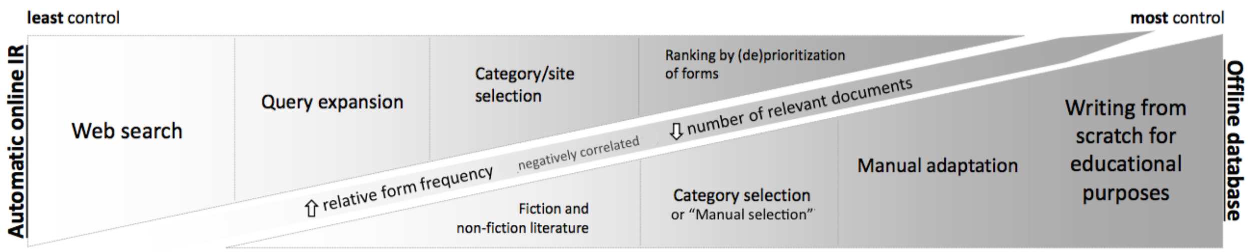

Figure 4: Strategies for input enrichment of an already existing corpus or during web search. The font size of the more common strategies is larger than that of the less common ones.

select a range of values for one or more constructions to precisely identify and retrieve documents.

4.2 Towards evaluating FLAIR in practice

Depending on a number of parameters, from the internet connection to the nature of the retrieved documents, the current *FLAIR* version takes 10 to 45 seconds to retrieve and analyze 20 web documents, making real-life use possible. Web crawling, text extraction, and NLP analysis are performed on the server in parallel for several documents, depending on the available memory and CPU power. It takes more than half of the total time (from entering the query till displaying a list of results) to fetch the results and extract the text. 20-30% of the total time are used for the NLP. Ranking is performed on the client side and takes 10-20% of the time.

As a pilot exploring whether *FLAIR* can support teachers in real-life scenarios, we asked three foreign language teachers to rank a list of six short documents taking into account the occurrences of two target forms, *the definite article* and *phrasal verbs*. We selected the documents by searching for news about the *Pulitzer Prize*, and we made sure that the distribution of the target constructions was different in each document. The teachers were completing this task on paper and did not have access to *FLAIR*.

Our assumption, in line with the common IR practice, was that high-ranked documents should balance the occurrences of all the items in the search query. That is, the most relevant document would ideally contain the same number of occurrences of all query items. Documents containing all query items but considerably more instances of one than the others would be ranked lower. Finally, the documents containing only one item, even if the number of occurrences is higher than in any other document, would be considered the least relevant.

In the pilot exploration with the three teachers, the general preferences of each of them confirmed these assumptions underlying the scoring algorithm implemented in *FLAIR*. For the *Pulitzer Prize* query results, the teachers agreed on the most relevant document, which was also ranked highly by *FLAIR*. In future work, we plan to follow up on this pilot with a study of *FLAIR* being used by teachers of English as a foreign language at the university level.

5 Input Enrichment Strategies

The analyses in section 4 confirmed a high variability in the occurrence of many of the targeted structures in the web documents retrieved, making a search reranking approach promising. At the same time, we also found that some (combinations of) constructions do not commonly occur in web documents. Figure 4 spells out a spectrum of input enrichment strategies for ensuring sufficient representation of the targeted linguistic forms in reading material. As an input enrichment tool originally designed with web search in mind, *FLAIR* can equally well be used to search through *Project Gutenberg*[8], the oldest digital library containing more than 50 thousand books, or in hand-curated text repositories for children or serving as resources for language teachers such as Time for Kid[9], BBC Bitesize[10], Newsela[11], or OneStopEnglish[12].

[8] https://www.gutenberg.org
[9] http://www.timeforkids.com/news
[10] http://www.bbc.co.uk/bitesize
[11] https://newsela.com
[12] http://onestopenglish.com

	REAP (Brown and Eskenazi, 2004)	**TextFinder** (Bennöhr, 2005)	**LAWSE** (Ott and Meurers, 2011)	**FLAIR**
Database	offline	offline	Web	Web
Third party tools	AltaVista	Lucene	Lucene	Bing API, Boilerpipe, Stanford Parser
Learner model	$+$	$+$	$-$	$-$
Reading interface	$+$	$-$	$-$	$+$
Text complexity	$+$	$+$	$+$	$+$
Vocabulary load	$+$	$-$	$+$	$+/-$
Grammar	$-$	$+/-$	$-$	$+$
Coverage of curriculum	$-$	$-$	$-$	$+$
Stated future work	grammar, cohesiveness	readability formula	syntactic features, grammar	vocabulary, large-scale testing

Table 2: Comparison of Information Retrieval systems for language learning.

