

卒業論文

日本語非母語話者のための 自動文書要約を用いた日本語対話支援システム

指導教官 村上 陽平 准教授

立命館大学 情報理工学部
先端社会デザインコース 4回生
2600170533-3

LI Chuang

2021年度（春学期）卒業研究3（CH）
令和3年7月2日

日本語非母語話者のための 自動文書要約を用いた日本語対話支援システム

LI Chuang

内容梗概

現在、日本の在留外国人数が増えている。日本統計局(e - Stat)が公開したデータでは、令和2年6月末に日本の在留外国人数は288万5千人以上になった。これらの外国人の半分以上は中国、韓国、ベトナムの出身である。日本にいる中国人、韓国人とベトナム人を合わせると、161万を超えている。中国語、韓国語、ベトナム語と日本語の間に、例えば、「愛人」という単語、日本語での発音は「愛人」、中国語での発音は「爱人 airen」、韓国では「애인 aein」、ベトナム語では「tình nhân」と発音の似ている単語はあるが、これらの言語間で文法は全く異なる。第二言語習得の過程では、すでに知っている言語が、新しい言語の学習過程に影響を与える言語転移が生じるため、在留外国人の母語と日本語の違いが大きいほど、日本語を学ぶのは難しくなる。このような問題に対して、チャットツールで即時翻訳によって日本語非母語話者の日本語理解の支援が行われている。しかしながら、この方法では、ユーザーの日本語レベルを向上させることができない。一方で、日本語による対話では日本語文が長くなるほど、外国人にとって理解はより困難になる。

そこで、本研究では、日本語能力の低い日本語非母語話者が日本語の長い対話を理解することを支援するシステムを提案する。具体的には、発話者の入力した日本語長文を要約ツールで処理し、各利用者にとって読解負荷を低減する長さの日本語に変換する。本提案手法の実現にあたり、取り組むべき課題は以下の2点である。

要約ツールとチャットツールの結合

日本語長文を要約するために、日本語自動要約ツールをチャットツールに組み込む必要がある。要約ツールには一文を短縮するものと、複数文から重要な文を抽出するものがあり、両方の自動要約ツールとチャットツールの結合によって、発話者からの日本語長文を理解しやすい形に変換する。

システムの有効性の検証

日本語能力レベルの異なる日本語非母語話者は、日本語長文を理解する能力も異なる。したがって、単文要約と重要文抽出のどちらのタイプの自動要約手

法がどの日本語能力レベルの日本語非母語話者に有効であることを明らかにする必要がある。

前者の課題で重要なのは、日本語長文の要約方式である。現在、要約方式は2種類ある。一つは複数文から重要文を抽出する方式である。日本語単文間の類似度を計算し、類似度が事前に設定した閾値以上の日本語単文を削除する。もう一方は、日本語長文を日本語単文に分割して、得られた日本語単文それぞれを短縮する方式である。前者はフィルタリングという特徴があり、非母語話者にとってトピックセンテンスを探し出しやすいという効果が期待できる。一方、後者は圧縮という特徴があり、非母語話者にとって単文のコアを探し出しやすいという効果が期待できる。

後者の課題に対しては、実験で2つの効果の検証を行う。1つ目は、提案システムがどの日本語能力レベルの日本語非母語話者に効果があるかを検証することである。2つ目は、どの要約方式が効果的なのかを検証する必要がある。これら2点を検証するために、ユーザーの発話が比較的に長くなるように、タブーゲームを用いて実験を行う。具体的には、キーワード1つと禁止ワード3つの単語セットを20個用意し、そのキーワードを禁止ワードを用いずに被験者に説明し、被験者にキーワードを推測させ、正答率と回答時間を計測する。被験者には、日本語能力がN3～N1の日本語非母語話者を選び、異なる要約方式で処理したキーワードの説明文に対して、その正答率と解答時間の差を比較する。

本研究の貢献は以下の通りである。

要約ツールとチャットツールの結合

自動文書要約を用いた日本語対話支援システムを提案した。文書要約ツールPYSUMMARIZATIONとSLPとチャットツールの結合により、日本語非母語話者が日本語長文の意味を理解できるようにする。

システムの有効性の検証

本研究では、タブーゲーム形式の日本語対話模擬実験を行った。実験データにより、自動文書要約を用いた日本語対話支援システムはN2とN3レベルの人のコミュニケーション効率を上げることがわかる。N2レベルの人にとって、約200時間の学習の程度を上昇した。また、複数文から重要文をフィルタリングする要約方式により、各単文を短縮の要約方式の方が良いことがわかる。

