

「教育アセスメント×言語処理シンポジウム」 の開催に当たって

乾 健太郎

理化学研究所 AIPセンター
東北大学 情報科学研究科



AIP 理研AIPセンター 自然言語理解チーム

応用テーマ：教育アセスメント×言語処理



教育における言語

思考力・判断力・表現力の
訓練・実践の媒体

- 記述式答案、作文・エッセイ
- 学習者のメモ、振り返りログ
- 教員の解説・フィードバック



(かなり高度な) 自然言語処理

採点・診断・添削
解説・分析・対話

教育アセスメント×言語処理

これまで 「自動採点」≡ エッセイの自動採点

学習者のエッセイ

Is air travel as safe as it should be ?

Statistically, air travel is reasonably safe when compared with road, rail and sea. Many countries have a high death and injury rate on the roads; train, ferry and ocean-going ship disasters are regular enough to be commonplace, taking a global view. yet air disasters are usually fatal to all or most concerned and are therefore more widely reported. Consequently there is international pressure to make air travel as safe as possible, and rightly so. Yet no form of travel can be made totally safe. The causes of disaster, whether natural or man-made, can never be completely eliminated.

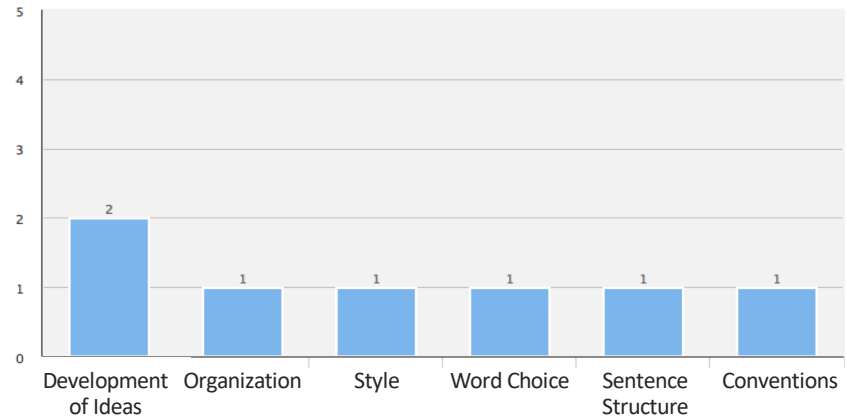
Looking first at natural causes, at least three can be identified. Bad weather is one. This includes storm, icing and air-pockets in the case of light aircraft; less so in the case of large jets with sophisticated instrumentation. One cause in the case of jets can be large flocks of birds or swarms of insects being sucked into the jet nacelles and thereby stopping the engines. This may happen at low altitudes. Little can be done to prevent this. A third is the alleged danger in what is known as the "Bermuda triangle" in which both ships and aircraft have been lost without trace. In the case of aircraft the reason is thought to be loss of horizon due to electrical disturbance.

Most dangers to aircraft however are man-made. The first and obvious danger is collision. In the busiest airports, especially in the tourist season, aircraft may take off as often as every twenty seconds. Much strain is imposed on aircraft dispatchers and traffic control generally. Clearance for dispatch is by radar and computer. Personal fatigue or mechanical failure in a radar center can be very dangerous. Aircraft often have to circle before being given landing permission, and collision or perhaps a near miss can result from mistakes in assigning heights. Most, though not all countries have strict regulations governing air traffic control. Any strike by operatives will cause grounding -- bad enough, but preferable to taking risks.

Another cause of trouble may be the age of the aircraft. Metal fatigue eventually sets in; cracks appear, bits of the aircraft may fall off, even whole engines, or the passenger cabin may lose compression. This means almost certain death to all passengers and crew, and there have been cases where an aircraft has been lost because one of the doors was not properly secured. The importance of safety checks before take-off is obvious. So also is the importance of regular and completely efficient servicing. Operatives' licenses can be withdrawn if government inspectors find inefficiency in this area. Most modern aircraft can shut down engines which catch fire, deal automatically with the fire, and proceed on three, or even two engines.

