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

Maria Chinkina, Ankita Oswal, Detmar Meurers
BEA13 in New Orleans, USA
06/05/2018
Outline

1. Motivation
2. Input enrichment
3. FLAIR system
4. The online study
5. Discussion and Outlook
Motivation

- **Input Hypothesis** (Krashen, 1977)
  - Exposure to input containing target constructions

- **Frequency of forms** (Slobin, 1985)
  - The more frequent the construction in input, the more it facilitates acquisition

- **Input Enrichment**
  (Related to *input flood* by Trahey and White, 1993)
  - Ensuring a high number of target constructions in text
Problem

Material of interest is readily available on the Web. But the linguistic forms are unevenly distributed across texts.

The top 60 results for the query *Brexit*:
Problem

Material of interest is readily available on the Web. But the linguistic forms are unevenly distributed across texts.

The top 60 results for the query Brexit:

1. How do teachers and learners get to the best results?
2. Are those results still good content-wise?
Solution: Automatic input enrichment

We re-rank the search results based on the representation of the selected linguistic forms in them.
FLAIR: Web Search for Language Teachers

Available online: purl.org/icall/flair

FLAIR (Chinkina & Meurers, 2016) helps teachers search for texts appropriate in form, content, and reading level. It:

- retrieves the search results from Microsoft Bing or lets the user upload their own texts,
- identifies 87 linguistic forms from the official curriculum,
- reranks the results based on the selected linguistic forms,
- thus, provides systematic automatic input enrichment.
FLAIR Interface

1. Brexit group fined for breaking spending rules in EU vote...
   https://www.reuters.com/article/uk-britain-evoting-brexit...
   Britain's Electoral Commission imposed a record-breaking £70,000 fine on Friday on one of the
   main groups that campaigned for Brexit, and said police might have to investigate possible criminal
   offences because of unreported campaign spending.

2. Exclusive - Jumping ship: Brexit-hit EU staff ditch UK...
   https://www.reuters.com/article/uk-britain-evoting-exclusives...
   About one in 10 British civil servants at the European Commission has taken another EU nationality
   since the Brexit vote, but are nonetheless resigned to scant prospects of future promotion.

3. Brexit seen threatening UK links in EU supply chain | Reuters
   https://www.reuters.com/article/uk-britain-evoting-brexit-sees...
   Mandy Richardson knew Brexit was going to be a challenge for her aviation components firm, but it was still
   a shock when she heard a French company bluntly ruling out British suppliers from an international bid for a contract in China.

4. UK's May facing local election losses, Brexit unity at ...
   https://www.reuters.com/article/uk-britain-political-elections-vote...
   Voters in England cast their ballots on Thursday in local government elections expected to show mixing
   support for Prime Minister Theresa May's opponents in London and add to questions about her ability to
   follow through on her Brexit plan.

5. Whistleblower questions Brexit result, says campaigners ...
   https://www.reuters.com/article/uk-britain-evoting-brexit ...
   A whistleblower at the heart of a Facebook data scandal on Monday questioned the results of Britain's
   2016 Brexit referendum as his lawyer presented evidence that they said showed the main campaign for
   leaving the EU had broken the law.

6. Trust me on Brexit, May says as ministers squabble | Reuters
   https://www.reuters.com/article/uk-britain-evoting-brexit ...
   British Prime Minister Theresa May said on Sunday she could be trusted to deliver Brexit, but that it
   could not be done without compromises on all sides – a possible warning to cabinet ministers who are

Brexit group fined for breaking spending rules in EU vote.

Referendum. The commission said it was enforcing the rules without suggesting the breach had altered the result.
"These are serious offences," told Rob Pope, the Electoral Commission's director of political finance and regulation.
"Leave.EU exceeded its spending limit and failed to declare its funding and its spending correctly."

Add on BBC radio if the commission was saying the breach was serious enough to have impacted the result, its
Chief executive, Claire Basset, said "No", but she added that the rules still needed to be enforced.
The commission said it suspected criminal offences may have been committed, and the person responsible,
Leave.EU CEO Liz Bittner, had been referred to the police.
Related Coverage.
"REMOANER SWAMP".
Aron Banks, the founder of Leave.EU who was pictured
with Donald Trump and leading Brexit Napolitano outside a closed elevator soon after the 2016 U.S.
Performance evaluation

Manual annotation of 351 sentences:

<table>
<thead>
<tr>
<th>Linguistic target</th>
<th>Prec.</th>
<th>Rec.</th>
<th>F₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes/no questions</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Irregular verbs</td>
<td>1.00</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td>used to</td>
<td>0.83</td>
<td>1.00</td>
<td>0.91</td>
</tr>
<tr>
<td>Phrasal verbs</td>
<td>1.00</td>
<td>0.61</td>
<td>0.76</td>
</tr>
<tr>
<td>Tenses (Present Simple, ...)</td>
<td>0.95</td>
<td>0.84</td>
<td>0.88</td>
</tr>
<tr>
<td>Conditionals (real, unreal)</td>
<td>0.65</td>
<td>0.83</td>
<td>0.73</td>
</tr>
<tr>
<td><strong>Mean</strong> (81 targets)</td>
<td>0.94</td>
<td>0.90</td>
<td>0.91</td>
</tr>
<tr>
<td><strong>Median</strong> (81 targets)</td>
<td>1.00</td>
<td>0.97</td>
<td>0.95</td>
</tr>
</tbody>
</table>
Use Cases of FLAIR

- It supports language teachers and learners in searching for linguistically rich reading material.

