

# 記述式答案の採点・評価を 支援する言語処理技術

乾健太郎

理化学研究所／東北大学

# 乾 健太郎

専門分野: **自然言語処理** (言葉がわかるAIをつくる研究)

理化学研究所 AIPセンター 自然言語理解チームリーダー

1995年東京工業大学情報理工学  
研究科博士修了、博士(工学)。  
同大学助手、九州工業大学助教授、  
奈良先端科学技術大学院大学准  
教授を経て、2010年より東北大学  
情報科学研究科教授。  
2016年より理研AIP現職兼任。

言語処理学会 副会長  
国立情報学研究所 客員教授  
日本学術会議 連携会員  
NPO FactCheck Initiative 理事



東北大学 自然言語処理学研究室

**国内最大級の言語処理研究チーム**

スタッフ(博士取得者)10名、技術スタッフ8名  
学生: 博士17名 / 修士14名 / 学部5名

**応用テーマ：教育アセスメント×言語処理**



**教育における言語情報**

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



**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 ?

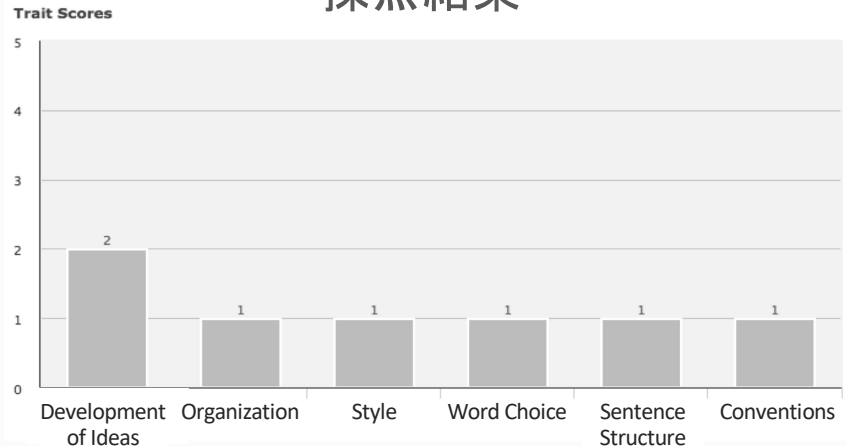
Statistically, air travel is reasonably safe when compared with road, rail and sea. Many countries have a high death and injury rate on the roads; train, ferry and ocean-going ship disasters are regular enough to be commonplace, taking a global view. yet air disasters are usually fatal to all or most 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



### 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  
(言語表現)

## 英文ライティング支援

**NEW**

Before

There are two main approaches to develop 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, 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. In

(INLG2019, J.CogSci2020, EMNLP2020)

## (解説付き) 英作文添削

△ since は既知の事実に基づく説明に使います

**NEW**

She spoke Japanese fluently since she had lived in Japan for thirty years.

She spoke Japanese very fluently because she has lived there for 30 years.

× 時制をそろえましょう

× 場所を正しく記述しましょう

(EMNLP2019, NAACL2019, EMNLP2019, ACL2020, EMNLP2020, LREC2020)

what-to-say  
(内容・論理)

**NEW**

## 記述答案項目評価



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使う「対決」のスタンスが西洋の文化を導いてきた。

A	B	C	D	減点対象	模範回答
西洋文化は、2000年を築いてきた。	西洋文化は、2000年を築いてきた。	西洋文化は、2000年を築いてきた。	西洋文化は、2000年を築いてきた。	西洋文化は、2000年を築いてきた。	西洋文化は、2000年を築いてきた。

(BEA2019, ACL-SRW2020)

## 論述の対話的指導

**NEW**

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)

# ねらい

- 先端技術を用いた**新しい学習／教育方法**の開発
  - 学習者のパフォーマンスの即時的な評価
  - 適応的なフィードバックによる対話的学習環境
  - 学習者のパフォーマンスの集合的な分析
  
- 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**」
    - 「（自分の判断を）説明できるAI」はAIの重要課題
  - 書き手の「**意図**」の推測
    - 「（文脈からの）意図の理解」は言語処理の重要課題

# 教育アセスメント×言語処理の課題例

短中期

中長期

how-to-say  
(言語表現)

## 英文ライティング支援

**NEW**

Before

There are two main approaches to develop 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, 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, EMNLP2020)

## (解説付き) 英作文添削

△ since は既知の事実に基づく説明に使います

**NEW**

She spoke Japanese fluently since she had lived in Japan for thirty years.

She spoke Japanese very fluently because she has lived there for 30 years.

× 時制をそろえましょう

× 場所を正しく記述しましょう

(EMNLP2019, NAACL2019, EMNLP2019, ACL2020, EMNLP2020, LREC2020)

what-to-say  
(内容・論理)

**NEW**

## 記述答案項目評価

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使う「対決」のスタンスが西洋の文化を導いてきた。

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使う「対決」のスタンスが西洋の文化を導いてきた。

(BEA2019, ACL-SRW2020)

## 論述の対話的指導

**NEW**

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)