採点結果

Trait Scores



Word Choice

★ ★ ★ ★ ★
You scored 1 out of 5

Evaluation

- Have you chosen the correct words to express your ideas and to make the meaning of each sentence clear?
- Are you sure that the verbs, nouns and adjectives you used were the correct ones?
- Do you repeat the same words or phrases over and over?
- Are the words you have chosen specific and grade-appropriate?

Feedback

- Choose words more exactly.

Lessons

- Appositives
- Choosing the Right Word for the Job

Sentence Structure

★ ★ ★ ★ ★
You scored 1 out of 5

Evaluation

- Are the sentences you have written correct ones, or were there run-ons, fragments, and/or comma splices?
- Have you used correct grammar throughout your essay?
- Are your sentences choppy or awkward?
- Do any of your sentences cause confusion?

Feedback

- Your use of different sentence beginnings makes your essay more interesting.

Lessons

- Sentence Variety
- Subject Verb Agreement

できることは、もっとあるはず

短中期

中長期

how-to-say
(言語)

what-to-say
(内容・論理)

文法誤り解説生成

tipic: Gardening
student essay:
*I had ever been to gardening.
Because my mother's hobby is
gardening. And I beyond
gardening club for 11 years ...*

<過去完了>は、ある状態を表す場合に用いられる「(これまでに) ...した」という経験を表す場合は、<現在完了>で表します。
「I had perfect」 is used with a certain point of time to talk about something that happened before that point. Use <present perfect> to talk about experiences.

NEW

(EMNLP2019, NLP2019言語資源賞, LREC2020)

文法誤り訂正評価

CoNLL-2014	CoNLL-2013	FCE	JFLEG	KJ
Transformer 48.62	Transformer 46.20	LSTM 41.41	Transformer 40.06	LSTM 45.16
LSTM 48.57	LSTM 43.76	CNN 40.06	CNN 37.49	CNN 45.40
SMT 46.80	CNN 33.67	Transformer 39.57	LSTM 37.47	Transformer 42.80
CNN 46.16	SMT 32.30	SMT 26.79	SMT 42.35	SMT 32.04
				SMT 32.51

NEW

(IJCNLP2017, NAACL2019)

論述採点

Prompt: Should shops be open on weekends?
Student essay:
*I find it practical to be able to shop on weekends.
Sure, other people have to work on the weekend,
but they will have vacation during the weekdays.
Therefore, shops should be open on the weekend!*
Scores: Overall: ?
Clarity: ?
Fluency: ?
Coherence: ?
Logical consistency: ?
Persuasiveness: ?

open-domain tasks

論述リライト

NEW

Before
There are two main approaches to building CLIR system. The modular approach involves a pipeline of two components: translation (machine translation or bilingual dictionary look-up) and monolingual information retrieval (IR). These approaches may be further divided into the document translation and query translation approaches. In the former, one translates all foreign-language documents to the language of the user query prior to IR indexing; in the latter, one indexes foreign-language documents and

After
There are two main approaches to develop CLIR system. The modular approach involves a pipeline of two components, namely translation (machine translation or bilingual dictionary look-up) and monolingual information retrieval (IR). These approaches may be subdivided into the document translation and query translation approaches. In the former, one translates all the foreign-language documents are translated to the language of the user query prior to IR indexing, whereas in the latter, one indexes the foreign-language documents and translates the query to

(INLG2019, J.CogSci2020)

記述答案項目評価

NEW

問 傍線部(1)「こうした～築いてきた」とあるが、それはどういうことか。70字以内で説明せよ。

西洋文化の基礎の「対決」は神対人間(宗教・契約)、人間対自然(科学・合理主義)、人間対人間(個人主義)という形で現れるということ。日本人のように余計なことを言わないのではなく、他人に分かってもらうために言葉を尽くす「対決」のスタンスが西洋の文化を築いてきた。

① 他人は自分と対決のスタンスで自分の意見を主張する
② 他人は自分と対決のスタンスで自分の意見を主張する
③ 他人は自分と対決のスタンスで自分の意見を主張する
④ 他人は自分と対決のスタンスで自分の意見を主張する

減点対象
① 減点を減くして
② 減点を減くして
③ 減点を減くして
④ 減点を減くして

(BEA2019, ACL-SRW2020)

closed-domain tasks

論述評価と批判的コメント生成

Smoking **should be banned**! It causes bad breath and makes your clothes stinky!