- It can serve as a front-end for input enhancement (Meurers et al., 2010) and exercise generation systems (Burstein et al., 2012).

- It provides a platform for Second Language Acquisition studies on input enrichment and enhancement.
Use Cases of FLAIR

- It supports language teachers and learners in searching for linguistically rich reading material.

- It can serve as a front-end for input enhancement (Meurers et al., 2010) and exercise generation systems (Burstein et al., 2012).

- It provides a platform for Second Language Acquisition studies on input enrichment and enhancement.
Studies on Input Enrichment

Studies on input enrichment with *language learners* have yielded mixed results (Trahey and White, 1993; Reinders and Ellis, 2009).

What about *language teachers*? Do they need automatic input enrichment when selecting reading material for their students?
Online Study with English Teachers

- **Read** pairs of news articles (one from Bing, one from FLAIR)
- **Rate** each of the news articles
- **Select** one news article in each pair as a reading assignment
Topic 1/10

Topic: Rogers Cup 2017

Linguistic forms: regular verbs (typed)
irregular verbs (bought)

Start
"Pliskova reaches Rogers Cup quarters after Osaka retires ..." (click the title to read)

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)

Next
"Canada"s Bouchard ousted in first round of home tournament ..." (click the title to read)

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: "Pliskova reaches Rogers Cup quarters after Osaka retires ...

Text 2: "Canada"s Bouchard ousted in first round of home tournament ...

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

Submit
Repeated-measures Design

12 participants produced 240 responses by rating 20 articles in pairs:

– one top news article from Bing (original)
– one top news article from FLAIR (re-ranked)

Each pair had the same topic, the same pair of linguistic forms, was of comparable length and readability level.
Topics

60 news articles from Reuters on 10 popular topics:

- Game of Thrones
- healthcare
- street artists
- Roger’s Cup 2017
- SpaceX
- electric cars
- Bitcoin
- Venezuela coup
- Brexit
- opioid epidemic
## Pairs of Linguistic Forms

<table>
<thead>
<tr>
<th>Category</th>
<th>Form</th>
<th>Mean relative frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>frequent (95%)</strong></td>
<td></td>
<td>Bing</td>
</tr>
<tr>
<td></td>
<td>regular verbs</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>irregular verbs</td>
<td>0.012</td>
</tr>
<tr>
<td><strong>mixed (50%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>present simple</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>present continuous</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>infrequent (4%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>comparative d. of short adj. and adv.</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>comparative d. of long adj. and adv.</td>
<td>0</td>
</tr>
</tbody>
</table>
Sanity check

**Linguistic representation:** was FLAIR rated higher than Bing?

<table>
<thead>
<tr>
<th>Linguistic representation</th>
<th>Bing</th>
<th>FLAIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1 – 5)</td>
<td>$M = 2.51$</td>
<td>$M = 3.22$</td>
</tr>
<tr>
<td></td>
<td>$SD = 1.15$</td>
<td>$SD = 1.07$</td>
</tr>
</tbody>
</table>

*Logistic regression* was significant: 

$b = 1.89$, $SE = 0.51$, $p < .001$

$\Rightarrow \text{FLAIR} > \text{Bing}$ with regard to representation of linguistic forms
Hypotheses

**Input enrichment:** FLAIR

**Baseline:** Microsoft Bing

**H1:** When selecting a *reading assignment*: FLAIR > Bing

**H2:** Relevance of the *content* to the topic: FLAIR < Bing

**H3:** The *more infrequent* the target linguistic forms are, the more FLAIR > Bing
H1: Selecting a reading assignment: FLAIR > Bing?

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

'Doesn't matter' items (25%) were excluded from the analysis.

- English teachers selected FLAIR 71% of the time, significantly more than Bing:
  \[ \chi^2(1) = 16.04, p < .001 \]

- 'Definitely' was selected 3 times more for FLAIR than for Bing:
  \[ \chi^2(1) = 12.60, p < .001 \]

\[ \Rightarrow \text{FLAIR} > \text{Bing} \] when selecting a reading assignment

A strong argument in support of automatic input enrichment.
H2: Relevance of content to the topic: FLAIR < Bing?

<table>
<thead>
<tr>
<th></th>
<th>Bing</th>
<th>FLAIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevance of content</td>
<td>$M = 3.58$</td>
<td>$M = 3.67$</td>
</tr>
<tr>
<td>(1 – 5)</td>
<td>$SD = 1.00$</td>
<td>$SD = 1.08$</td>
</tr>
</tbody>
</table>

Logistic regression was not significant:

$b = 0.53$, $SE = 0.74$, $p = .470$

Was it due to chance or is FLAIR = Bing with regard to content?
H2: Relevance of content to the topic: FLAIR < Bing?