# 記述答案採点（短中期課題）

## 問題文



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。

① 他人は自分と違うという説明：3点  
② 自分と異なる考え方を持つ：2点

① 自分の意見に同意を得るために：3点

① 言葉を尽くして：3点  
② 他人を説得する：3点

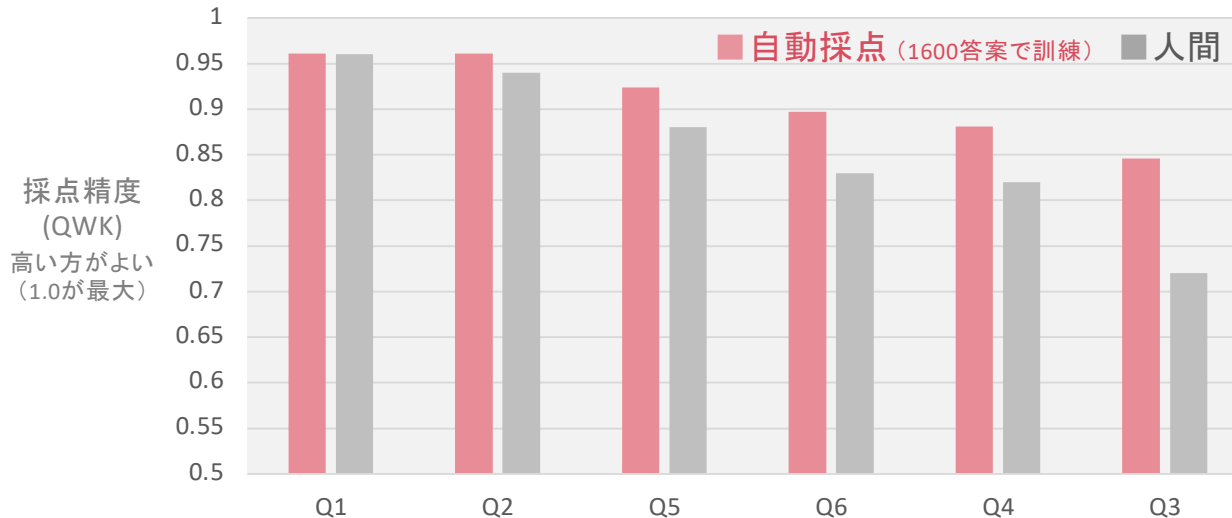
誤字 脱字 1点減点  
文末が「こと終わってないもの1点減点

減点対象

模範回答

## 採点基準（ループリック）

1問当たり1000答案程度の訓練データがあれば  
人間と同等程度の採点精度が得られる



(Mizumoto+ BEA 2019, Funayama+ ACL-SRW 2020)



# 記述答案採点（短中期課題）

## 日本語記述式問題自動採点デモ

次の文章を読んで、後の問いに答えよ。

西洋文芸の本質は（饒舌）にある。それに対して日本文芸の本質は黙説（寡黙）にある。黙説とは聞かない言葉だが、省略、含意、暗示、ほのめかしなどを能うかぎり言葉を抑制した表現全体を指す。西洋人は基本的には他人は自分と異なる人間（異人）と見なす。だから自分の考えに他人を同意させる必要があると考える。他人に分かってもらうために言葉を尽くして自分の考えを伝えようとする。説得である。それに対して日本人は他人は自分とおなじ人間（同類）と考える。だから余計なことは言わない。黙黙の前提に寄りかかる。私にいわせれば、西洋文化の基底には「対決」のスタンスがある。その「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。こうした緊張のスタンスこそが饒舌な西洋文化を導いてきたのである。（野内良三 「偶然を生きる思想」による）

問：下線部「こうした緊張のスタンスこそが饒舌な西洋文化を導いてきた」とあるが、それはどういうことか。句読点とも七十字以内で説明せよ。

解答入力欄:

西洋人は基本的に他人は自分とは異なる人間と見なすので、言葉を尽くして自分の考えを伝え、他人に自分の考えを同意させる必要があること。

I

0文字

Submit

合計点:  
12 / 16

西洋人は基本的に他人は自分とは異なる人間と見なすので、言葉を尽くして自分の考えを伝え、他人に自分の考えを同意させる必要があること。

採点基準A 基準 根拠事例

2 / 2

「西洋（では）」（＝話題の中心）・・・2点

※「ヨーロッパ」も可

※「外国」は不可

採点基準B 基準 根拠事例

3 / 5

「他人を自分とは異なる考え方をもつ人間と（みなす）」・・・5点

① 他人は自分とは違うという説明・・・3点

※「対決のスタンス」はBのポイント2点

②（異なる）考え方（を持つ）・・・2点

※Bのの違いの中身の具体化

採点基準C 基準 根拠事例

3 / 3

自分の意見に同意を得るために・・・3点

※「他人にわかってもらうため」なども可

採点基準D 基準 根拠事例

4 / 6

「言葉を尽くして他人を説得する」・・・6点

① 「言葉を尽くして」・・・3点

※「手を替え品を替え」や「あれこれ工夫して」などは2点

※「饒舌」そのままは1点

② 「（他人を）説得する」・・・3点

※「（自分の考えを）説明する」や「自分の考えを伝えようとする」は1点

誤字・脱字減点 基準

0 / 1

※誤字・脱字・転記ミス・主述のねじれなどは1点減点

文末減点 基準

0 / 1

※文末が「こと」、「事」でないものは1点減点

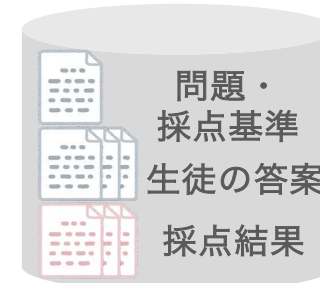
# 記述答案採点（短中期課題）

採点＝情報包含判別タスクと見なして解く

- － 答案が(採点基準で指定された)特定の情報を含んでいるか否かを判別するタスク

## 技術的課題

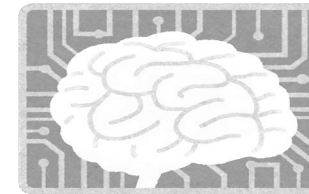
- － 答案の表現には幅があるため柔軟な意味的マッチングを自動化する必要がある  
e.g., 「他人に分かってもらう」～「同意を得る」
- － 個別の問題ごとに採点モデルを作る必要があるため、モデル作成のコスト削減が重要



訓練データ  
(採点済みのデータ)



機械学習  
(解答パターンの汎化)



自動採点モデル

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。

**A**

西洋(では)……2点  
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自分と異質な考え方を持つ……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

