Smart Dictionary for E-book Reading Analytics

Louis Lecailliez, Brendan Flanagan, Alice Mei-Rong Chen, Hiroaki Ogata

Kyoto University, Japan

Learning and Educational Technologies Research Unit

Authors

Louis Lecailliez

🇫🇷 France

Brendan Flanagan

🇦🇺 Australia

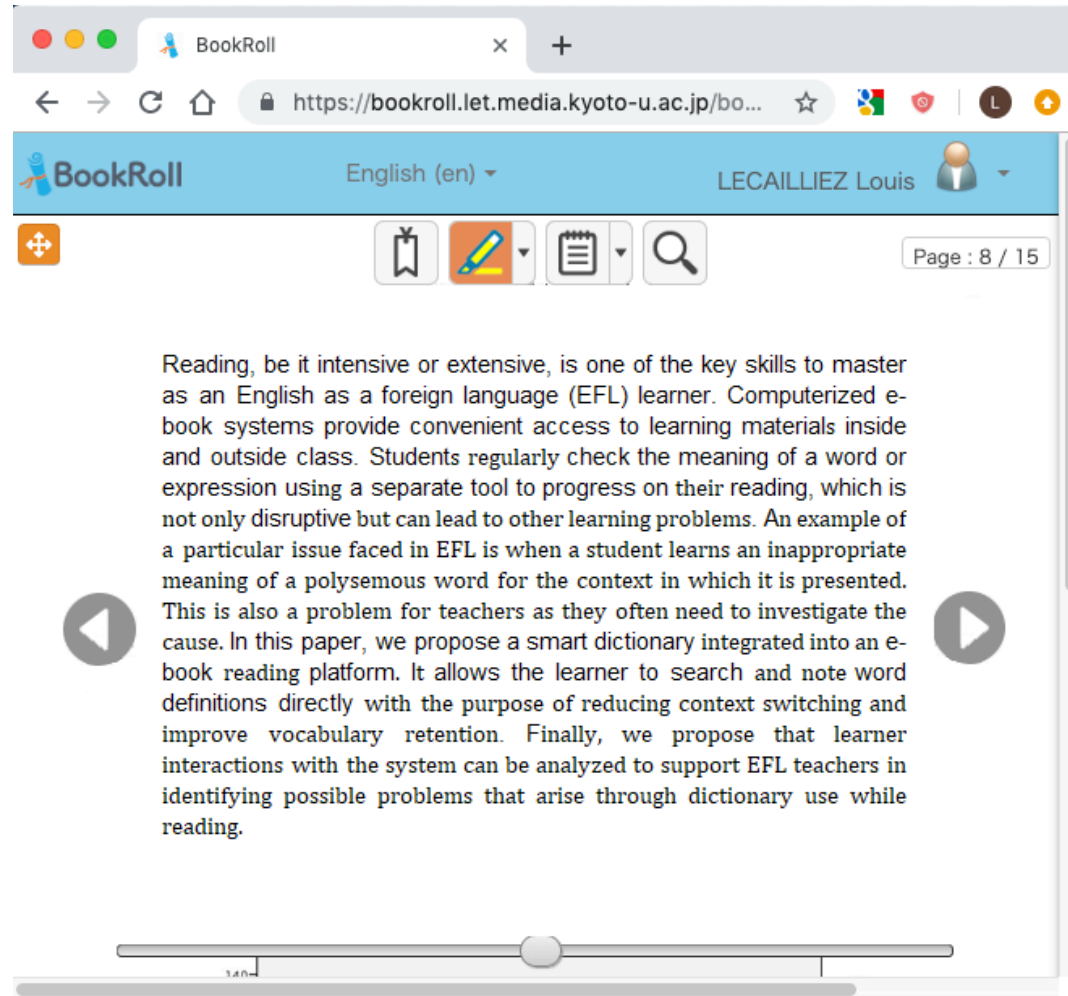Mei-Rong Alice Chen

陳玫蓉

🇹🇼 Taiwan

Hiroaki Ogata

緒方広明

🇯🇵 Japan

Presentation Plan

1) The Need for Dictionary System

2) Research Results on Dictionary Usage (Related Work)