While your point is **logically valid**, it does not justify banning the substance. Can you provide a more relevant reason?

Oh, I know that many people die of second-hand smoke each day!

Indeed, people dying from second-hand smoke is much more convincing than causing bad breath. **Where did you learn about this information?**

(AIED 2019, FEVER 2019)

できることは、もっとあるはず

例① 記述式答案の採点・解説（短中期課題）

問題文

① 西洋文化の基底の「対決」は神対人間(宗教=契約)、人間対自然(科学=合理主義)、人間対人間(個人主義)という形で現れるということ。日本人のように余計なことを言わないのではなく、他人に分かってもらうために言葉を尽くす「対決」のスタンスが西洋の文化を導いてきた。

問 傍線部(1)「こうした～築いてきた」とあるが、それはどういうことか。70字以内で説明せよ。

学習者の答案

(代ゼミ模試データ)

西洋文化の基底の「対決」は神対人間(宗教=契約)、人間対自然(科学=合理主義)、人間対人間(個人主義)という形で現れるということ。日本人のように余計なことを言わないのではなく、他人に分かってもらうために言葉を尽くす「対決」のスタンスが西洋の文化を導いてきた。

模範回答

他人を自分とは異なる考え方をもつ人間と見なす西洋では、自分の意見に同意を得るため、人を説得する技術が求められたこと。

減点対象

誤字・脱字1点減点
文末が「こと」終わってないもの1点減点

A 西洋(では)も可
西洋人は「も」可

B ①他人は自分と違うという説明:3点
対決のスタンス:2点
②自分と異なる考え方を持つ:2点

C (自分の意見に)同意を得るために:3点

D ①言葉を尽くして:3点
②他人を説得する:3点

採点基準(ルーブリック)

例② 英作文の添削・解説（中期課題）

トピック: ガーデニング

学習者の解答:

*I had ever been to gardening.
Becouse my mother's hobby is gardening. And I beyond gardening club for 11 years ...*

「have been to」は「...に行ったことがある」を表します。「ガーデニングをする」という表現で用いられる「動詞」を調べてみましょう。
«have been to» is used to express "an experience in visiting" Use another verb for "gardening".

<過去完了>は、ある過去の時点を示した上で、それまでの行為・状態を表す場合に用いられます。単に「(これまでに)・・・したことがある」という経験を表す場合は、<現在完了>で表します。
<Past perfect> is used with a certain point of time to talk about something that happened before that point. Use <present perfect> to talk about experiences.

«beyond»は<前置詞>で「...の向こうで;...を越えて」です。ここは<動詞>belongを用いて「...に所属している」と表してみましょう。
<<beyond>> is a <preposition> meaning "further away" or "outside a limit". Use the <verb> "belong" instead.

例③ 論述/ディベートの指導（中長期課題）

Smoking **should be banned**! It causes **bad breath** and makes your **clothes stinky**!

While your point is **logically valid**, it **does not justify** banning the substance. Can you provide a more relevant reason?

Oh..I know that **many people die of second-hand smoke** each day!

Indeed, **people dying from second-hand smoke** is much more convincing than **causing bad breath**. **Where did you learn** about this information?

