

産業日本語研究会 文書作成支援分科会活動

20230209 橋田 浩一

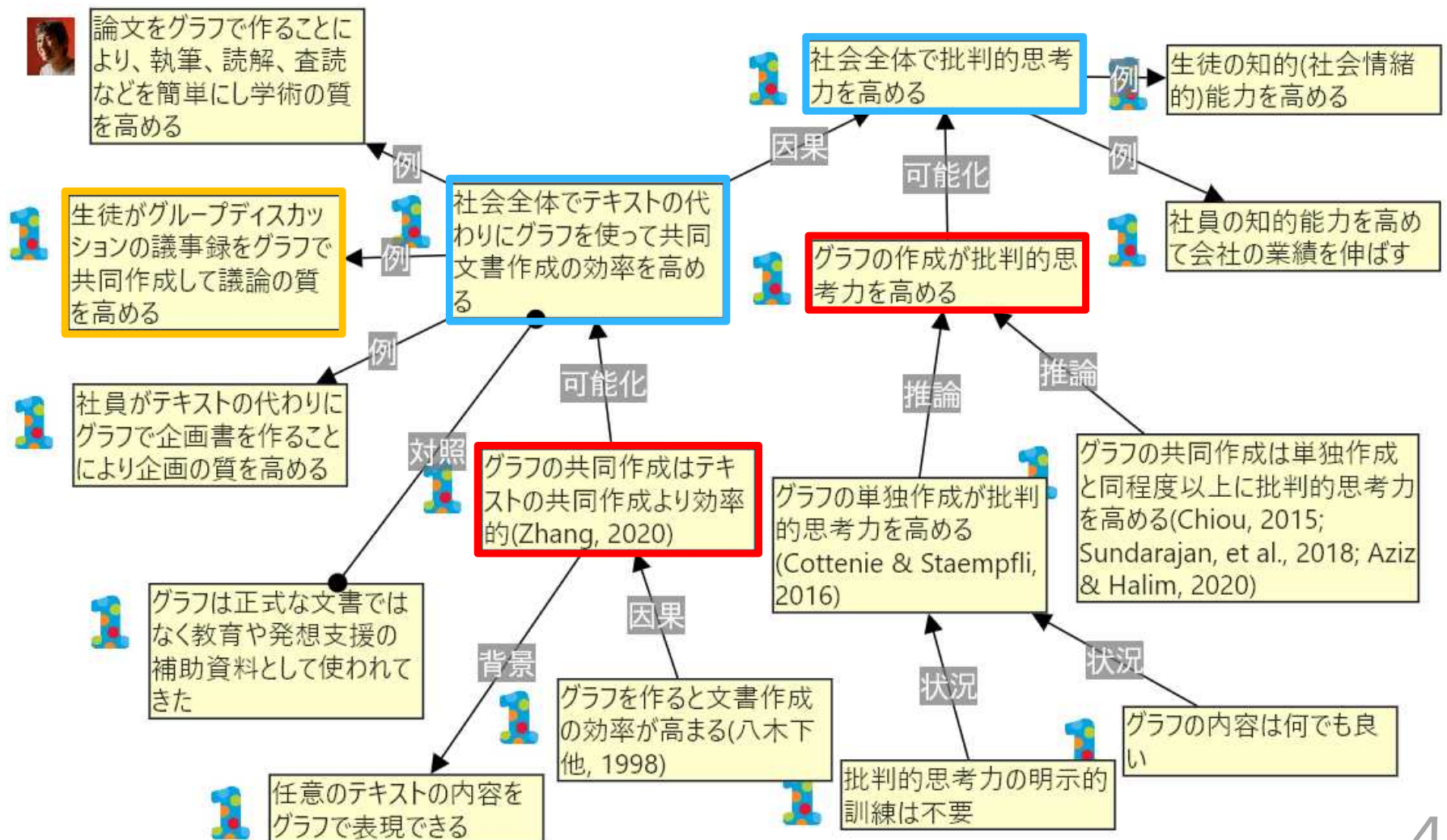
グラフ文書の普及のための実験

研究の概要

- **目的：グラフ文書を普及させて知的共同作業の生産性と批判的思考力を高める**
- **先行研究：グラフ文書作成の効果**
 - ◆ **長期的**
 - * 批判的思考力が高まる(Chiou, 2015; Cottenie & Staempfli, 2016; Sundarajan, et al., 2018; Aziz & Halim, 2020)
 - * 学業成績が上がる(Baig, et al., 2016; Marutirao & Patankar, 2016; Malekzadeh, et al., 2020)
 - ◆ **一時的**
 - * 文書・議論の質が高まる(Zhang, 2020)
 - 授業等の効果も高まるはず
- **実証：グラフ文書を学校教育の現場で運用できる**
 - ◆ **学習指導要領等との整合性**

グラフ文書

- ノードとリンクの型が規格化されているため作成も読解も楽なのでテキスト文書を代替可
- セマンティックエディタ(Personaryの機能)で作成・編集



教育へのグラフ文書導入の実証

目標: グラフ文書によって授業の効率と生徒の批判的思考力を高める

- 教員は多忙なので、学習指導要領等に沿った通常の授業に余分なコストをかけずに導入して授業の効率を高める必要がある
- 批判的思考力の向上は教育の目的に合致し、測定も比較的容易(標準的なテスト)
 - ◆ 企業等の業務の目的には合わない

手順(2022-10/2023-01)

	実験群	統制群
1.事前準備	<ul style="list-style-type: none">● CT試験(1回目)● セマンティックエディタの設定と練習	<ul style="list-style-type: none">● CT試験(1回目)