3) Features of our Dictionary

4) Our Learning Analytics Infrastructure Architecture

5) New Visualization Widgets

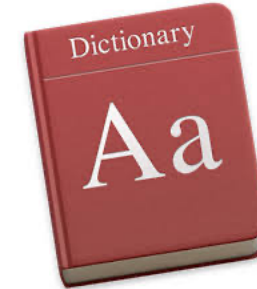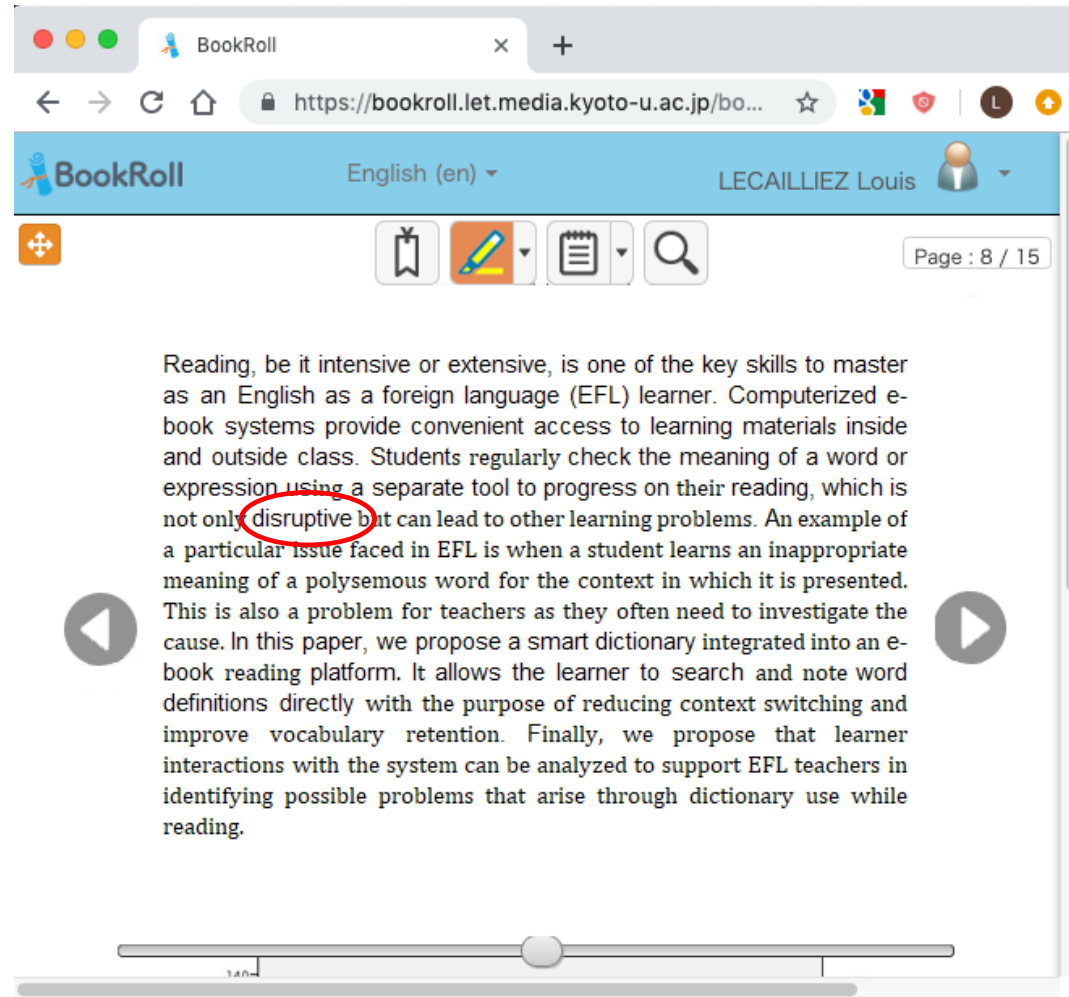6) Evaluation and Deployment (Future Work)