ねらい

- 先端技術を用いた**新しい教育 / 学習方法**の開発
 - パフォーマンスの即時的な評価
 - インタラクティブなフィードバック、など
- 自然言語処理の**未解決重要課題**（骨のある問題）
 - 不完全な文章の頑健な解析
 - 採点/評価結果の説明
 - 書き手の「意図」の推測、など

ところが、研究者の参入少

〔問題〕 教育 × 言語処理：研究利用可能な**データが決定的に不足**

[希望] **データと技術を流通させる仕組み**を作ることができれば、
世界をリードできる可能性も

[illegible]

問 傍線部(1)「こうした～築いてきた」とあるが、それはどういうことか。70字以内で説明せよ。

西洋文化の基底の「対決」は神対人間(宗教=契約)、人間対自然(科学=合理主義)、人間対人間(個人主義)という形で現れるということ。
日本人のように余計なことを言わないのではなく、他人に分かってもらうために言葉を尽くす「対決」のスタンスが西洋の文化を導いてきた。

模範回答

減点対象

① 言葉尽くして 3点

② 他人を説得する 3点

③ 自分の意見に同意を得るために 3点

④ 他人は自分と違う 3点

⑤ 解決の手段を3点

⑥ 自分と異なる考え方を持つ 3点

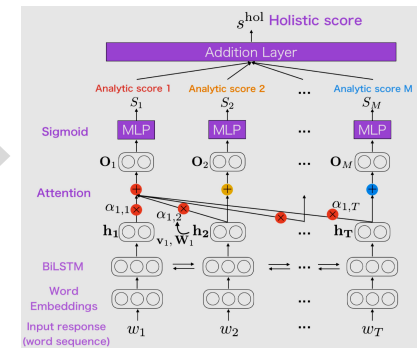
⑦ 西洋人とは違う 2点

⑧ 西洋人はよく可

12点



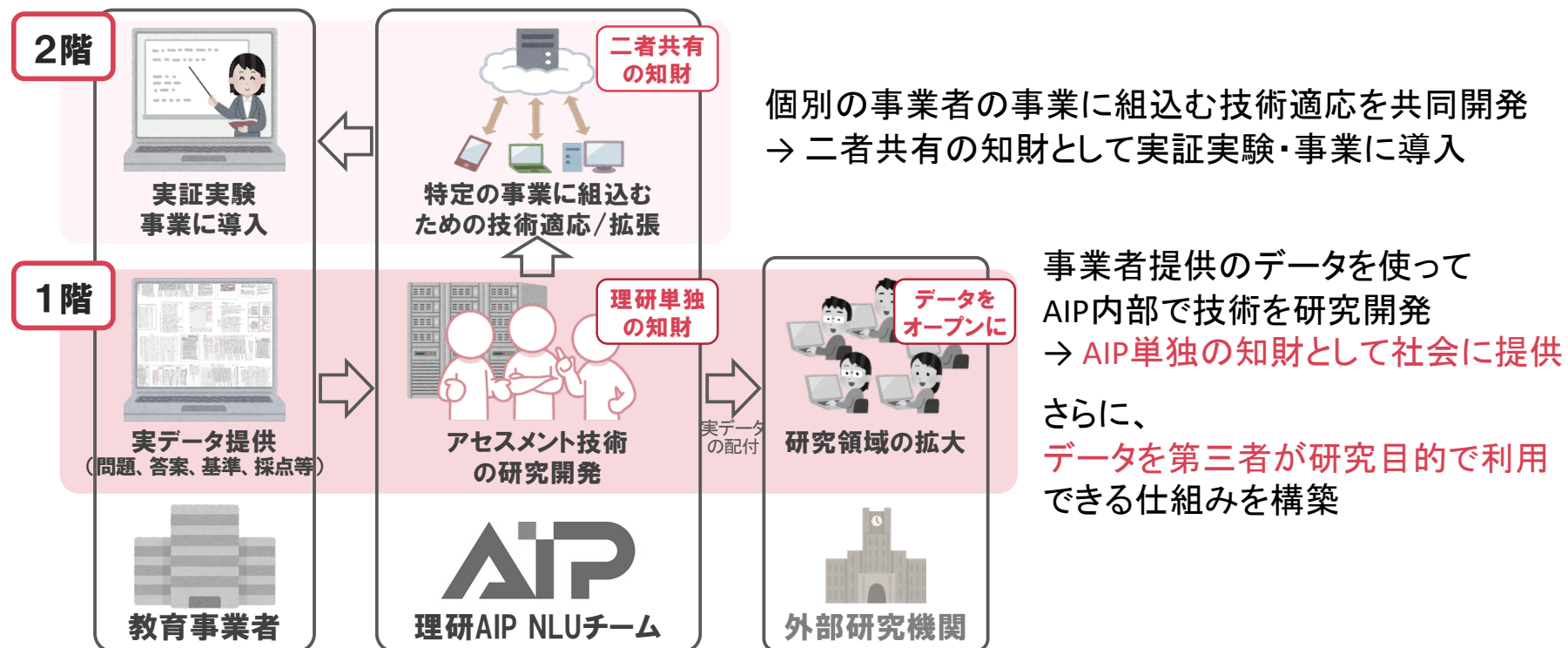
技術の研究開発



多様な教育事業者と2階建て連携

[問題] 教育 × 言語処理: 研究利用可能な**データが決定的に不足**

[希望] **データと技術を流通させる仕組み**を作ることができれば、
世界をリードできる可能性も



研究用データの公開 @IDR/NII

- データ提供の可否を提供元がコントロール
- 提供先ごとにデータ利用の覚書を交わして管理



国立情報学研究所 情報学研究データリポジトリ

