Template-guided Grammatical Error Feedback Comment Generation

Steven Coyne

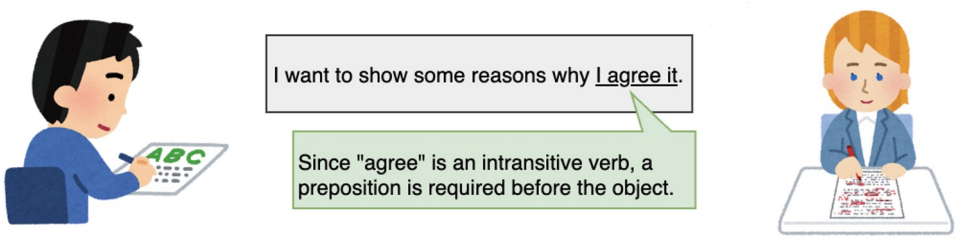
Tohoku University
RIKEN

EACL 2023 Student Research Workshop
Thesis Proposal

coyne.steven.charles.q2@dc.tohoku.ac.jp

Feedback Comment Generation: Task and Motivation

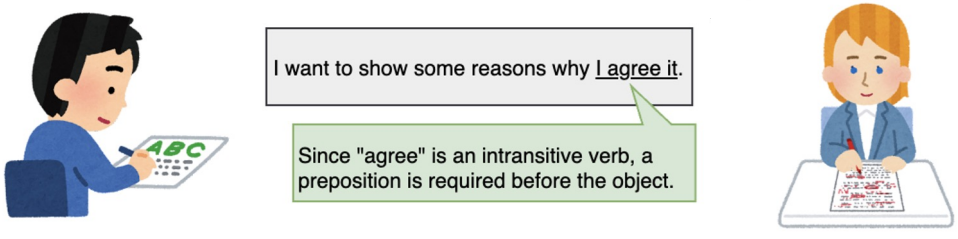
- Setting: language education. Learners write, instructors check and comment.



Example based on Nagata et al., 2020

Feedback Comment Generation: Task and Motivation

- Setting: language education. Learners write, instructors check and comment.



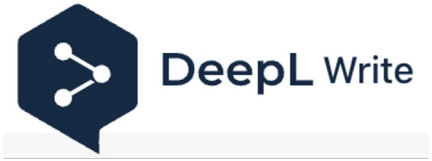
- Issue: time and labor-intensive for instructors.
- Some learners do not have access to an “instructor.”

Example based on Nagata et al., 2020

Automatically Generate Feedback Comments

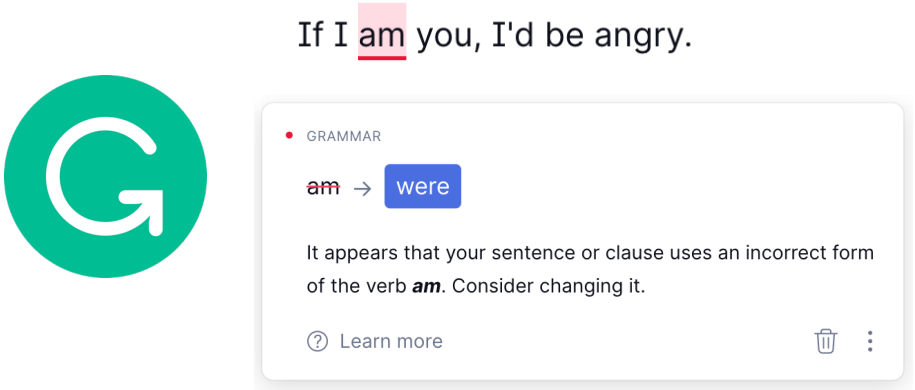
- A system can be added to a learning environment or writing application.
- Instructors can work faster with comment suggestions based on the text.
- They can accept, edit, or reject the suggestions (human mediation).
- Can help independent learners in a language learning app.

What about Grammarly (and others)?



- These systems are for **writing support,** correcting issues to improve the text.
- Our goal: **learning support**. We want learners to absorb each comment and learn from it.
- Our comments can be **more detailed** and **less direct**, and should not always give the exact correct answer to the user.