**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

模範回答

わ	人	に	意	な	考	他
れ	を	言	見	す	え	人
た	を	言	に	西	方	を
と	説	葉	同	洋	も	自
い	得	を	意	で	分	分
う	す	を	得	は	と	異
こ	技	得	る	、	人	質
と	術	し	た	自	間	な
。	が	め	の	見		
培	他					

文末-1

採点基準(ルーブリック)

(代ゼミ模試データ)

# 鍵の一つはデータ

[問題] 研究利用可能な**データ**が**決定的に不足**  
(研究者も少ない)

[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  
→ 複数の教育事業者と連携関係を展開中

NIIデータレポジトリで  
一部のデータを提供中(2020年8月～)



研究利用可能なデータがほとんどない

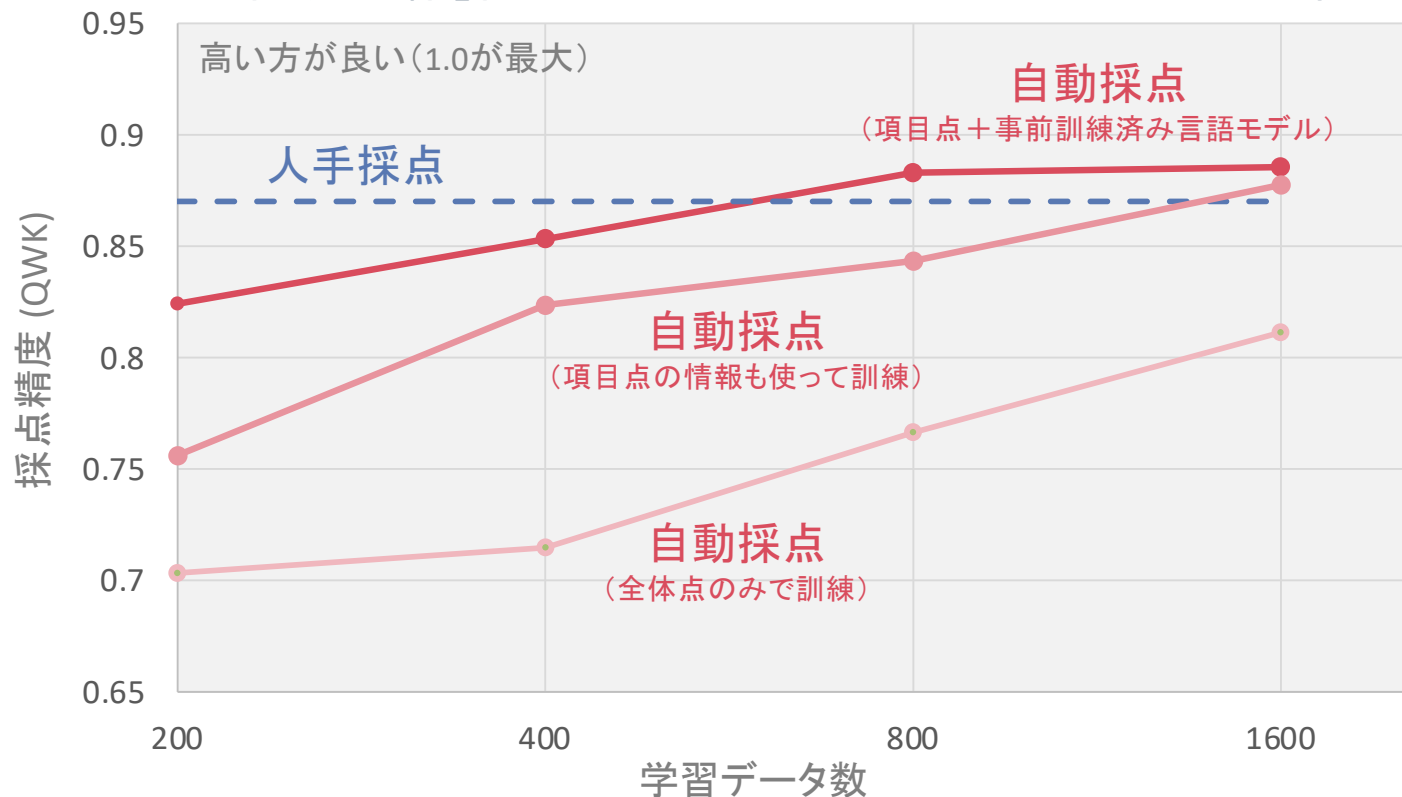
順次構築・共有を進めていく予定

	ASAP-SAS	X-CSD (Mohler+11)	SciEntsBank (Dzikovska+13)	国語記述 (高校国語)	英語記述 (準備中)
言語	英語	英語	英語	日本語	英語
ドメイン	自然科学、 読解など	計算機科学	自然科学	読解	準備中
問題数	10	80	135	24 (拡充予定)	準備中
答案数/問	～2000	28	36	～2000	準備中
点数	0-2 or 0-3	0-5	2 or 5 段階	0-16	準備中
項目点	×	×	×	✓	✓
採点基準	一部のみ	×	×	✓	✓
根拠箇所	×	×	×	✓	✓