HOME データ一覧 研究成果一覧 ユーザーフォーラム 組織 関連リンク 問い合わせ

> HOME > データ一覧 > 理研記述問題採点データセット

民間企業提供データ

- Yahoo!データセット
- 楽天データセット
- ニコニコデータセット
- リクルートデータセット
- クリックパッドデータセット
- LIFULL HOME'Sデータセット
- 不満調査データセット
- Sansanデータセット
- インテジデータセット
- オリコンデータセット

理研記述問題採点データセット

国立情報学研究所が研究者提供データセットとして理化学研究所から受け入れたデータセットです。

2020/07/14 作成

データ概要

大学受験予備校や教育関連会社にて行われたテストの記述問題の解答に対し、理化学研究所において採点アノテーションを付与したデータセットです。今後も順次同様のデータの追加を予定しています。

1. 代々木ゼミナールデータ（2020年4月時点、9問各約2000件）

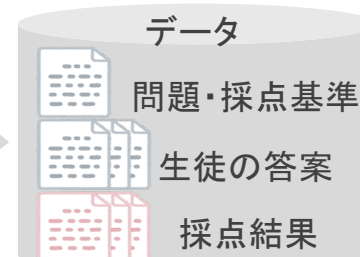
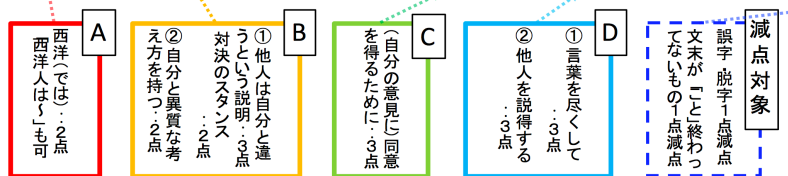
2014年～2015年に行われた高校生対象模擬試験の国語記述問題に採点アノテーションを付与したものです。全体点および採点項目ごとの部分点や、その根拠となる単語・フレーズがアノテーションされています。

データの詳細は理化学研究所の「記述答案採点の自動化に関する研究」のページもご参照ください。

また本データを用いた研究成果として理化学研究所が発表した国際会議ACL併催ワークショップBEA14の論文もご参照ください。

第1弾：代々木ゼミナールデータ

西洋文化の②基底の「対決」は③神対人間(宗教=契約)、人間対自然(科学=合理主義)、人間対人間(個人主義)とい
日本人のように余計なことを言わないのではなく、他人に分かってもらうために③言葉を尽くす「対決」のスタンスが②



高校現代国語
記述21問 × ～2,000答案
(これでも世界最大)