2.授業	<ul style="list-style-type: none">● 5回以上の授業で小グループのディスカッション中に議事録をグラフ文書で共同作成	<ul style="list-style-type: none">● 通常の授業
3.事後調査	<ul style="list-style-type: none">● CT試験(2回目)	<ul style="list-style-type: none">● CT試験(2回目)
4.データ提供	生徒本人と保護者の同意が必要 <ul style="list-style-type: none">● グラフ文書のデータ● 各生徒の学科試験の成績とCT試験の解答	同意は不要 <ul style="list-style-type: none">● 学科試験とCT試験の統計情報(人数と平均と標準偏差)

通常の授業へのグラフ文書導入は可能

グラフ文書の導入により授業のコストは増えず効率が高まる

- 教員が準備する資料は以前から作っていたのと同様のものであり、グラフ文書の導入によって準備のコストが増すことはないと考えられる
- 教員は生徒が作ったグラフ文書を見て即座にコメントしていたが、テキスト文書の場合はもう少し時間がかかるはず
- 生徒はグラフ文書エディタ(セマンティックエディタ)を苦もなく使いこなしており、リンクの種類を選択入力できることによってグラフの作成が簡単になることを認識していた
- 試験のデータの分析はまだ

グラフ文書のさまざまな活用

特許請求項のグラフ文書化

● 請求項の構造化

◆ クレーム分説支援ツールは請求項の全体構造を視覚化するわけではない。「構成要素単位に分割」とのことだが、実際は構成要素が長くて複雑。そもそも最近の請求項はほとんど改行入り。

◆ 要望1：現状の手法、精度、課題を知りたい

◆ 要望2：バッチツールとして公開してもらえると、その後処理ツール（詳細解析&グラフ文書化）の研究を進めやすい

● 発明拡張(augmented inventing)