The Need for a Dictionary System in our LA platform

BookRoll: our Lab E-Book Platform

- Used internationally
  - Japan, Taiwan, India, Turkey
- Any kind of traditional class
  - Mathematics, science, languages

- English as a Foreign Language:
  - English lessons
  - Extensive reading
  - Intensive reading

When a Student Doesn't Understand a Word?

Dictionary App or Device

OS function lookup

External Website

PDF are converted to images

Some schools forbid internet use

Learning Analytics Opportunity

- Acquire search data from learners

- Allow specific features / research question to be addressed
  - in this paper, experimenting with definition reordering

  => end goal being better vocabulary retention

- In the (near) future, other questions or problems are to be addressed

  => multi-word expression search

  => personalized example sentence selection

  => search strategy

Research Results on Dictionary Usage

Research Results on Dictionary Usage

Some difficulties exist

- Dictionary usage is a skill that must be learned (Nesi, 2014)

- Selection the right meaning is a problem for users (Luppescu, 1993)
  => we will try to guide the learner to make to correct selection

- Dictionary usage can sometimes cause lexical errors (Ard, 1982)

- Users don't read the full entry (Cubillo, 2002)
  => "They then became frustrated because they had not read a second entry for the same form belonging to a different grammatical class [...]"
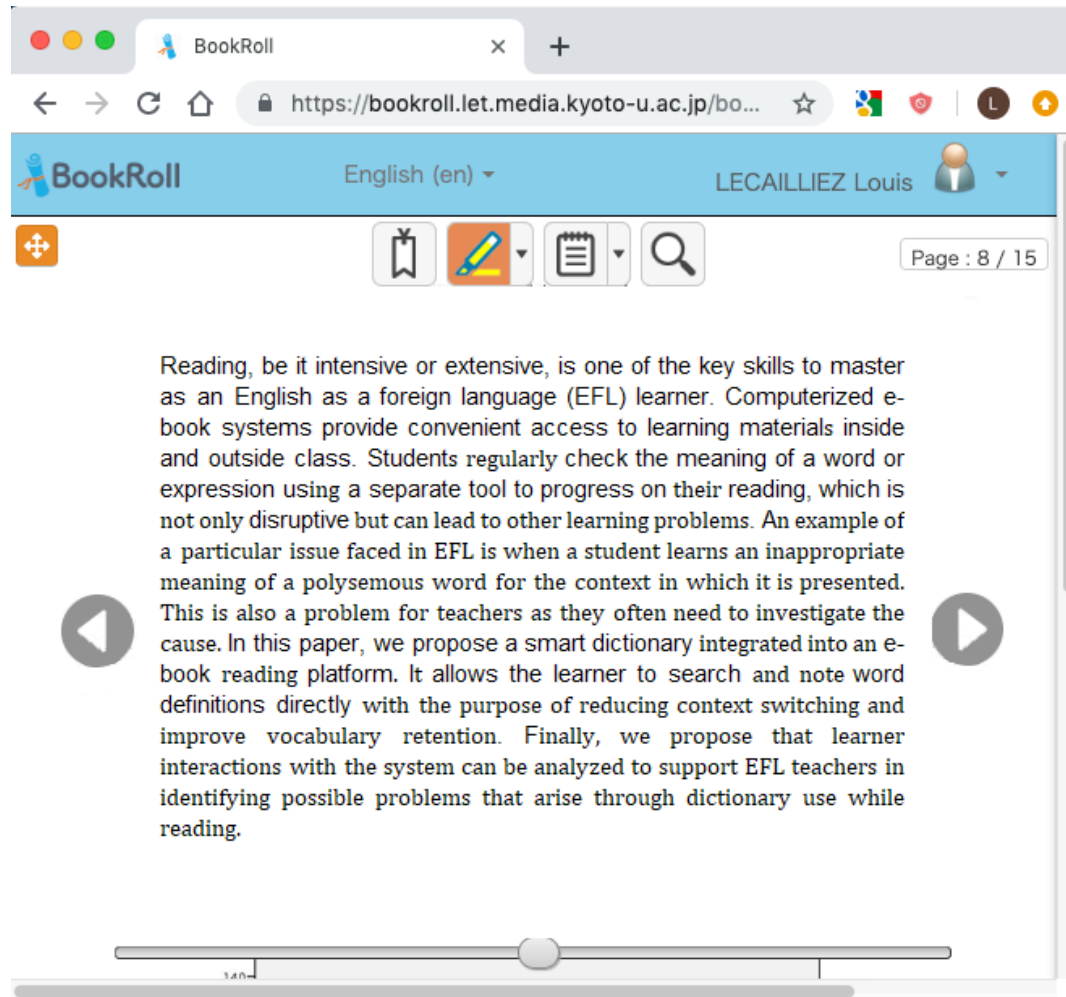
Research Results on Dictionary Usage