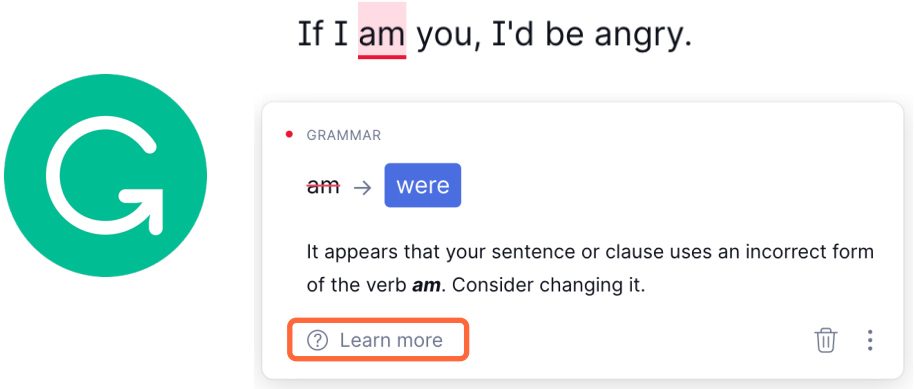
Writing vs. Learning Support

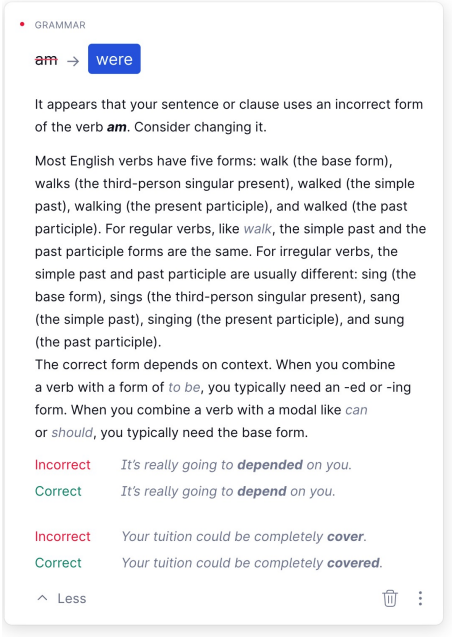


- May not explain **why** there is an error, or what to study.
- One-click fixes can be applied without understanding.

Content from https://grammarly.com/

Writing vs. Learning Support





- May not explain **why** there is an error, or what to study.
- One-click fixes can be applied without understanding.

Content from https://grammarly.com/

Writing vs. Learning Support

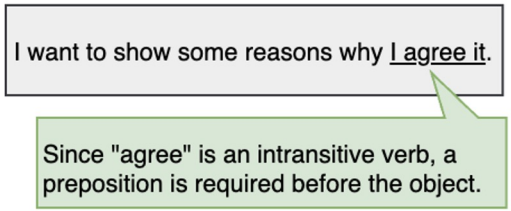
If I am you, I'd be angry.

Learning Support System

In a conditional sentence like this one, the first verb should be in the past tense or subjunctive.

- Explains **why** “am” is wrong and **what to change.**
- Keywords for learning about the rule (conditionals).
- Writer has a chance to learn and apply the rule.

Related Work



- Nagata (2019), Nagata et al. (2020):
  Corpus of (learner sentence, comment text, location)
  - This is our main resource as of now

- Hanawa et al. (2021): Experiments showing difficulties with generation.
  - Mixed generations
  - Unreliable generations and edits

Target sentence	Retrieved feedback comment	Output of original retrieve-and-edit
… should declare **about** area for …	Since the verb claim is a transitive verb, the object does not require a preposition.	Since the verb area is a transitive verb, the object does not require a preposition.

