

AI を使った中国語研究の可能性

山崎直樹（関西大学外国語学部／外国語教育学研究科）

5th International Symposium on Academic Writing and Critical Thinking 2024

AI: Deliverance, diversion, or dystopia

16th, February, 2024 (Nagoya University)

今日の話

1. 大規模言語モデルによる生成系AI
の特徴と、その上手な使いかた
2. 分析をさせる
3. 生成をさせる



言及した理論、フレームワーク、参考文献等
【↓】

<https://scrapbox.io/2024-02-16/INDEX>

生成系AIの特徴

- 大規模言語モデルによる対話生成系AI
 - 「いかにもありそうなことを無責任にべらべらしゃべる」
 - 言語的な確率に基づいている（意志や思考があるように見えるだけ）
 - 「**幻覚 hallucination**」に注意

上手な研究への利用方法

1. 自分が検証できる範囲で分析の手伝いをさせる
2. あれこれ生成させて、その中から自分が気づかなかったことを発見し、研究の端緒を見出す

(余談ですが……)

言語学習において学習者に生成系AIを利用させるときにも、同じことがいえます

生成系AIに分析をさせる

- （ありがたいことに） ChatGPTなどは、さまざまな言語理論や分析のフレームワークの知識をもっている。
- 新たに学習をさせることも可能
- 人手ではたいへんな分析を、いろいろさせてみたら、新しい発見があるかも

- 次は、文法の誤用のタイプと誤用の深刻度を規定した資料（自作）を使って学習したAIが、意図どおりに誤用を判定できるかどうかを試した例です。

- “今天比昨天冷一點”の句構造【↓】

(IP(NP-TMP(NT今天))(VP(PP(P比))(NP(NT昨天)))

(ADJP(AD冷))(DVP(QP(ADVP(AD一))(QP(CD点))))))

「学習」の例
(比較構文)

- 誤用の例(a)

(1) 誤用文: 今天冷一點比昨天

(2) 原因: (PP)が(DVP)の後にある

(3) TYPE: “前置詞句の位置の誤り

(4) 深刻度: 2

- 誤用の例(b)

(1) 誤用文: 今天昨天比冷一點

(2) 原因: (PP)の下の(P)がNPの後にある

(3) TYPE: 前置詞と前置詞の目的語の語順の誤り

(4) 深刻度: 3

この誤用判定器を何に使うか？

- 「階層構造の深いところでの誤りほど深刻である」という仮定で、深刻度の数値を決めて判定させています
- この（勝手に決めた）深刻度と、他の指標が一致するかしらないか？
- 「他の指標」とは……？

- 次は、AIが生成したglossを使って、複数の言語の間に意味的に対応する文の構造を比べたり、「構文の意味」を考えたりする試みです。

2つの言語の対訳文の構成要素

- ChatGPTは、Leipzig Glossing Rules に従った分析もしてくれます。
- (1) 対訳コーパスの中の一対の文を分析し、(2)そのglossの異同を報告する……という作業を、AIにさせたら？

(1)

彼	が	亡くなっ て	もう	3年	に	なる
he	NOM	die-PST	already	3-year	DAT	become

(2)

他	去世	已经	3年	了
he	die	already	3-year	ASP

「異同を分析して」と指示すると

- (1)には、(2)にあるアスペクトマーカ―「ASP」がありません。これは、中国語で行為や状態が完了したこと、または変化があったことを示すために用いられます。
- (2)には、(1)にある主格を示す助詞「NOM」と、時間的経過を示す構造「になる」がありません。また、「亡くなって」のような動作の連用形も見られません。これらは日本語で主語の役割、時間の経過、および動作の継続性や順序性を示すために重要です。

(1)

今日 は 飛行機 が 低く 飛んで いる

today TOP airplane NOM low fly PROG

(2)

今天 飛機 飛 得 很 低

today airplane fly DE very low

「異同を分析して」と指示すると

- (1)では進行形を示す「PROG」が用いられており、動作が現在進行中であることを示していますが、(2)ではこの概念が直接示されていません。
- (2)では、動作の程度を示す「DE」構造が用いられており、動詞の後に動作の質や程度を具体化するために使われていますが、(1)ではこのような構造は見られません。
- (1)における「低く」は副詞として動作を修飾していますが、(2)では「很低」という形容詞的表現を用いて動作の程度を直接表しています。

「構文の意味」

- 構成要素に還元できない「意味」は、「構文の意味」？
- A語のある構文が、B語の複数の構文に翻訳されるとしたら、「構文の意味」が近いものと遠いものがわかるかもしれない？
- 「A語話者にとって理解しやすい／習得しやすいB語の構文」に根拠が与えられる？