Results informing our design

- Bilingual dictionaries better for beginner; monolingual better for more advanced learners (Lew, 2004)

- Forcing less advanced user to make a choice (meaning selection) improves long term vocabulary retention (Chang, 2018)

=> as we target beginners, we will force action and use a bilingual dictionary

- User tend to select the first meaning they encounter (Nesi & Haill, 2002)

=> we could exploit this tendency by re-ranking the meaning listed for a word
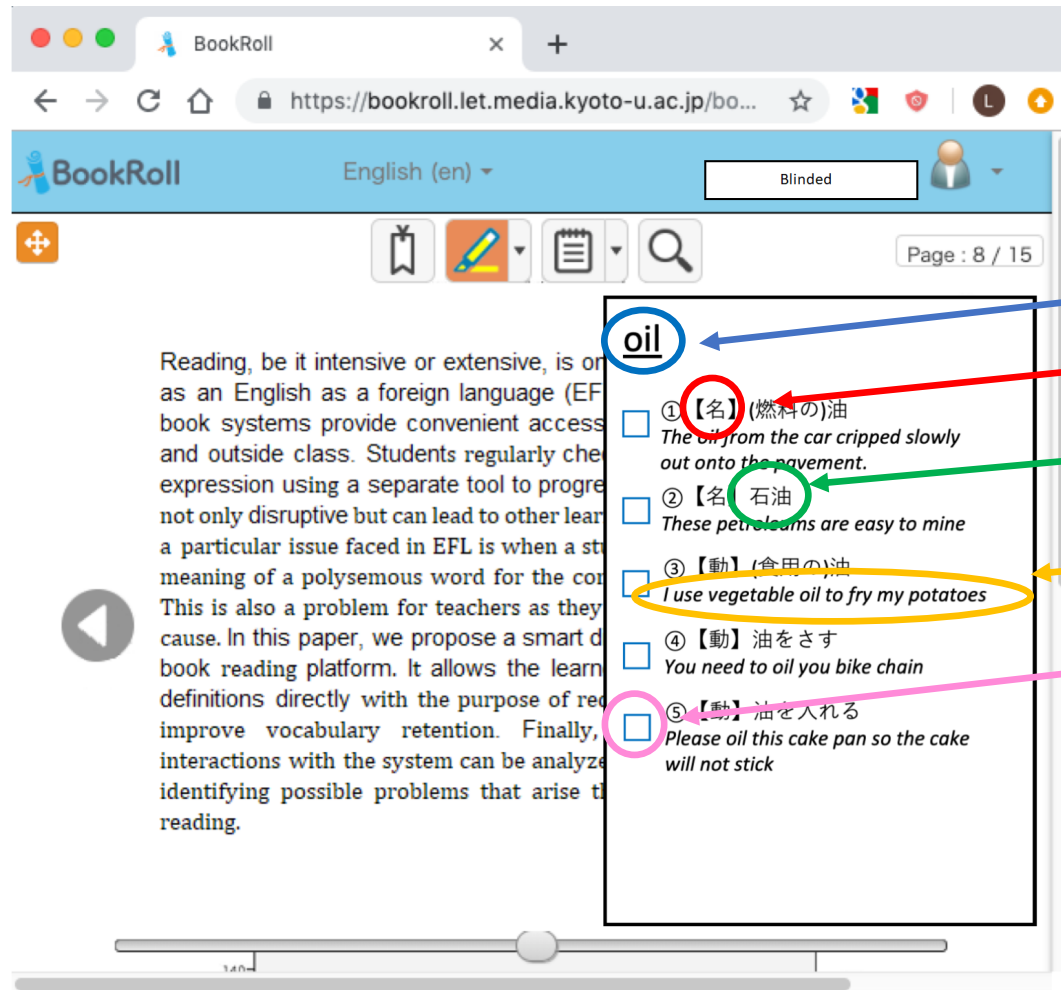
Features of the Dictionary