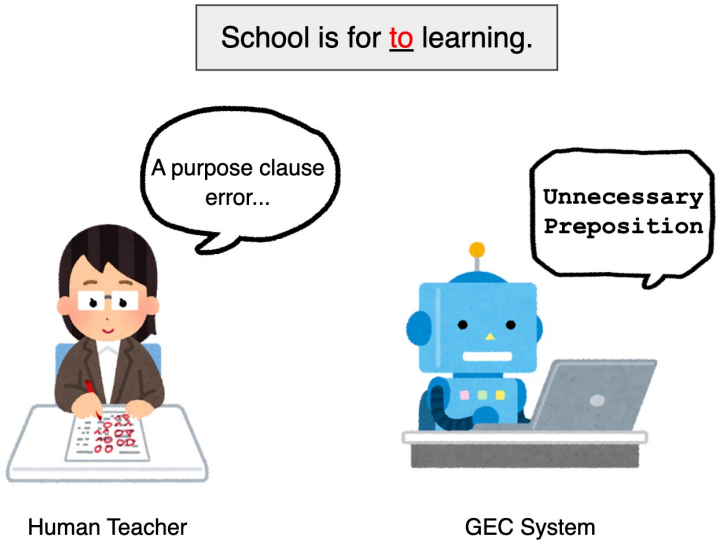
Example of problematic editing in Hanawa et al. (2021)

A Feedback Comment Typology

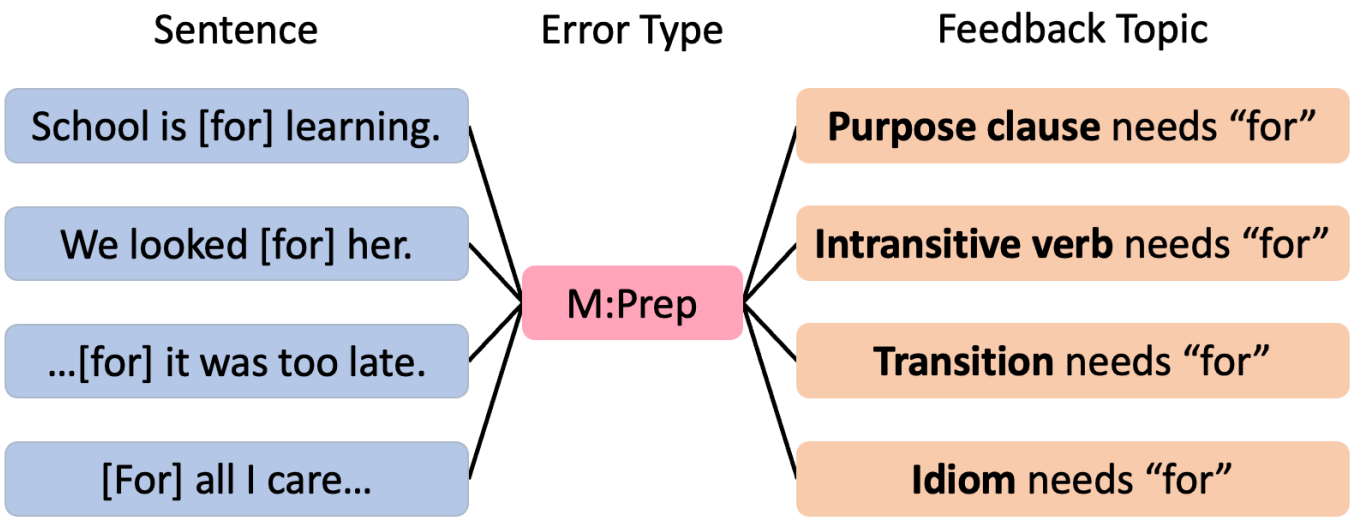
- It seems that some “kinds” of comments have a higher rate of issues, or have unique issues.
- What “kinds?” **There is no well-defined typology for feedback comments.**
- Grouping by manual labels enables better analysis of outputs and their issues.
- Enables different approaches for different types of comments.

Why not use Error Types?

- Error types from grammatical error correction (GEC)
- Ultimately for a different domain.
- Too operation-focused.
- Do not capture the broader concepts that motivate an instructor to comment on an issue.
- Complex mappings to such topics.