6 Related Work

The computational linguistic research targeting the provision of reading material to learners has generally focused on vocabulary and lexical properties or readability (Miltsakaki and Troutt, 2008; Collins-Thompson et al., 2011; Vajjala and Meurers, 2014), with some of the researchers mentioning the integration of grammar modules as future work (Brown and Eskenazi, 2004; Ott and Meurers, 2011).

Table 2 puts our *FLAIR* approach into the context of three learner-oriented IR systems: *REAP* (Brown and Eskenazi, 2004), *TextFinder* (Bennöhr, 2005), and *LAWSE* (Ott and Meurers, 2011). While each of the four systems implements a text complexity module, they differ in how they treat vocabulary and grammar. Vocabulary models are built using either word lists (*LAWSE*) or the information from the learner model (*REAP*). Grammar is given little attention, apart from Bennöhr (2005) taking into account the complexity of different conjunctions as an aspect related to discourse coherence that she directly integrates into her readability formula.

A distinguishing feature of FLAIR aimed at making it usable in real-life language teaching and learning is the comprehensive coverage of the grammatical phenomena contained in a complete curriculum of English, as spelled out in the real-life English curriculum for schools in the state of Baden-Württemberg (Germany).

Finally, most of the IR tools delegate full control over the reading material to one user – either the learner or the educator. This can also be justified for language test developers (cf., *SourceFinder*; Sheehan et al., 2007), but many language learning contexts include more of a mixture of teacher-led and learner-driven, data-driven learning. *FLAIR* addresses this issue by allowing teachers to configure the linguistic form preferences determining the reranking, while letting the learner enter queries based on their content interests to identify the base set of texts being reranked.

7 Conclusion and Outlook

The paper presented *FLAIR*, a linguistically aware IR approach supporting automatic input enrichment maximizing exposure of language learners to constructions currently being taught or likely to be learned next. The *FLAIR* tool can be characterized in terms of (i) the coverage of 87 linguistic constructions implemented to meet the requirements of the official curriculum for the English language in German schools, (ii) the use of efficient IR methods for the retrieval and reranking of relevant documents based on the occurrences of selected linguistic constructions in them, and (iii) the option to preconfigure the settings to direct rather than control learners' choice of reading material.

While in this paper we have mainly focused on supporting language teachers in their search for reading material richly representing the forms to be taught, what a language learner is likely to learn next is heavily researched in Second Language Acquisition Research in terms of Krashen's $i + 1$, Vygotsky's *Zone of Proximal Development*, or Pienemann's *Teachability*, so that future research could explore combining input enrichment with learner models determining the construction to be enriched.

Based on the feedback obtained from the foreign language teachers taking part in the discussed pilot studies, we identified several strands for future development. Teachers requested expanding the functionality of the tool to include more linguistic, cultural and social text characteristics that would help them get a more complete grasp of each text retrieved by *FLAIR*. Such factors as a language variety, text register, and the use of formulaic language were prominently mentioned. As a first step, we will integrate a vocabulary module: Integration of the Academic Word List (Coxhead, 2000) is currently being implemented to estimate aspects of text register. An alternative approach we are considering is to check the percentage of words from the core general vocabulary (Brezina and Gablasova, 2013). Yet another type of word list functionality, in line with Krashen's (1977) input hypothesis and similar to the learner model implemented in the *REAP* system (Brown and Eskenazi, 2004), could keep track of the words that the learner has already encountered and take this into account in ranking the retrieved documents.

Full user studies with language teachers and learners will be necessary to evaluate the overall approach as well as the effectiveness of distinct components of *FLAIR*, including what the interactive visualization offers to the teacher. On the technical side, it would be worthwhile to explore other IR algorithms that could be more directly linked to the lexical and grammatical aspects of the linguistic system we are focusing on. On the quantitative side, the analysis of linguistic forms identified by *FLAIR* could be taken one step further by running large text corpora through the parsing module of our system. Analyzing texts in Project Gutenberg, for instance, could show whether it is possible to identify appropriate reading passages from its collection of thousands of books and shed light on the linguistic nature of such collections. Considering FLAIR in the broader research context, the system also holds promise for conducting SLA research on *input enrichment* and *input enhancement*.

Acknowledgments

This research was funded by the LEAD Graduate School [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.