Japanese Dialogue Support System Using Automatic Document Summarization for Non-native Japanese Speakers

LI Chuang

Abstract

According to the data released by the Statistic Bureau of Japan, at the end of June 2020, the number of foreign residents in Japan had already reached 2,885,000. More than half of these foreign residents are from China, Korea, and Vietnam. The total number of Chinese, Koreans and Vietnamese in Japan have been exceeded 1.61 million. Even if there are similar words between Chinese, Korean, Vietnamese, and Japanese. For example, the word "愛人", the pronunciation in Japanese is "あいじん", the pronunciation in Chinese is "airen", the pronunciation in Korea is "애인 aein", and the pronunciation in Vietnamese is "tình nhân", but the grammatical structure is different among those languages. Normally, in the process of language acquisition, the native language has a great influence on second language learning, this phenomenon is called language transfer. For such problems, chat tools are used to support the understanding of Japanese by non-native speakers of Japanese through immediate translation. However, this method cannot improve the user's Japanese level. And the greater the difference between the second language and our first, the harder it will be for most people to learn.

Therefore, in this research, we design a system that helps foreigners who has a low proficiency to understand long Japanese sentences. Specifically, we summarize the long Japanese sentences by the summarization tool and convert it to the Japanese that is easy for every user to understand. To realize this system, there are two issues to work on.

Combine the summarization tools and chat tools

In order to summarize long Japanese sentences, it is necessary to search for a Japanese summarization tool and combine it with a multilingual chat tool. By combining the summarization tool and the chat tool, long Japanese sentences from the speaker can be easily understood.

Verification of the effectiveness of the system

People with different levels of Japanese proficiency have different abilities to understand long Japanese sentences. The Japanese proficiency of foreign residents in Japan differs depending on the actual situation. Therefore, it is necessary to verify whether this is effective

for foreigners with different Japanese abilities.

For the former issue, the first important thing is the summarization method for long Japanese sentences. Now there are two types to summarize the sentence. The first one is to extract important sentences from multiple sentences. The other is to divide the Japanese long sentences into single sentence, and to summarize each single sentence. The former has the characteristic of filtering and can be expected to have the effect of finding topic sentences for non-native speakers. On the other hand, the latter has the characteristic of compression, and can be expected to have the effect of finding the core of a simple sentence for non-native speakers.

For the latter issue, we conducted experiments to validate two hypotheses. First, it is necessary to verify whether this is effective for foreigners with different Japanese abilities. Therefore, the experiment is performed by a method like a taboo game. Specially, we prepared 20 word sets which include one keyword and three prohibited words and send the description of the keyword to the examinee, then let the examinee guess the keywords, and count the correct rate and time of completing the exam. Second, it is necessary to verify which summarization method is better. We invited the foreigners with different Japanese proficiency, and count the accuracy and completing time of foreigners with different Japanese proficiency in different tools. The following is the contribution.

Combine the summarization tools and chat tools

We proposed a Japanese dialogue support system using automatic summarization tool. By combining a sentence summarization tool with a chat tool, the system enables non-native speakers of Japanese to understand the meaning of long Japanese sentences. This system improves the efficiency of Japanese communication for non-native speakers of Japanese.

Verification of the effectiveness of the system

We conducted a simulated Japanese dialogue experiment in the form of taboo game. The experimental data shows that the Japanese dialogue support system with automatic summarization tool can improve the communication efficiency of N2 and N3 level users. The summarization method that shortens each single sentence is better than the method that filters the main sentence from multiple sentences.

日本語非母語話者のための 自動文書要約を用いた日本語対話支援システム

目次

第 1 章 はじめに	1
第 2 章 関連研究	3
2.1 非母語話者へのコミュニケーション支援	3
2.2 テキスト要約と文圧縮	4
第 3 章 自動文書要約を用いた日本語対話支援システム	6
3.1 システム概要	6
3.1.1 システムワークフロー	6
3.1.2 ユーザーインタフェース	6
3.2 システム構成	8
3.2.1 pysummarization	8
3.2.2 SLP	9
3.2.2 pysummarization と SLP の統合	12
第 4 章 実験	13
4.1 実験目的	13
4.2 実験説明	13
4.2.1 実験フロー	13
4.2.2 被験者の種類	14
第 5 章 評価	15
5.1 評価方法	15
5.2 実験結果	15
5.3 分析	16
5.4 考察	20
5.4.1 有効性の検証	20
5.4.2 問題点	20
5.4.3 これからの課題	21