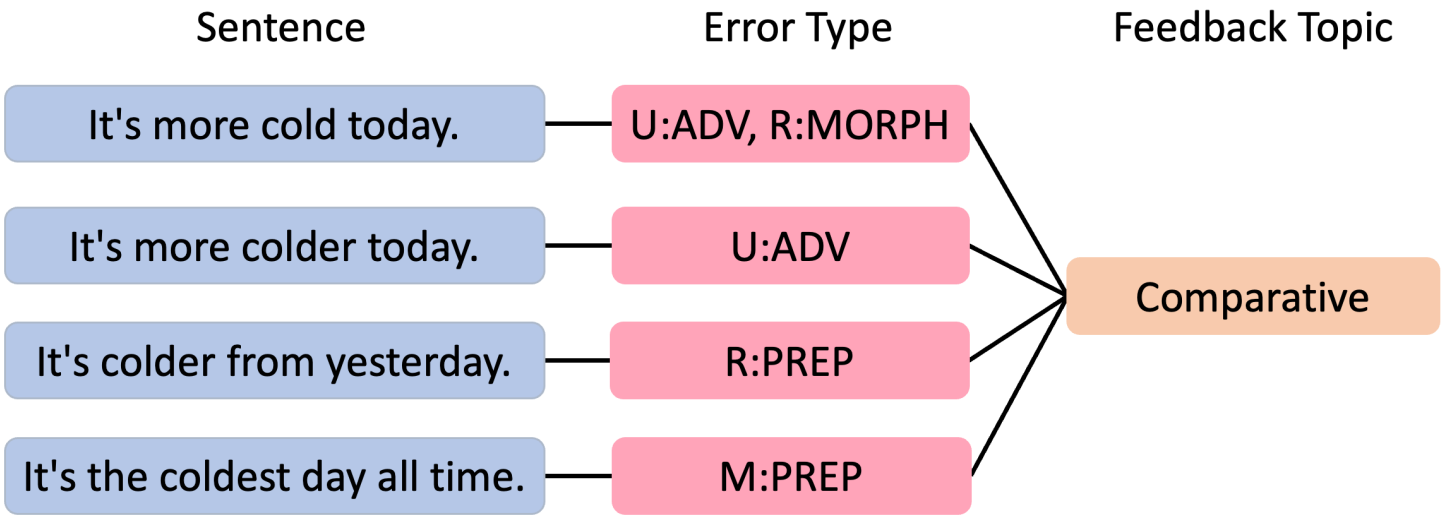


Example: One Error Type, Many Topics



Error types from ERRANT (Bryant et al., 2017)

Example: Many Error Types, One Topic



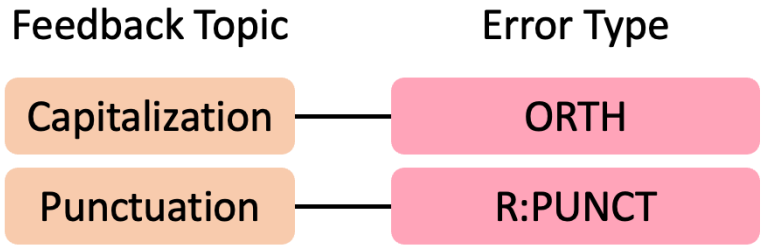
Error types from ERRANT (Bryant et al., 2017)

Typology: Principles

- We have decided to tag by “feedback topic” to capture this perspective.
- The basic principle is “What would an English teacher say is the issue here?”
- Three levels of abstraction:
  - Operational Topics
  - Grammatical Pattern Topics
  - Abstract Topics

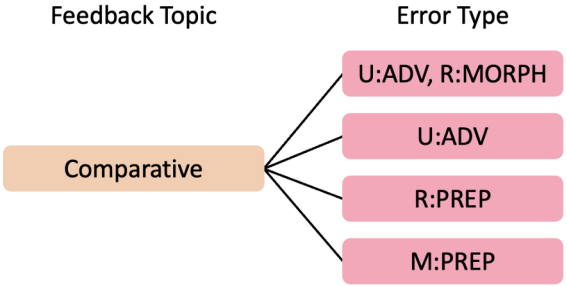
Operational Topics

- The comment is directly about a local error.
- These are expected to match closely with GEC error types. Some share names.
- Sometimes this is the best level to characterize an error-comment pair.
- Can otherwise be used as a fallback.



Grammatical Pattern Topics

- Represent broader grammatical constructions.
- Expected to map to a set of error types or particular mixtures of error types.
- Many resemble “grammar points” from education.



Tag Name	Example
Comparative	Study **[more hard → harder]** at school.
Preposition+Transitivity	I agree **[∅ → with]** that idea.
Purpose Clause	We go to school **[for → to]** learn.
Transition	**[But → However,]** tuition is expensive.