*Equivalence tests* \((d = 0.5, \alpha = .05)\) were significant:

\[
t_1 = 4.55, p_1 < .001
\]
\[
t_2 = -3.19, p_2 < .001
\]
\[
CI [-0.13; 0.31]
\]

\[\Rightarrow \text{FLAIR} = \text{Bing}\] with regard to relevance of content to topic

**No trade-off** between content and linguistic representation in the top 20 results for popular topics in English.
H2: Relevance of content to the topic: FLAIR < Bing?

*Equivalence tests* \((d = 0.5, \alpha = .05)\) were significant:
\[ t_1 = 4.55, p_1 < .001 \]
\[ t_2 = -3.19, p_2 < .001 \]
\[ CI [-0.13; 0.31] \]

\[ \Rightarrow \text{FLAIR} = \text{Bing} \] with regard to relevance of content to topic

**No trade-off** between content and linguistic representation in the top 20 results for popular topics in English.
H2: Relevance of content to the topic: FLAIR < Bing?

Equivalence tests \((d = 0.5, \alpha = .05)\) were significant:

\[ t_1 = 4.55, p_1 < .001 \]
\[ t_2 = -3.19, p_2 < .001 \]
\[ CI [-0.13; 0.31] \]

\(\Rightarrow\) FLAIR = Bing with regard to relevance of content to topic

No trade-off between content and linguistic representation in the top 20 results for popular topics in English.

BUT: Correlation of the answers to the two questions (form and content) was 0.2 \((p < .001)\).
H3: The More Infrequent, the More FLAIR > Bing?

Mean preference for FLAIR (1 – 5):

<table>
<thead>
<tr>
<th></th>
<th>Frequent</th>
<th>Mixed</th>
<th>Infrequent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Regular and</td>
<td>Present Simple</td>
<td>Comparative degree of short</td>
</tr>
<tr>
<td></td>
<td>irregular verbs</td>
<td>and Present</td>
<td>and long adj. and adv.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Continuous</td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>3.46</td>
<td>3.92</td>
<td>3.69</td>
</tr>
<tr>
<td>$SD$</td>
<td>1.39</td>
<td>1.99</td>
<td>1.30</td>
</tr>
</tbody>
</table>

$\Rightarrow$ Automatic input enrichment seems particularly beneficial for targeting linguistic forms of lower frequency levels.
H3: The More Infrequent, the More FLAIR > Bing?

Mean preference for FLAIR (1 – 5):

<table>
<thead>
<tr>
<th>Frequent</th>
<th>Mixed</th>
<th>Infrequent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular and</td>
<td>Present Simple</td>
<td>Comparative degree of short</td>
</tr>
<tr>
<td>irregular verbs</td>
<td>and Present</td>
<td>and long adj. and adv.</td>
</tr>
<tr>
<td></td>
<td>Continuous</td>
<td></td>
</tr>
<tr>
<td>$M = 3.46$</td>
<td>$M = 3.92$</td>
<td>$M = 3.69$</td>
</tr>
<tr>
<td>$SD = 1.39$</td>
<td>$SD = 1.99$</td>
<td>$SD = 1.30$</td>
</tr>
</tbody>
</table>

=> Automatic input enrichment seems particularly beneficial for targeting linguistic forms of lower frequency levels.

2-way rANOVA was not significant: $F(2, 90) = 0.87, p = .419$

=> No statistically significant linear relationship between frequency of linguistic forms and preference for FLAIR
H3: The More Infrequent, the More FLAIR > Bing?

Check for pairwise differences (*paired 2-samples Wilcoxon tests*):

<table>
<thead>
<tr>
<th>Pairs of linguistic forms</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequent – Mixed</td>
<td>157</td>
<td>.643</td>
</tr>
<tr>
<td>Infrequent – Mixed</td>
<td>128</td>
<td>.352</td>
</tr>
<tr>
<td>Infrequent – Frequent</td>
<td>217</td>
<td>.727</td>
</tr>
</tbody>
</table>

=> **No statistically significant differences** in preference for FLAIR when targeting linguistic forms of different frequencies
Conclusion

- Teachers preferred FLAIR over Bing when selecting a reading assignment.

- There was no trade-off between relevance of content and linguistic representation: FLAIR was rated as high as Bing with regard to content.

- The preference for FLAIR did not significantly differ when targeting frequent and infrequent linguistic forms.
Outlook

Possible empirical studies:

- A set of independent studies with English teachers looking at content, linguistic forms and selecting reading material separately

- A randomized controlled field study to compare the learning outcomes of students using web search vs. automatic input enrichment

Possible system extensions:

- Providing a variety of contexts in which linguistic forms are used, with their different meanings

- Integration of a component that automatically generates exercises targeting the selected linguistic forms
Take-Home Messages

1. Develop systems for English teachers – they appreciate it!

2. Let them give you feedback – you will appreciate it!
References


References


Thank you.

Contact:

Maria Chinkina
maria.chinkina@uni-tuebingen.de