- 次は、既存のフレームワークを使って、与えられた文にメタ情報を付与させる試みです。

メタ情報付きコーパス

- TreeBank: 品詞と階層構造の情報をもったコーパス
- PropBank: 項構造の情報をもったコーパス

……どちらもChatGPTがアノテーションを代行してくれます（仕様どおりかどうかは、チェックが必要ですが）

Propbank

- 自然言語処理の分野で使われるアノテーション付きコーパス
- どのようなアノテーションかというと……文中の動詞とその動詞に関連する引数（主語、目的語など）およびその動詞の意味的役割にラベルを付けることにより、文の意味構造を明確にする
- 動詞ごとに定められた「どんな意味役割の項を取るか」という定義→フレーム frame

Propbankの方式で分析して…

我打了一個電話給他。

- Predicate: 打
- Arg0: 我
- Arg1: 一个电话
- Arg2: 他

- これだけではぜんぜんおもしろくない……各項の「意味役割」が指定されていない
- 「どのフレームに基づくか」という情報がないと、意味役割の指定ができない
- 動詞に対して1つずつフレームを設定しているコーパスはすでにある

フレーム意味論

……ある語の理解にはその語と関連する世界知識へのアクセスが不可欠であるということである。例えば「買う」という語を理解するには、「〈売り手〉が〈買い手〉の提供する〈商品〉を同意した〈金銭〉と引き換えに交換する」という商取引についての背景知識が必須になる。この背景知識はフレームと呼ばれ、フレームを用いることによって言語を用いることで「何が理解されるのか」という情報を記述することができる。（「フレーム意味論」, Wikipedia)

AIに抽象的なフレームを作らせる

フレーム意味論のアプローチで次の3つの文を分析した場合、この3つの文に共通したフレームを設定することができますか？

1. 媽媽寄給我一封信。
2. 我打了一個電話給他。
3. 你可以給我發電子郵件。

これらの文に共通するフレームは「コミュニケーション」または「メッセージ伝達」と設定することができます。このフレームは、情報やメッセージが送信者から受信者へと伝達されるプロセスを捉えるものです。

フレーム意味論からPropbankへ

「メッセージ伝達」を共通のフレームとして用いて、この(1)(2)(3)の文をPropbankの方式でアノテーションしてください。

1. 媽媽寄給我一封信。
2. 我打了一個電話給他。
3. 你可以給我發電子郵件。

媽媽寄給我一封信。

Predicate: 寄

Arg0: 媽媽: メッセージの送信者

Arg1: 一封信: 伝達されるメッセージ

Arg2: 我: メッセージの受信者

AIに抽象的なフレームを作らせる(2)

フレーム意味論のアプローチで……この5つの文に共通したフレームを……

1. 她的話給了我不少啓發。
2. 我打了一個電話給他。
3. 請給我叫一輛汽車。
4. 媽媽給孩子講故事。
5. 我給妹妹做早餐。

共通のフレームは「サービス提供」または「援助」になります。このフレームは、あるエージェント（行為者）が別のエージェント（受益者）に何らかの形でサービスや援助を提供する状況を捉えます。

フレーム意味論からPropbankへ

「サービス提供」を共通のフレームとして用いて……

1. 她的話給了我不少啓發。
2. 我打了一個電話給他。
3. 請給我叫一輛汽車。
4. 媽媽給孩子講故事。
5. 我給妹妹做早餐。

她的話給了我不少啓發。

Predicate: 給

Arg0: 她的話: サービスの提供者

Arg1: 我: サービスの受益者

Arg2: 不少啓發: 提供されるサービスの内容

AIにテキストの生成をさせる

- AIの生成したテキストにgender biasはあるか？
- Metadiscourseを生成させたら？
- Semantic formulaを指定したテキストはどうなるか？

AIにgender biasがあるか

中国語を学習する学生に、次の条件を与え、「何を買ってやりたいか」を続けさせました。

- 如果我有1個6歲的兒子，我想給他買……
- 如果我有1個14歲的女兒，我想給她買……

その後、「自分の選んだ物にgender biasがあると思うか」と尋ねました

生成系AIに

- 我有1個5歲的女兒。我想給她買（ ）。
- 我有1個5歲的兒子。我想給他買（ ）。

結果

「女の子」

- ……。我想給她買一些兒童書。
- ……。我想給她買一套美麗的衣服。
- ……。我想給她買一個可愛的娃娃。
- ……。我想給她買一頂漂亮的帽子。
- ……。我想給她買一雙舒適的鞋子。

「男の子」

- ……。我想給他買一些積木玩具。
- ……。我想給他買一輛小型自行車。
- ……。我想給他買一套運動衣服。
- ……。我想給他買一本有趣的兒童書。
- ……。我想給他買一個遙控車。