# 現在の到達点

1問当たり1000答案程度の訓練データがあれば  
(平均的には) **人間と同等**の採点精度が得られる

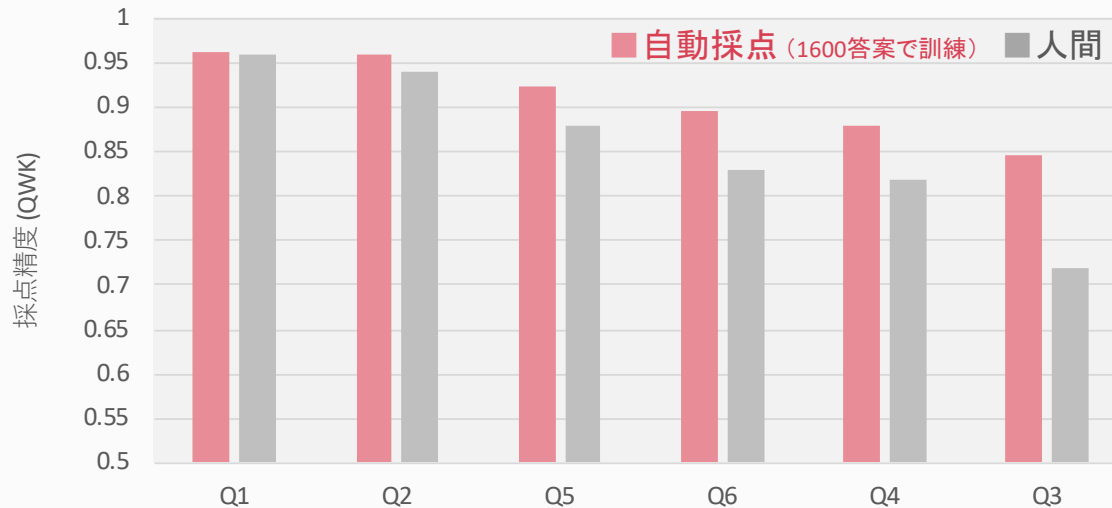
高校「国語」模試データ(代々木ゼミナール提供)での結果



(Mizumoto+ BEA 2019, Funayama+ ACL-SRW 2020)

# 採点が容易な問題と困難な問題

人間にとって採点が難しい問題は機械にとっても難しい



Q1 採点基準より一部抜粋

- ① 「他人は自分と違う」という説明 3点
- ② 「違い」の中身が具体化できていること 5点  
(「他人と異質な考え方」)

含むべき内容の**基準が具体的**  
 = 情報包含判別タスク

Q3 採点基準より一部抜粋

- ① 「未練が断ち切れていない」ことを表現 4点
- ② 困惑した心情の表現が入っていれば 2点  
(ネガティブなニュアンスの表現を広く認める)

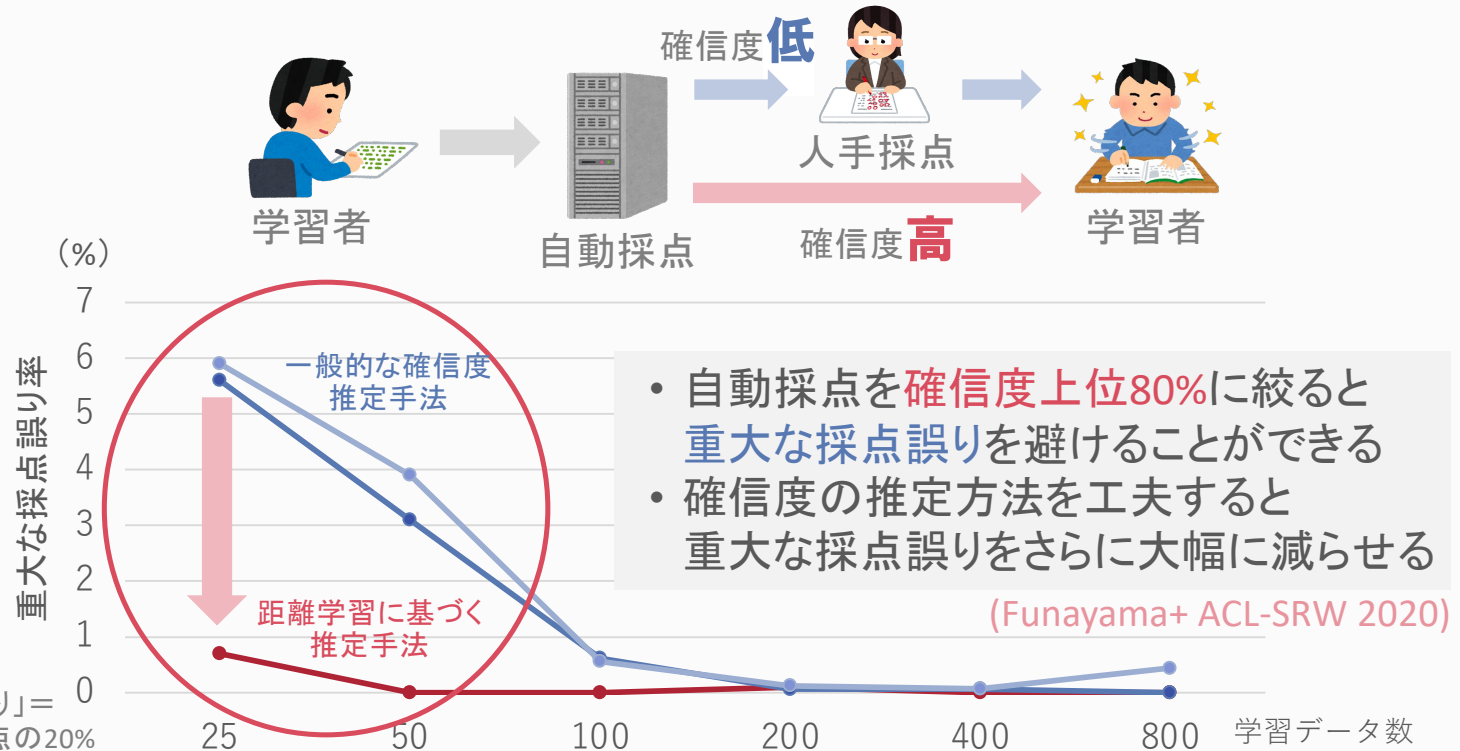
得点になる**表現が多様**、境界が**主観的**  
 → 人の判断もゆれる、自動採点も困難



# 確信度推定の導入

自動採点の個々の結果には確信度(**confidence**)を付与できる

- ユーザ(学習者・教員)に確信度を開示 → 納得感・信頼の向上
  - 何も説明がないと、「なんでこの答案でこの点数なのか？」
  - せめて確信度情報があれば、
  - 「モデルも『自信がない』と言っているので、採点が間違っているのかも」
- 「採点確信度が低い答案は人手採点に回す」という使い方も可能