E-book Reader Software: Bookroll

- Navigation
  - next page
  - previous page
- Bookmarking
- Note taking
  - typing
  - handwriting
- Every action is logged with xAPI
- Logs are analyzed in a dashboard

Dictionary in Bookroll

- Based on memo taking feature

- Dict. Micro-structure:
  - headword on top
  - part of speech in brackets
  - Japanese translation
  - English sentence example

- Checkbox to select meaning

Meaning Reordering: Sources of Data

- Dictionary search exist in commercial products
    - => query are made by sending only the searched string
    - => light processing can happen: lemmatization

- Here, we use three additional source of data:

- First, we send the surrounding context of the queried word
    - => we send the **surrounding** context of the [window=N]

- Second, we have learner query and action logs

- Third, optionally, teacher override

Entry Re-ordering #1: Context

oil

☐ ① 【名】(燃料の)油
The oil from the car cripped slowly out onto the pavement.

☐ ② 【名】石油
These petroleums are easy to mine

☐ ③ 【動】(食用の)油
I use vegetable oil to fry my potatoes

☐ ④ 【動】油をさす
You need to oil you bike chain

☐ ⑤ 【動】油を入れる
Please oil this cake pan so the cake will not stick

Figure 1

oil

☐ ① 【動】(食用の)油
I use vegetable oil to fry my potatoes

☐ ② 【動】油をさす
You need to oil you bike chain

☐ ③ 【動】油を入れる
Please oil this cake pan so the cake will not stick

☐ ④ 【名】(燃料の)油
The oil from the car cripped slowly out onto the pavement.

☐ ⑤ 【名】石油
These petroleums are easy to mine

Figure 2

- Word + the part of sentence it appears are send to dictionary server
- Part of speech (PoS) analysis
- Example
  - Figure 1: noun first
  - Figure 2: verb first
- Future: Word Sense Disambiguation (WSD)
  - (Navigli, 2009)