- 次は、教科書にあるような会話を与え、「強い関心」あるいは「無関心」が伝わるように、談話を修正させる試みです。

Metadiscourse

話者や書き手が、聞き手や読み手との関係を構築し、自らの意図や態度を明確に伝えるために使用される言語的手段

- **Interpersonal** meta discourse: 話者や書き手が自分の立場を表明したり、読み手や聞き手の態度や反応を予測したりする手段
- **Interactive** meta discourse: テキストや発話の組織化をする手段

AIに生成させる

李老師: _____, 我把中國雜誌帶來了。

高木愛: _____, 我看看。

李老師: _____, 怎麼樣？都看得懂嗎？

高木愛: _____, 有些地方看不懂。

李老師: _____, 我不看了，你可以帶回去。

高木愛: _____, 我在家慢慢看。

1. ……その雑誌にとっても興味を持ちました。李老師への敬意と、高木さんの強い関心を表現してください。
2. 中国語の学習への興味を失っています。しかし、李老師には感謝をしています。李老師への敬意と、高木さんの無関心を……

李老師: 請看，我把中國雜誌帶來了。

高木愛: 謝謝您，李老師，我看看。

李老師: 那麼，怎麼樣？都看得懂嗎？

高木愛: 實話說，有些地方看不懂。

李老師: 不用擔心，我不看了，你可以帶回去。

高木愛: 好的，非常感謝，我在家慢慢看。

李老師: 請看，我把中國雜誌帶來了。

高木愛: 哇，真是太感謝您了，我看看。

李老師: 那麼，怎麼樣？都看得懂嗎？

高木愛: 大部分都懂了，但是，有些地方看不懂。

李老師: 沒問題，我不看了，你可以帶回去。

高木愛: 非常感謝您，李老師，我在家慢慢看。

- 次は、状況を与えて談話を生成させるだけではなく、意味公式を指定して、条件を統制する試みです

「誘いを断る談話」の生成

宮本（2023）

- 食事の誘いを断る談話にはどのような意味公式semantic formulaが見られるか
- その談話の中で、意味公式はどのように配列されているか
- 意味公式Semantic formula: 談話の中で特定の意味や機能を果たす言語表現のパターン

結論，理由，謝罪，感謝，提案，残念，呼称，保留，約束，条件，フィラー，承諾，要求……

こんなふうに生成させることもできますが……

あなたは、××の授業で出されたレポートの締め切りを、うっかり忘れていました。

レポートのテーマは、「××」です。

そこで、授業の担当教師である王先生にレポートの締め切りを3日延ばしてもらうお願いをするメールを書こうと思いました。

嘘でもよいので、何か理由を探して、お願いをしてみようと思っています。

……もっと条件を統制して、複数の談話を比較しやすくすることはできるだろうか？

意味公式の配列を指定して生成させられないか？

1. 宮本（2023）の結果を学習させる
2. 意味公式の配列を指定
 - A. [結論]→[理由]→[謝罪]→[提案]
 - B. [残念]→[理由]→[謝罪]→[結論]→[感謝]→[提案]
 - C. [フィラー]→[呼称]→[残念]→[理由]→[謝罪]→[感謝]→[提案]

A. 結論先行型

- あなたは、職場の上司の李主任に食事に誘われました。次の順で意味公式を配列して、拒絶してください。
- [結論]→[理由]→[謝罪]→[提案]
- 不方便出來吃飯。因為我今天晚上有家庭聚會，真的很抱歉。我們改天吧。

B. 理由先行型

- [残念]→[理由]→[謝罪]→[結論]→[感謝]→[提案]
- 真不巧，今晚我有家庭聚會。非常抱歉，不能陪你吃飯了。謝謝你的邀請，我們改天再約吧。

こんな型もありうる……

- 意味公式[結論]を使わない
- [フィラー]→[呼称]→[残念]→[理由]→[謝罪]→[感謝]→[提案]
- 哎呀，李主任，真是不巧，今晚已經有其他安排了。對不起給您帶來不便，非常感謝您的邀請。能否改日再約？

このように生成された談話から何を考えるか？

- 拒絶、依頼……などの談話のpolitenessは、意味公式の配列とも関係が深い
が、配列によるpolitenessと、言語表現自体のpolitenessとどちらがよく「効く」か？
- 意味公式の配列には文化的な嗜好がある……配列を変えた場合にはどのような違和感があるのか？

- 時間と手間のかかる分析を（そのために学習させた）AIにやらせて、その結果から何かを考える
- 条件を統制して、AIに言語表現を生成させ、考える端緒にする



言及した理論、フレームワーク、参考文献等
【↓】

<https://scrapbox.io/2024-02-16/INDEX>