…

Abstract Topics

- Can not be mapped to any given error type or mixture of errors.
- Generating these may require specialized approaches.

Tag Name	Example
Fragment	**Obligation at home and at campus.**
Idiom	**[There's → That's]** the way it goes.
Language Transfer	I like riding **[jet → roller]** coasters.
Praise	(Various kinds of praise/encouragement)
Rewrite	(Explicit, complex revision suggestions)
Tone	It's **[cause → because]** of my experience.
Unclear	(Meaning unknown, even if error-free)

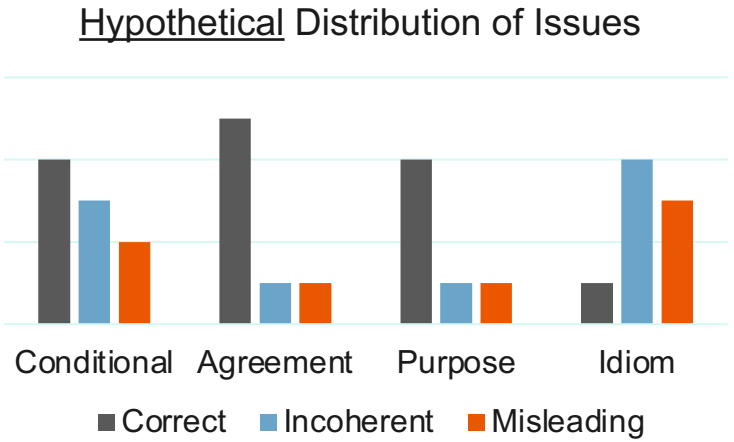
Analysis: Grouping

Once we have tagged data, we can analyze examples by category. This applies to both the data and generations.

Key Questions:

- Do different topics have different distributions of generation issues?

- Do some topics face unique issues?

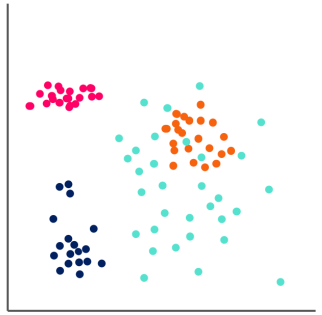


Analysis: Clustering

We can also compare the manual labels to clusters

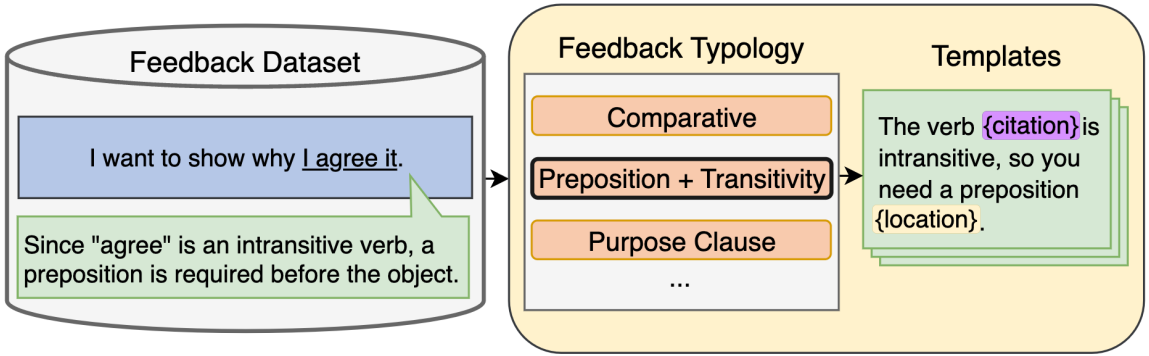
.

- Do the categories have different levels of sparsity in clustering?
- Will clusters resemble the human labels?