# 採点項目ごとの採点と根拠箇所の提示

採点結果に対する**説明性** → 利用者の**納得感**

- 問題全体の得点の他、**評価項目ごとの点数**の推定も可能（世界初）
  - holistic scores vs. analytic scores
- さらに、項目ごとの**採点の根拠箇所**の提示も可能（世界初）
  - 根拠箇所 = 答案のどの部分が得点に繋がったか

2点
3点
3点
3点

採点の根拠箇所
 評価項目ごとの点数

西洋人は自分と異なる人間を異人と見なして考えを同意させるために、言葉を尽くして自分の考えを伝えようとする説得を行うこと。



# 使い方の想定例①：自動採点問題集

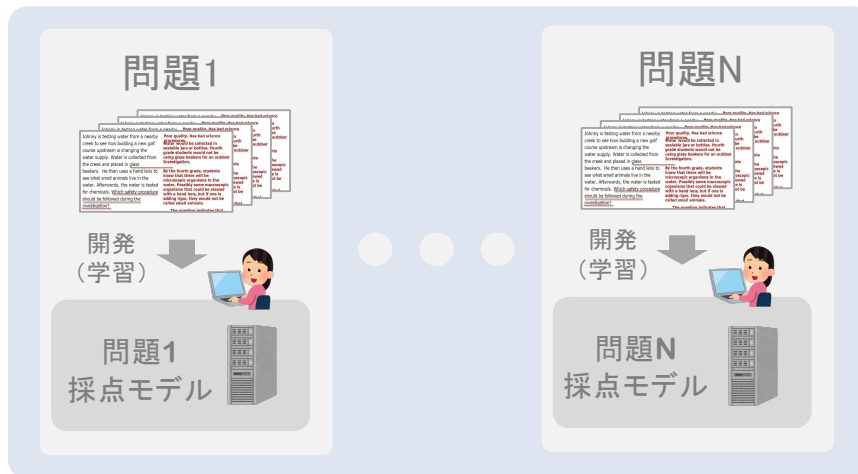
日常の学習に使える「問題集」の形で提供

- －クローズドな問題集合に対して採点モデルを訓練
- －問題ごとに訓練データが必要だが、一度開発すれば使いまわしも可

→ 即時的な採点による対話的学習環境の提供へ

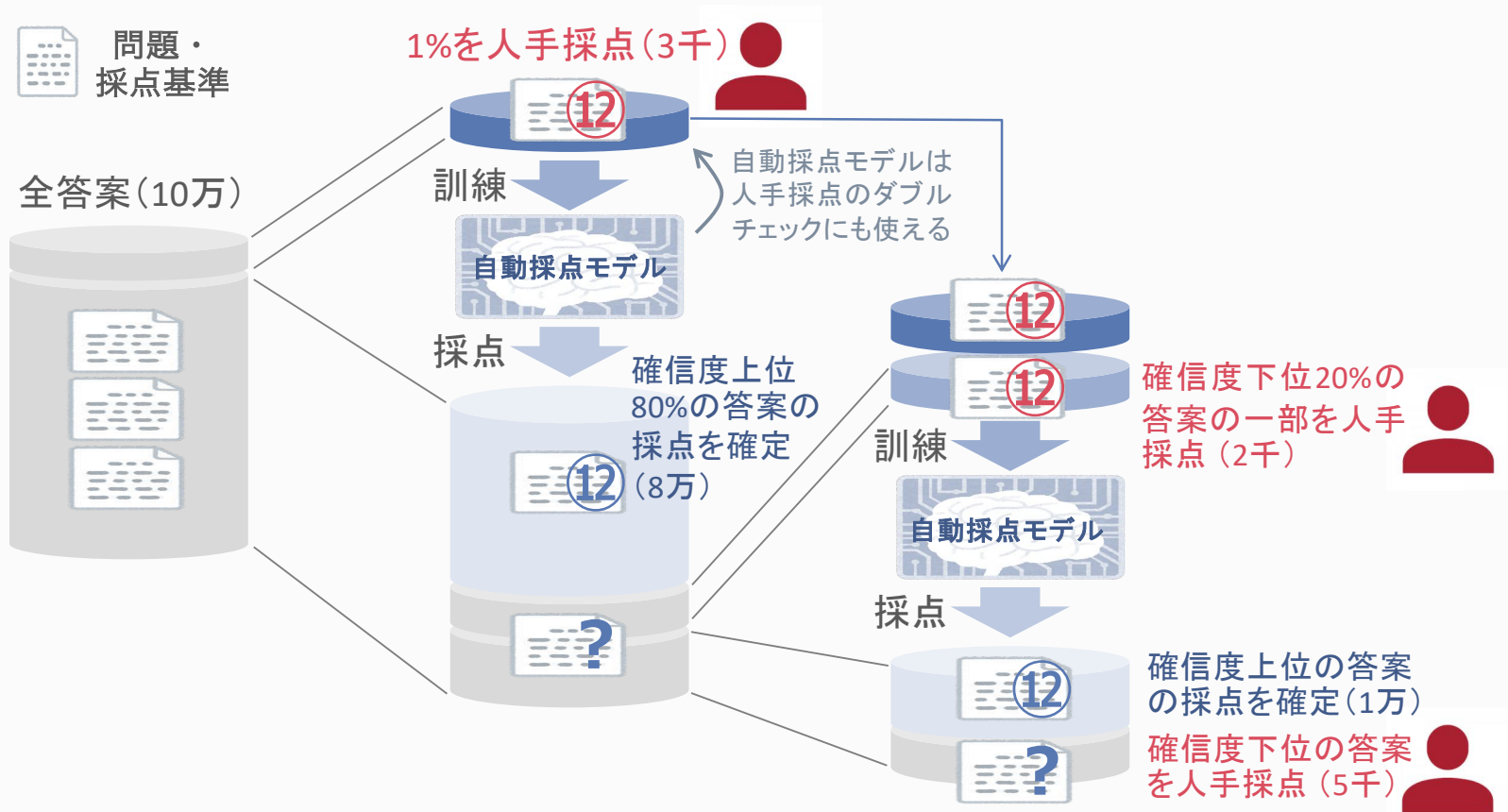
自動採点サービス付き問題集

インタラクティブな学習環境



# 使い方の想定例②：大規模試験の採点支援

## 人手採点と自動採点の併用：考えられるフローの例



自動採点の機械学習(=答案の類型化) → 答案の分析を支援できる可能性も  
今後少なくとも数万答案規模のデータで多角的に検証していく必要がある

# 英語ライティングのアセスメント

短中期

中長期

how-to-say  
(言語表現)