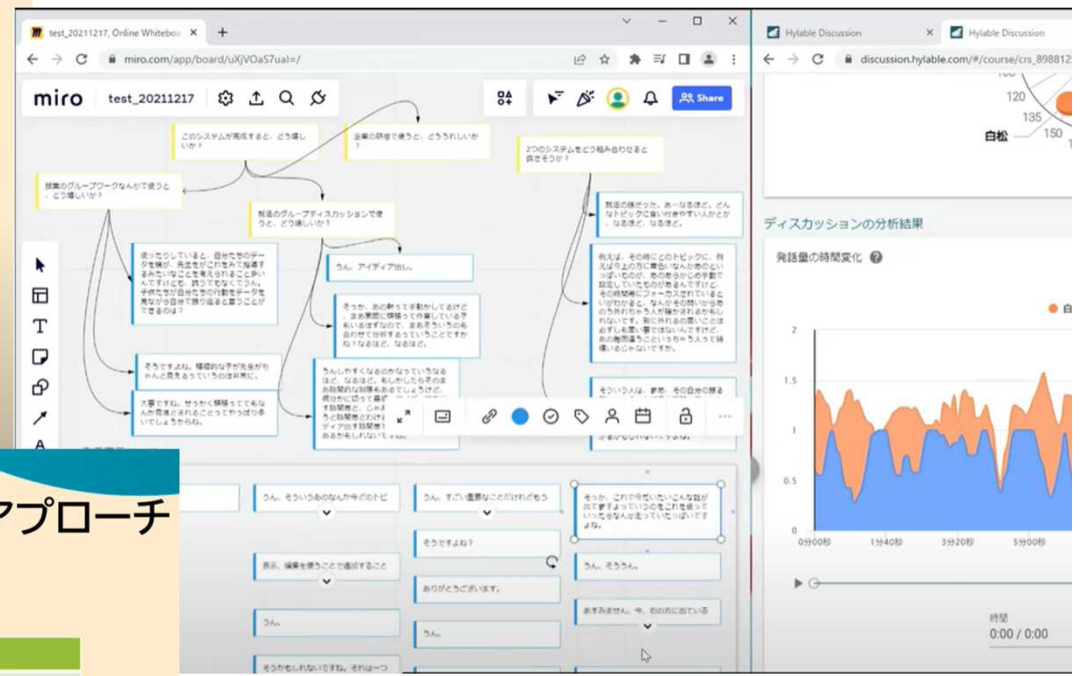
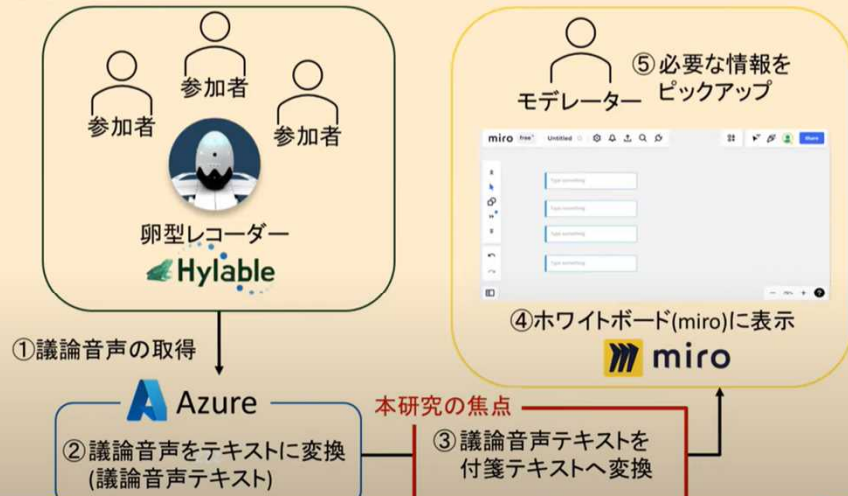
◆ GPT等の利用

◆ グラフ文書を入力すれば良い?