Entry Re-ordering #2: User Log
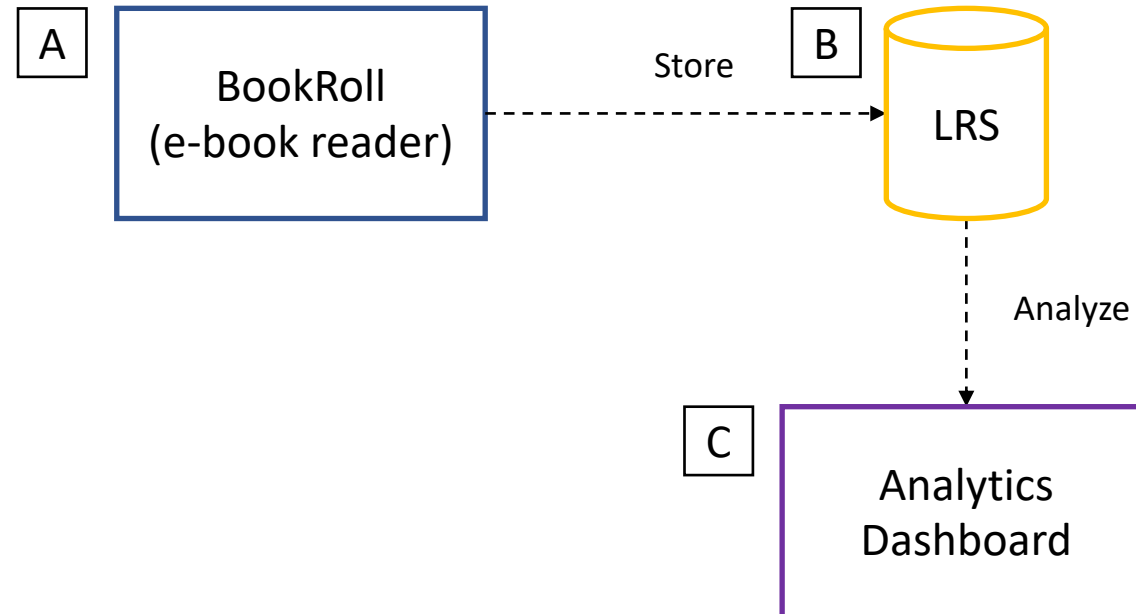
Checkbox acts
as poll items

oil

☐ ① 【動】(食用の)油
I use vegetable oil to fry my potatoes

☐ ② 【動】油をさす
You need to oil you bike chain

☐ ③ 【動】油を入れる
*Please oil this cake pan so the cake
will not stick*

☐ ④ 【名】(燃料の)油
*The oil from the car cripped slowly
out onto the pavement.*

☐ ⑤ 【名】石油
These petroleums are easy to mine

- Threshold based re-ranking

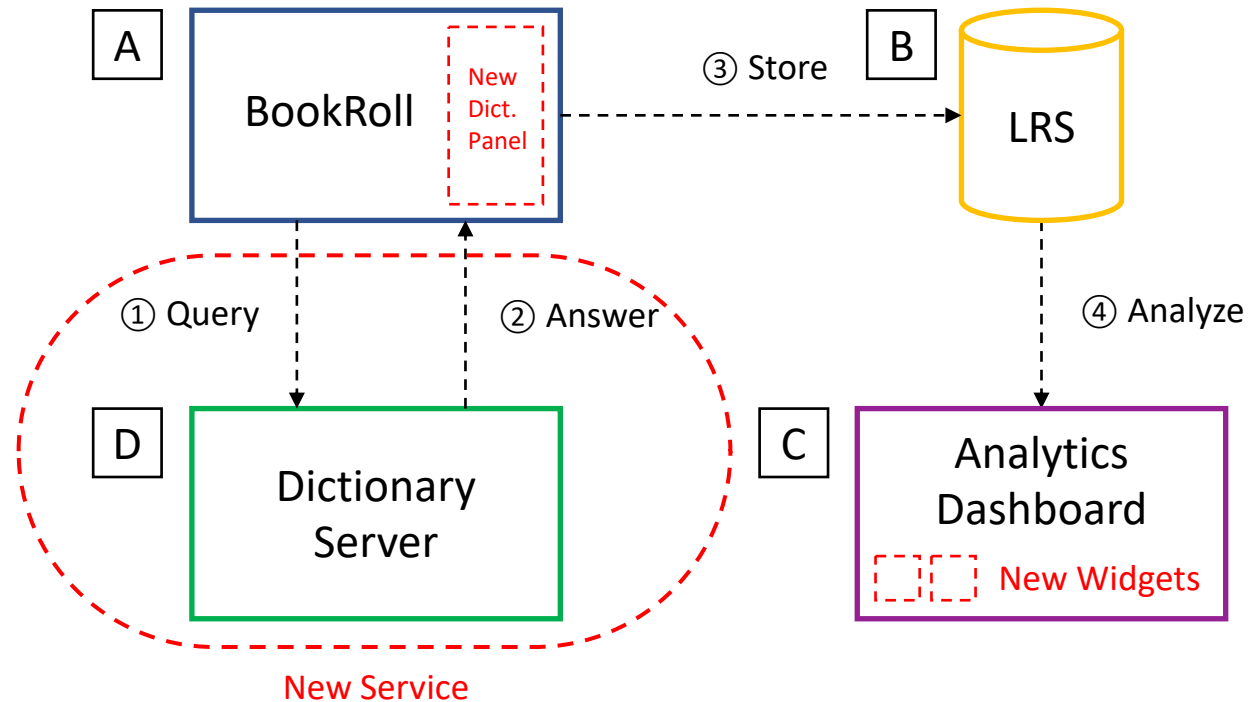# Our LA Platform Infrastructure Architecture