## 英文ライティング支援

**NEW**

Before

There are two main approaches to the 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, 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. In

(INLG2019, J.CogSci2020, EMNLP2020)

## 解説付き英作文添削

△ since は既知の事実に基づく説明にします

**NEW**

She spoke Japanese fluently since she had lived in Japan for thirty years.

She spoke Japanese very fluently because she has lived there for 30 years.

× 時制をそろえましょう

× 場所を正しく記述しましょう

(EMNLP2019, NAACL2019, EMNLP2019, ACL2020, EMNLP2020, LREC2020)

what-to-say  
(内容・論理)

## 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」のスタンスが西洋文化を築いてきた。

模範回答

減点対象

① 言葉を尽くして ② 他人を説く ③ 他人は自分と違う ④ 他人は自分と違う ⑤ 他人は自分と違う ⑥ 他人は自分と違う ⑦ 他人は自分と違う ⑧ 他人は自分と違う ⑨ 他人は自分と違う ⑩ 他人は自分と違う

(BEA2019, ACL-SRW2020)

## 論述の対話的指導

**NEW**

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)



# 英作文の解説付き自動添削

## 従来の教育現場での採点

① 次の日本語の下線を英語に訳しなさい。ただし、解答欄に与えられた語句で文を始めること。(10点)

私が最初に英語を教わった先生はアイルランド人でした。彼女は日本で30年暮らしていて、日本語がペラペラでした。

She spoke Japanese fluently since <sup>-2</sup> (8)  
she had lived in Japan for thirty years.

She spoke Japanese very fluently because  
she has lived <sup>-2</sup> there <sup>-3</sup> for 30 years. (5)

- × 採点コストが高い
- × 個別にコメントを返すのは大変
- × なぜその点数か、どう書けば良かったかが学習者に伝わりにくい

## 解説付き自動採点 (構想)

採点根拠の明示と追加学習項目のアドバイス

△ since は既知の事実に基づく説明に使います

She spoke Japanese fluently since she had lived in Japan for thirty years.

She spoke Japanese very fluently because she has lived there for 30 years.

× 時制をそろえましょう

× 場所を正しく記述しましょう

- 採点コストが低い
- 瞬時に採点結果がわかる
- 答案に応じた個別アドバイスが可能

# 基盤となる世界最先端の翻訳・訂正技術

- 自動翻訳国際コンペ (WMT 2020)、参加4種目すべて **人手評価1位**  
(理研・東北大・NTT)
- 英語文法誤り訂正国際コンペ (BEA 2019) で **世界2位** (理研・東北大)

Rank ▲	User ▲	Team Name ▲	TP ▲	FP ▲	FN ▲	P ▲	R ▲	F0.5 ▲
1	goo2go	LAIX	2618	960	2671	73.17	49.50	<b>66.78</b>
2	tomoyamizumoto	AIP-Tohoku	2589	1078	2484	70.60	51.03	<b>65.57</b>
3	arahusky	UFAL, Charles University, Prague	2812	1313	2469	68.17	53.25	<b>64.55</b>
4	hsamswcc	BLCU	3051	2007	2357	60.32	56.42	<b>59.50</b>
5	gurunathp	Aparecium	1585	1077	2787	59.54	36.25	<b>52.76</b>
6	mengyang	Buffalo	699	374	3265	65.14	17.63	<b>42.33</b>
7	nihalnayak	Ramaiah	1161	8062	3480	12.59	25.02	<b>13.98</b>

*The system is design to help student.*

**文法誤り訂正**

誤りが含まれる文から  
誤りが含まれない文への  
「翻訳」として実現

*The system is designed to help students.*

- トップ国際会議採択: NAACL 2019, EMNLP 2019, INLG 2019,  
ACL 2020, EMNLP 2020 × 2

# 英文ライティング支援

初期段階の下書きから流暢な表現を提案

## 誤りを含む不完全な文

*Our model (\*) to the SemEval-2007 task and (\*) good results.*

## 流暢で完全な文

*We **applied** our model to the Semeval-2007 task and **obtained** good results.*

*We **demonstrate that** our model **is able to achieve** good results **on** the Semeval-2007 task **and outperforms other baselines.***

***Experimental results showed that** our model **achieves comparable performance on** the Semeval-2007 task and **yields** good results.*

# 英語論文ライティング支援サービス

2021年春商用化予定

<https://editor.langsmith.co.jp/>



Langsmith

## Generating Fluent Sentences from Early-Stage Drafts for Academic Writing

1 \* Introduction

2 There are several stages in writing, including drafting, reviewing, editing, proofreading. Studies on existing writing assistance, such as Grammatical Error Correction (GEC), have focused only the early stages of writing. On the other hand, few studies have focused on supporting the early revision phase, which requires

On the other hand, only few studies have focused on supporting the early revision stage which requires the correction and extensive rewriting of the information contained in the text.

On the other hand, few studies have focused on supporting the early revision stage which requires the correction and extensive rewriting of the information contained in the text.