データ・技術の集積・共有へ オープンなスキームを創っていききたい



言葉の意味がわかっているわけではない

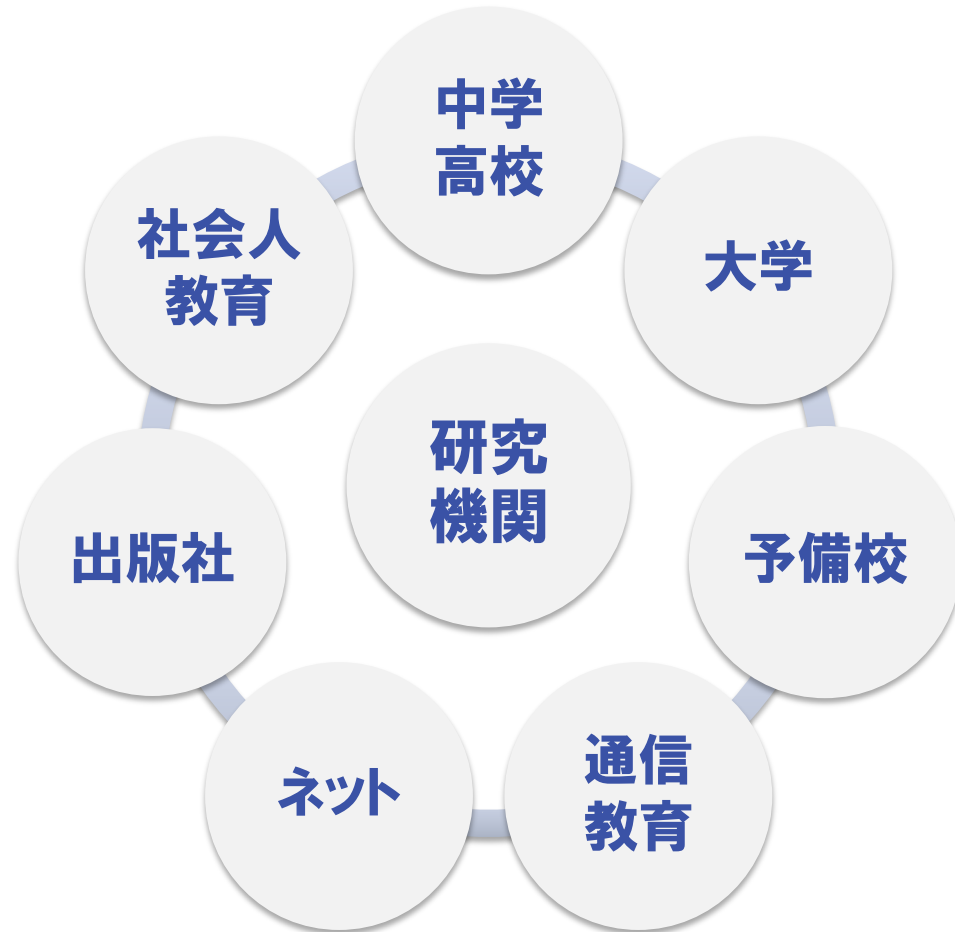


The screenshot shows the DeepL website interface. At the top, there's a navigation bar with the DeepL logo, links for '翻訳ツール' (Translation Tools), 'DeepL Pro', 'プランと価格' (Plans and Pricing), and 'アプリ' (App). On the right, there are buttons for 'Macにダウンロード 無料!' (Download on Mac - Free!), 'ログイン' (Login), and a menu icon. Below the navigation bar, there are two tabs: 'テキストの翻訳' (Text Translation) and '文書ファイルの翻訳' (Document File Translation). The main area is divided into two columns. The left column is labeled '原文 日本語 (自動検出)' (Original Japanese (Auto-detected)) and contains a text box with the following Japanese text: 「教育アセスメント」と「自然言語処理」の接点をテーマとしたシンポジウムを開催します。理化学研究所では、記述式答案や論述文など、学習者の記述・論述を自動解析し、「どこがどう良いか/不足か、どう直すとよいか」を評価・説明・助言する新しいAI技術の研究開発を進めています。本シンポジウムでは、下記のような方々に向けて、我々のチームで進めている研究開発を紹介するとともに、教育事業者をお招きしてパネル討論を行い、教育アセスメント×言語処理の可能性を参加者の皆さんと一。 The right column is labeled '訳文 英語 (US)' (Translation English (US)) and contains the English translation: A symposium on the interface between "educational assessment" and "natural language processing" will be held. Our research and development at RIKEN is focused on the development of new AI technology that automatically analyzes learners' written answers and essays and evaluates, explains, and advises on what is good, what is lacking, and how to fix it. In this symposium, we will introduce our team's research and development to the