Special thanks to Madeeswaran Kannan, a computational linguistics student and research assistant at the University of Tübingen, for his excellent support in optimizing the *FLAIR* prototype, making the system useful in real life.

References

Jasmine Bennöhr. 2005. A web-based personalised textfinder for language learners. Master's thesis, University of Edinburgh.

Adriane Boyd, Whitney Gegg-Harrison, and Donna Byron. 2005. Identifying non-referential it: A machine learning approach incorporating linguistically motivated patterns. In *Proceedings of the ACL Workshop on Feature Engineering for Machine Learning in Natural Language Processing*, pages 40–47, Ann Arbor, Michigan, June. Association for Computational Linguistics.

Vaclav Brezina and Dana Gablasova. 2013. Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, page amt018.

Jonathan Brown and Maxine Eskenazi. 2004. Retrieval of authentic documents for reader-specific lexical practice. In *InSTIL/ICALL Symposium 2004*.

Jill Burstein, Jane Shore, John Sabatini, Brad Moulder, Steven Holtzman, and Ted Pedersen. 2012. The language musesm system: Linguistically focused instructional authoring. *ETS Research Report Series*, 2012(2):i–36.

Kevyn Collins-Thompson, Paul N Bennett, Ryen W White, Sebastian de la Chica, and David Sontag. 2011. Personalizing web search results by reading level. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 403–412. ACM.

Averil Coxhead. 2000. A new academic word list. *TESOL quarterly*, 34(2):213–238.

Susan M Gass and Evangeline Marlos Varonis. 1994. Input, interaction, and second language production. *Studies in second language acquisition*, 16(03):283–302.

Christian Kohlschütter, Peter Fankhauser, and Wolfgang Nejdl. 2010. Boilerplate detection using shallow text features. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 441–450. ACM.

Stephen Krashen. 1977. Some issues relating to the monitor model. *On Tesol*, 77(144-158).

Shawn Loewen, Rosemary Erlam, and Rod Ellis. 2009. The incidental acquisition of third person-s as implicit and explicit knowledge. *Implicit and explicit knowledge in second language learning, testing and teaching*, pages 262–280.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.

Detmar Meurers, Ramon Ziai, Luiz Amaral, Adriane Boyd, Aleksandar Dimitrov, Vanessa Metcalf, and Niels Ott. 2010. Enhancing authentic web pages for language learners. In *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 10–18. Association for Computational Linguistics.

Detmar Meurers. 2015. Learner corpora and natural language processing. In Sylviane Granger, Gatanelle Gilquin, and Fanny Meunier, editors, *The Cambridge Handbook of Learner Corpus Research*, pages 537–566. Cambridge University Press.

Eleni Miltsakaki and Audrey Troutt. 2008. Real-time web text classification and analysis of reading difficulty. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications*, pages 89–97. Association for Computational Linguistics.

Raymond Murphy. 2012. *English grammar in use*. Ernst Klett Sprachen.

Ramanathan Narayanan, Bing Liu, and Alok Choudhary. 2009. Sentiment analysis of conditional sentences. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 180–189. Association for Computational Linguistics.

Niels Ott and Detmar Meurers. 2011. Information retrieval for education: Making search engines language aware. *Themes in Science and Technology Education*, 3(1-2):pp–9.

Manfred Pienemann. 1989. Is language teachable? psycholinguistic experiments and hypotheses. *Applied Linguistics*, 10(1):52–79.

Cristina Puente and José A Olivas. 2008. Analysis, detection and classification of certain conditional sentences in text documents. In *Proceedings of the 12th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU*, volume 8, pages 1097–1104.

Hayo Reinders and Rod Ellis. 2009. The effects of two types of input on intake and the acquisition of implicit and explicit knowledge. *Implicit and explicit knowledge in second language learning, testing and teaching*, pages 281–302.

Stephen E Robertson and Steve Walker. 1994. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 232–241. Springer-Verlag New York, Inc.

Richard W Schmidt. 1990. The role of consciousness in second language learning1. *Applied linguistics*, 11(2):129–158.

M Kathleen Sheehan, Irene Kostin, and Yoko Futagi. 2007. Sourcefinder: a construct-driven approach for locating appropriately targeted reading comprehension source texts. In *SLaTE*, pages 80–83. Citeseer.

Dan I Slobin. 1985. Crosslinguistic evidence for the language-making capacity. *The crosslinguistic study of language acquisition*, 2:1157–1256.

Merrill Swain. 1985. Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in second language acquisition*, 15:165–179.

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. 2013. On the applicability of readability models to web texts. In *Proceedings of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations*, pages 59–68.