On the other hand, few studies have focused on supporting the early revision stage which requires the correction and extensive rewriting of the information contained in the texts.

On the other hand, few studies have focused on supporting the early revision phase which requires a correction and an extensive rewriting of the information contained in the text.

On the other hand, few studies have focused on supporting the early revision phases which requires the correction and extensive rewriting of the information contained in the text.

On the other hand, few studies have focused on supporting the early revision stages which requires the correction and extensive rewriting of the information contained in the text.

On the other hand, few studies have focused on supporting the early revision stage which requires the correction and extensive rewriting of the information contained within a given text.

In contrast, few studies have focused on supporting the early revision phase which requires the correction and extensive rewriting of the information contained in the text.



Langsmith

+ New document

+ New document with example

Generating Fluent Sentences from Early-Stage Drafts for Academic Writing

B I H2 H3

フィードバックが**すぐに返ってくる** → **対話的な学習環境へ**（菊池+ NLP 2021）



## フィードバック付き英作文自動採点デモ

次の日本語の下線部を英語に訳しなさい。ただし、解答に与えられた語句で文を始めること。

私が最初に英語を教わった先生はアイルランド人でした。彼女は日本で30年暮らしていて、日本語がペラペラでした。

解答入力欄:

She spoke Japanese well as she had lived in Japan for thurty years.

-語

Submit

She spoke Japanese well as she had lived in Japan for thurty years.

### 採点基準A

基準 アドバイス



「流暢な日本語」を適切に表現できる副詞として fluently などがあります。

### 採点基準B

基準 アドバイス



理由の従属接続詞 because を使いましょう。

### 採点基準C

基準 アドバイス



従属節内の時制が適切に書けている

### 採点基準D

基準 アドバイス



場所を表す表現(前置詞+名詞)が適切に書けている

### 採点基準E

基準 アドバイス



期間を表す語が適切に書けている

### スペルミス

アドバイス



thurty → thirty



# 論述の対話的指導（中長期課題）

ディベートの論理構造を内容（意図）レベルで解析し  
フィードバックを生成する技術（世界初の試み）

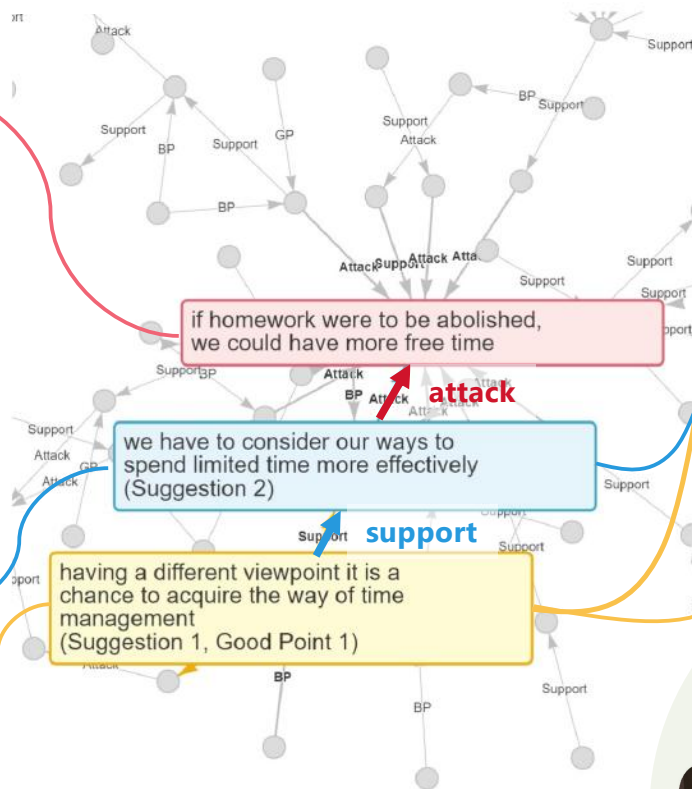
## 立論

Today's topic is "Homework should be abolished". We believe if homework were to be abolished, we could have more free time. As a result, we could do more of what we really wanted like club activities, hobbies, or playing with friends. In my case, I go to tennis club after class and then I go to cram school. After this full day, I start my homework at nearly 10:00 pm. I have a lot of homework, so I go to bed late at night and I don't have the opportunity to sleep for a long period of time. It is not healthy. Therefore, homework should be abolished.

## 反論（学習者）

I think club activities, hobbies, and playing with friends are necessary, but I also think being able to plan ahead and be able to do homework is also a necessary ability. It is because you can improve your ability to manage time later in life by planning it yourself. So I think homework which I can manage and do should not be abolished.

## システムの解釈



## フィードバック

### Suggestion 1:

タイムマネジメント力が将来にもいかに重要になってくるかまで説明できるとさらによいでしょう

### Suggestion 2:

どのように効率的に時間を使うか、具体的に話せるとよりよいでしょう。

### Good Point 1:

時間が限られていることは必ずしも悪いことではないと反論できたのがよかったです

## 論述の対話的指導へ

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?



# 論述の対話的指導（中長期課題）

Topic

School Uniforms: Good or Bad?

I think uniforms are good. Uniforms can remove the wall of status. They oppose the spirit of unity and its celebration. They also look very nice and respectful.