Existing Learning Infrastructure

- Deployed internationally in Japan, Taiwan, India, Turkey

A | BookRoll (e-book reader) — Store → B | LRS

LRS — Analyze → C | Analytics Dashboard

Extended Architecture

Dictionary Server

- Allow independent development/maintenance of the dictionary

- Language agnostic: can load monolingual or bilingual dict. of any lang.

- Support multiple packages for a given language pairs

- xAPI dictionary profile developed

- Web UI and Semantic Web capabilities => not used in this paper

New Visualization Widgets

Visualization Widget #1: Search Heat-Maps

a) Search Heat-Map

The Tahiti rail was 9 inches (23 centimetres) long, and its colouration was unusual for a rail. The underparts, throat, and eyebrow-like supercilium were white, and the upper parts were black with white dots and bands.

underparts: 5

b) Error Rate Heat-Map

The Tahiti rail was 9 inches (23 centimetres) long, and its colouration was unusual for a rail. The underparts, throat, and eyebrow-like supercilium were white, and the upper parts were black with white dots and bands.

Visualization Widget #2: QWIC

Query Word in Context

Search	Context	# queries	Students	Definition
tomorrow	ince tomorrow. **tomorrow** the Cat will be	3	admin, admin, admin	Yes
here	e Cat will be **there** for you.	1	admin	Yes
tomorrow t	ince tomorrow. **tomorrow** the Cat will be	1	admin	No

Lab Experiment #1

Search	Context	# queries	Students	Definition
permafrost	locked in the **permafrost** layer under the	7		Yes
relics	are probably **relics** of catastrophi	5		Yes
Martian surface	er on the early **Martian surface**. A computer-ge	5		No

Evaluation and Deployment

Evaluation of Word Meaning Ranking

- Heavily depends on dictionary (data source) used

- Can be tested locally

- Will be done on both real world learning material

- Also on a set of crafted sentences

Future Work

Short term

- Deploy the software in a Japanese school
  - Two groups, one with re-ranking activate, the other deactivate
- Do a preliminary study to weed out issues (done!)
- Run a real study to test the research question

Long term

- Test other research questions such as multiword search or example sentence selection from a corpus

Contributions

- Language agnostics dictionary server pluggable in LA platform

- Combined two research results to adress an existing issue

- Designed two dashboard vizualisation widgets for teacher to take action

Thanks for your attention

Questions, comments, contact:

louis.lecailliez@outlook.fr

Website

louis.lecailliez.net

Additional References

- Cubillo, M. C. C. (2002). Dictionary use and dictionary needs of ESP students: An experimental approach. *International Journal of Lexicography*, 15(3), 206-228.

- Navigli, R. (2009). Word sense disambiguation: A survey. *ACM computing surveys (CSUR)*, *41*(2), 1-69.

  https://dl.acm.org/doi/pdf/10.1145/1459352.1459355