Sowmya Vajjala and Detmar Meurers. 2014. Assessing the relative reading level of sentence pairs for text simplification. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL-14)*, Gothenburg, Sweden. Association for Computational Linguistics.

Lev Semenovich Vygotsky. 1986. *Thought and Language*. MIT Press, Cambridge, MA.

A Evaluation of the identification of the 87 linguistic constructions

Linguistic form	P	R	$\mathbf{F_1}$
to- infinitives	1.00	0.98	0.99
simple prepositions (in, at, on, with, after)	1.00	0.97	0.98
copular verbs	1.00	0.97	0.98
auxiliary verbs	1.00	0.96	0.98
irregular verbs (past participle)	1.00	0.96	0.98
advanced modals (might, ought to,able, etc.)	1.00	0.94	0.97
regular plural nouns (cats)	0.99	0.94	0.96
comparative d. of short adj. (nicer)	0.71	1.43	0.95
positive d. of adv. (fast)	0.91	1.00	0.95
Past Simple Tense	1.00	0.90	0.95
Past Time	1.00	0.90	0.95
existential there	0.90	1.00	0.95
regular verbs (past participle)	0.98	0.91	0.94
positive d. of adj. (nice)	0.94	0.93	0.94
full verb forms (am, have, etc.)	0.88	1.00	0.94
direct object	0.91	0.93	0.92
advanced prepositions (during, through, etc.)	0.90	0.93	0.91
used to	0.83	1.00	0.91
Present Simple Tense	0.94	0.87	0.90
wh- questions	0.92	0.88	0.90
Past Perfect Tense	0.82	1.00	0.90
complex sentences (with subordinate clauses)	0.85	0.95	0.90
Present Time	0.97	0.83	0.90
-ing verb forms (gerund and pr. participle)	0.86	0.92	0.89
be- questions	0.80	1.00	0.89
subjective pronouns (I, you)	1.00	0.79	0.88
subordinate clauses reduced	0.83	0.94	0.88
Simple Aspect	0.85	0.92	0.88
ing- noun forms	0.90	0.82	0.86
Past Progressive Tense	1.00	0.75	0.86
comparative d. of long adv. (more often)	1.00	0.75	0.86
Progressive Aspect	1.00	0.73	0.84
adverbial clauses	0.83	0.83	0.83
Present Progressive Tense	1.00	0.71	0.83
superlative d. of long adv. (most often)	1.00	0.71	0.83
incomplete sentences	1.00	0.67	0.80
imperative verb forms	1.00	0.67	0.80
have- questions	0.67	1.00	0.80
real conditionals	0.68	0.96	0.79
passive voice	1.00	0.64	0.78
absolute possessive pronouns	1.00	0.63	0.77
relative clauses	0.71	0.83	0.77
Perfect Aspect	0.89	0.67	0.76
phrasal verbs	1.00	0.61	0.76
Present Perfect Tense	0.88	0.64	0.74
complex prepositions (according to, etc.)	0.56	0.83	0.67
comparative d. of short adv. (faster)	1.00	0.50	0.67
indirect object	1.00	0.50	0.67
unreal conditionals	0.63	0.71	0.67
comparative d. of long adj. (more interesting)	1.00	0.40	0.57
simple sentences	0.80	0.44	0.57
Degrees of comparison (adj)	0.93	0.95	0.89
Tenses	0.95	0.84	0.88
Conditionals	0.65	0.84	0.73
Mean (81 targets)	0.94	0.90	0.91
Median (81 targets)	1.00	0.97	0.95

Data: nine news articles with an average length of 28 sentences.

28 constructions with $\mathbf{F_1}$ **of 1**:
questions, do- questions, yes-no questions, tag questions, Future Simple Tense, Future Time, going to, irregular plural of nouns (*children*), emphatic *do*, contracted verb forms, simple modals (*can, must, need, may*), short negation (no, not, never, n't), partial negation (hardly, barely), simple conjunctions (*and, but, or*), advanced conjunctions, objective pronouns, possessive pronouns, reflexive pronouns, some, any, many, much, a, an, the, superlative form of short adjectives (*nicest*), superlative form of long adjectives (*most interesting*), superlative form of short adverbs (*fastest*).

As the texts for the evaluation were selected randomly, we found **few to no instances** of the following six constructions:
Present Perfect Progressive Tense, Past Perfect Progressive Tense, Future Perfect Progressive Tense, Perfect Progressive Tense, Future Progressive Tense, Future Perfect Tense.