Suggestion: Change 'oppose' to 'can oppose'

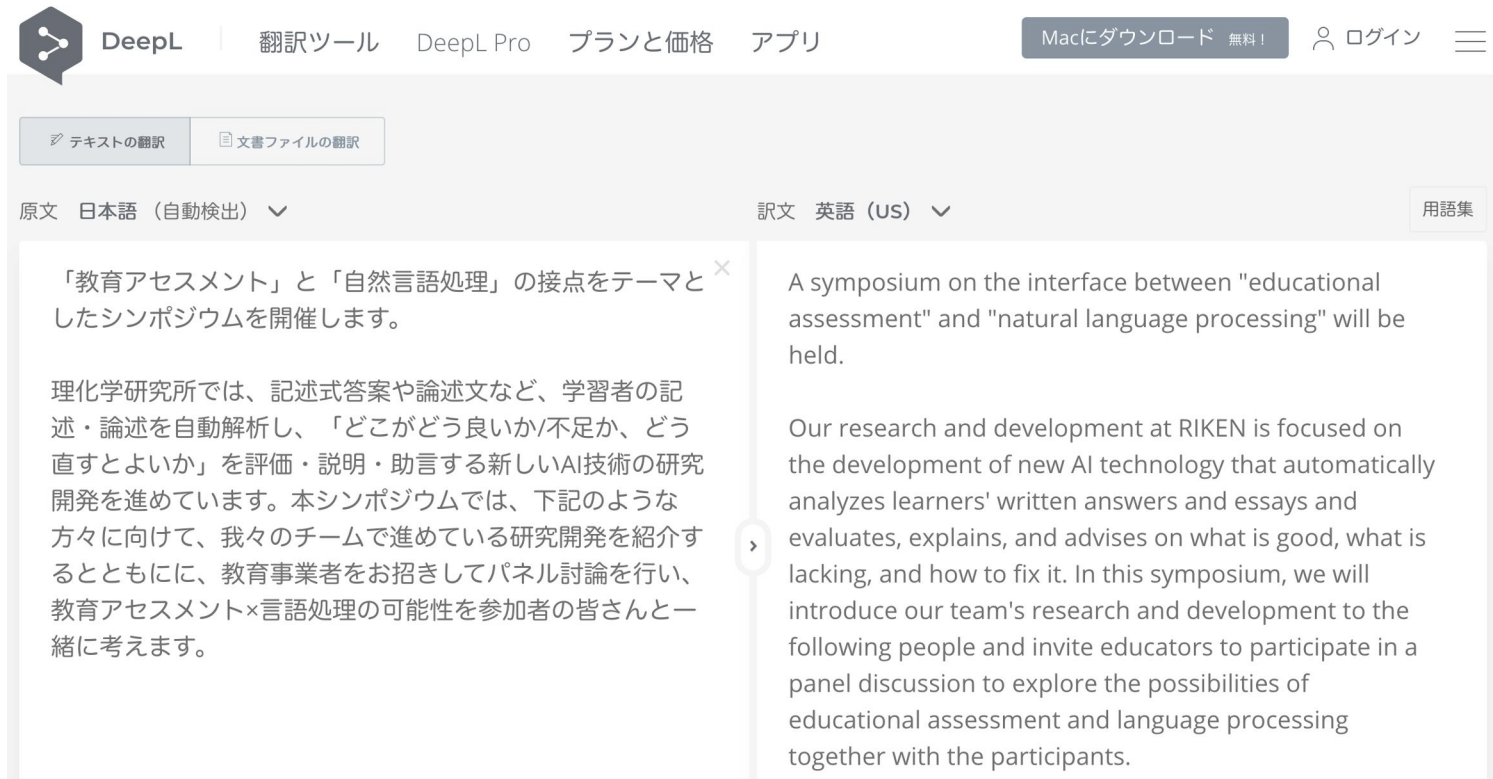
Suggestion: Add the phrase 'Based on my personal experience' to your argument.

Your argument: They also look very nice and respectful.

If you are making an argument based on your own experience, consider explicating this.

# AI は言葉の意味がわかっているわけではない

例えば自動翻訳はずいぶん流暢になったけれど、  
問題・答案を「理解」して採点・添削する AI が出現する訳ではない。  
人間の教師の代わりができるような技術にももちろんならない。



The screenshot shows the DeepL website interface. At the top, there are navigation links: "翻訳ツール" (Translation Tool), "DeepL Pro", "プランと価格" (Plans and Prices), and "アプリ" (App). There are also buttons for "Macにダウンロード 無料!" (Download on Mac - Free!) and "ログイン" (Login). 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 content area shows a translation from Japanese to English. The Japanese text on the left discusses a symposium on the interface between educational assessment and natural language processing, mentioning RIKEN's research. The English translation on the right is a direct translation of the Japanese text.

DeepL | 翻訳ツール | DeepL Pro | プランと価格 | アプリ | Macにダウンロード 無料! | ログイン

テキストの翻訳 | 文書ファイルの翻訳

原文 日本語 (自動検出) ▼ | 訳文 英語 (US) ▼ | 用語集

「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緒に考えます。

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 following people and invite educators to participate in a panel discussion to explore the possibilities of educational assessment and language processing together with the participants.

# 学習・教育を支援する仕組みのデザイン

「機械に何をさせて人間は何をするか」の丁寧な設計が必要  
 多様なプレーヤーが連携して検討・試行を積み重ねていくことが重要

- 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

公開シンポジウムを理研AIP主催で企画・開催（2020年11月27日）



TOP ▶ イベント ▶ 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

セミナー

いいね！

ツイート

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

Doorkeeper

2020/10/30 14:45

## 要旨

### ■趣旨

「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。

理化学研究所革新知能統合研究センター自然言語理解チームでは、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緒に考えます。

- 先端技術を用いた新しい学習方法の開発に関心をお持ちの教育関係者
- 先端技術を用いた新しい教育事業の展開に関心をお持ちの教育事業者
- 教育関連の新しい応用の開拓に関心をお持ちの言語処理研究者・技術者

皆様の積極的なご参加をお待ちしています。

### ■日時

2020年11月27日（金）13時～17時

### ■主催

理化学研究所革新知能統合研究センター（AIP）自然言語理解チーム

### ■共催

東北大学データ駆動科学・AI教育研究センター  
東北大学大学院教育学研究科

## 参加者324名 → 当領域への関心の高まり

教育施策機関	8	研究機関	22
大学	109	民間教育事業者	79
小中高校・高専	10	民間一般企業	81
		その他	17

## 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）
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）
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）
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）
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）
6. 全体討論・意見交換