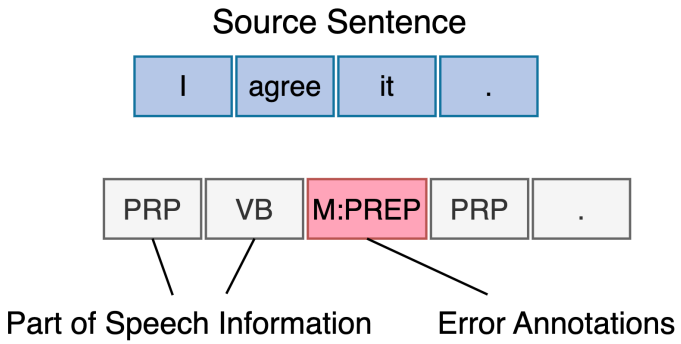
Templates

- Some types of errors are based on **particular words**, not **generalizable rules**.
- **Hypothesized to have a large number of issues.**
- Proposal: Create templates for these problem categories:



Linguistic Tags and Errors as Features

- Hypothesis: Adding structured information such as part-of-speech tags or GEC outputs can improve generation reliability.
- We will experiment with several combinations of annotations and report the impact on performance.
- Some involve creating a hypothesis correction.



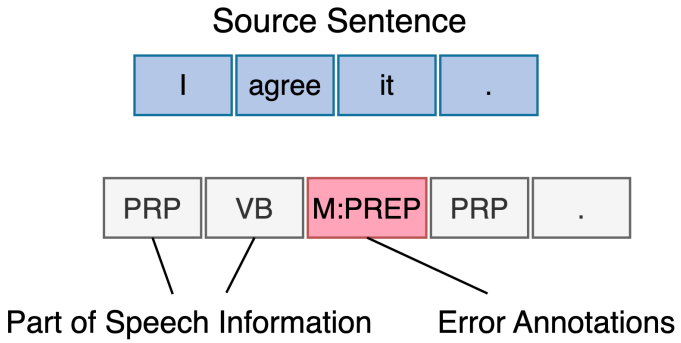
Simple example: Penn PoS tags plus error annotations from ERRANT.

Linguistic Tags and Errors as Features

- Hypothesis: Adding structured information such as part-of-speech tags or GEC outputs can improve generation reliability.
- We will experiment with several combinations of annotations and report the impact on performance.
- Some involve creating a hypothesis correction.

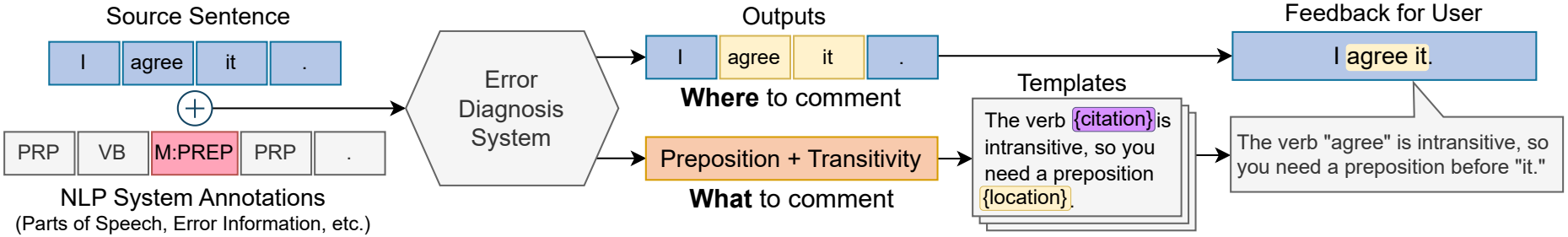
Candidates include:

- GECToR transformations (Omelianchuk et al., 2020)
- Edits from ERRANT or SERRANT (Choshen et al., 2021)
- Error-specialized parsing techniques (Morgado da Costa et al., 2022)
- CEFR-J Grammatical Items (Ishii and Tono, 2018)



Simple example: Penn PoS tags plus error annotations from ERRANT.

Example System Pipeline



We can adapt our generation strategies by comment type, e.g., neural text generation for general rules and template-guided generation for word-based issues.

Future Directions

- Adapt outputs to the learner's language level.
  - Feedback is only useful if the learner can understand it.

- Create additional corpora for this task.
  - We only have access to one dataset with a limited subset of English learners and comment writers. More data from more contexts is ideal.

Conclusion

This research presents:

- A typology of corrective feedback for grammatical errors.
- Analysis of feedback data and generations by type.
- Experiments with templates and supplemental inputs.

Thank you!

Inquiries:
coyne.steven.charles.q2@dc.tohoku.ac.jp