音声によるグラフ文書作成の支援

● 音声認識の結果をグラフ文書のノードに変換するシステムの開発

システム構成図



GPT-3で理解できる文へと言い換え・要約するアプローチ

議論音声テキストを要約して、付箋テキストとする

以下に例を示す

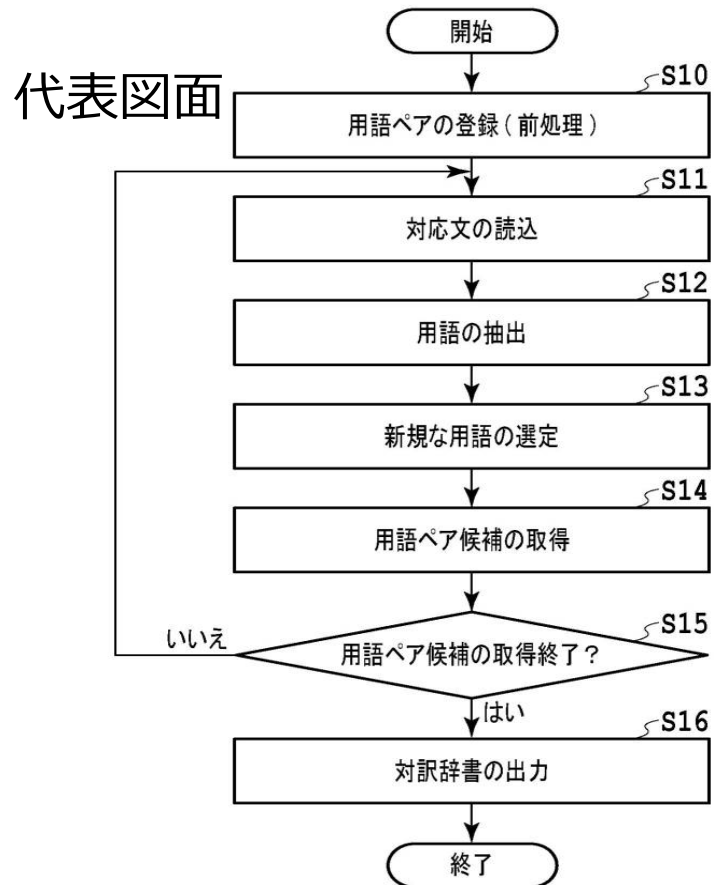
議論音声	議論音声テキスト	付箋テキスト
ゼミでの発表はいつにしましょうか。	ゼミでの発表はいつにしましょうか。	ゼミでの発表はいつにするか。
そうですね。	そうですね。	(なし)
うーん、学会の発表前がいいですね。	うーん、国会の発表前がいいですね。	ゼミでの発表は、学会の発表前がいい。

議論音声テキストから付箋テキストに変換するモデルの入力には、ターゲットとなる発言に加え、その前後の発話も含む

特に、前の発話に含まれる可能性が高いため、前は1分、後ろは30秒の発話を入力に加える

特許要約と代表図面の対応付け

- 日本国特許には技術分野の分類コードFタームが付与されている。
- 特許出願の際に、要約と代表図面が記載されている「要約書」を添付して提出することが義務付けられている。
- 特許の中には数多くの図表があり(1特許あたり平均21件)、その中には手順を説明するフローチャートがある。
- フローチャートに着目し、Fタームの分野ごとに、要約および代表図面から手順情報をまとめることで手順オントロジーを構築する。



要約

対訳コーパスから複数の対応文を読み込み S 1 1、複数の対応文から用語を抽出し S 1 2、抽出された用語が用語ペアテーブルに登録されている用語ペアを構成する用語以外である場合には、当該用語を、新規な用語として選定する S 1 3。複数の対応文のマッチングに基づいて、新規な用語のペアを用語ペア候補として取得し S 1 4、用語ペア候補の出現頻度に応じて、当該用語ペア候補を構成する新規な用語ペアを対訳辞書として出力するステップ S 1 6。取得するステップでは、複数の対応文の順序をランダムに変更して前記マッチングを繰り返し行う。

グラフ文書からテキスト文書への自動変換

GPT-3のfew-shot学習