例えば自動翻訳はずいぶん流暢になったけれど、

どのような技術をつくるか？ どう使うか？

- 使い方
 - 日常の学習の支援
 - テストの採点支援
- フィードバック
 - 答え？ ヒント？ 解説？
 - インタラクション
- 作題技術
 - 自動化に適した問題？
- 信頼性
 - 品質保証の方法
- 倫理問題
 - 技術導入の条件、ガイドライン



教育アセスメント×言語処理シンポジウム： 自動採点、英文添削、論述評価の可能性

技術紹介：教育アセスメント×言語処理の取り組み 司会：奥村学（東工大/理研）

1. 国語記述式答案自動採点の現状と実応用に向けて 舟山弘晃（東北大/理研）
2. 自動採点技術に基づく効果的な和文英訳学習 松林優一郎（東北大/理研）
3. ライティング学習支援のための文法誤り訂正技術の現状と今後の展望 三田雅人（理研/東北大）
4. ライティング学習のための解説文生成 永田亮（甲南大/理研）
5. 論述の構造解析による反論スピーチへのフィードバック生成 内藤昭一（リコー/理研）

パネル討論：教育アセスメント×言語処理の可能性 司会：乾健太郎（理研/東北大）

1. 代ゼミ自動採点における今後の取り組み
佐藤雄太郎（代々木ゼミナール教育事業推進本部長）
2. 思考力の評価・育成と言語処理技術への期待
伊藤素江（ベネッセコーポレーション 学校カンパニー GPS-Academic問題開発リーダー）
3. 模範解答をなぞる英作文学習から卒業しよう！弱点を知り主体的なやり直しを促す
支援ツールの開発
岡田健志（増進堂・受験研究社 NEXT LEARNING Labs主任研究員）
4. Classiにおける自然言語処理の取り組みと言語処理基礎研究に対する期待
伊藤徹郎（Classi データAI部長・データサイエンティスト）
5. ディープラーニングで経営教育のディープラーニングを目指す
鈴木健一（グロービス経営大学院 教授 兼GAiMERi所長）
6. 全体討論・意見交換

教育アセスメント×言語処理シンポジウム： 自動採点、英文添削、論述評価の可能性

アンケートについて

アンケートにご協力をお願いします。（所要時間3分程度）
シンポジウムの感想・意見をぜひお聞かせください。

メーリングリストへの参加

今後メーリングリストによって「教育アセスメント×言語処理」に関する情報共有や交流を図りたいと考えています。

参加を希望される方は、Googleフォームにて[参加登録](#)をお願いします。後日ご案内のメールをお送りします。

アンケートおよびメーリングリスト参加登録用のリンクは、本日お送りしたDoorkeeperの[メッセージ](#)をご覧ください。

教育アセスメント×言語処理シンポジウム: 自動採点、英文添削、論述評価の可能性

注意事項

- ① 発表者以外全員マイク&ビデオをオフをお願いします。
- ② 誹謗中傷はチャット、直接の発言ともに厳に慎んでください。

質疑応答について

質問される際は、チャット欄(送信先: 全員)に下記3点をご記入ください。

1. 所属
2. 氏名
3. 質問内容(簡潔に)

質疑の時間およびパネル討論時に、チャットで頂いた質問の中から司会が適宜ピックアップします。

指名された方は、マイクのミュートを解除し、直接ご質疑いただきます。