第 6 章 おわりに	23
謝辞	24
参考文献	25
付録 :	26
A.1 実験問題集	26

第1章 はじめに

ワクチンパスポートの発行に伴い、時間がたつにつれて、日本に来る外国人の数も増えることが予想される。そして、日本統計局(e-Stat)が公開したデータでは、令和2年6月末に、日本の在留外国人288万5千人以上なり、これらの外国人の半分以上は中国、韓国、ベトナム出身である。中国語、韓国語、ベトナム語と日本語の間に、例えば、「愛人」という単語、日本語での発音は「愛人」、中国語での発音は「爱人 airen」、韓国では「애인 aein」、ベトナム語では「tình nhân」と発音の似ている単語はあるが、これらの言語間では文法は全く異なる。第二言語習得の過程では、すでに知っている言語が、新しい言語の学習過程に影響を与える言語転移が生じるため、在留外国人の母語と日本語の違いが大きいほど、日本語を学ぶのは難しくなる。この状況では、日本語能力不足により、コミュニケーションがうまくできない場合もある。このような問題に対して、多言語チャットツールによるリアルタイム翻訳によって日本語非母語話者の日本語理解の支援が行われている。しかしながら、この方法では、ユーザーの日本語レベルを向上させることができない。一方で、日本語による対話では日本語文が長くなるほど、外国人にとって理解はより困難になる。

そこで、本研究では、日本語能力の低い日本語非母語話者が日本語の長い対話を理解することを支援するシステムを提案する。具体的には、発話者の入力した日本語長文を要約ツールで処理し、日本語非母語話者にとって読解負荷を低減する長さの日本語に変換する。本提案手法を実現するにあたり、取り組むべき課題は以下の2点である。

要約ツールとチャットツールの統合

日本語長文を要約するために、日本語自動要約ツールをチャットツールに組み込む必要がある。要約ツールには一文を短縮するものと、複数文から重要な文を抽出するものがあり、両方の自動要約ツールとチャットツールの統合によって、発話者の日本語長文を理解しやすい形に変換することを可能にする。

システムの有効性の検証

日本語能力レベルの異なる日本語非母語話者は、日本語長文を理解する能力も異なる。したがって、単文要約と重要文抽出のどちらのタイプの自動要約手法がどの日本語能力レベルの日本語非母語話者に有効であるかを明らかにする必要

がある。

以下、本論文では 2 章において本研究に関連する非母語話者のコミュニケーション支援システムおよび自動文書要約技術の関連研究について述べる。次に 3 章では、提案手法である自動文書要約を用いた日本語対話支援システムについて、システムの概要およびシステム構成を説明する。続いて 4 章では提案システムの有効性を検証するための実験の目的と詳細、評価方法について述べる。そして、5 章では 4 章で行った実験結果について分析を示し、本システムの有効性と問題点について述べる。最後に本稿をまとめて結論する。

第2章 関連研究

本章では、従来の関連研究の概要と本研究で使用した自動要約技術について説明する。

2.1 非母語話者へのコミュニケーション支援

これまで、非母語話者の負担軽減を目的としたコミュニケーション支援ツールの研究が行われている。支援方式により大きく分けて2種類の支援方式がある。まずは日本語非母語話者の母語への翻訳である。例を挙げると、マイクロソフトにより開発されたリアルタイムの自動キャプション機能では、PowerPointで発表する時に、発表者の発話を識別し、目標言語に翻訳して、画面上に表示することができる。有名なSNSであるWeChatやTwitterには、翻訳機能が搭載されており、翻訳ボタンをクリックするだけで、今読んでいる内容を母語に変換することができる。また、対面型会議を前提として、多言語間の討論支援システムPaneLiveが開発されている[1]。このシステムでは、多言語辞書を用いて多言語会議での討論を図解化することによって討論の支援を行っている。

もう一つの支援方式は非母語文の理解支援である。例を挙げると、日本語文の平易化の研究がある。やさしい日本語ニュースの制作支援システムが開発されている[2]。このシステムでは、日本語能力試験の出題基準の3級と4級に合わせて約1600語で、書き換え辞書を作成して、ニュース中に外国人達が読んで理解しづらい部分、漢字、単語、文法などを全部書き換えている。また、重要箇所の高ライトによるコミュニケーションの改善[3]。この研究では、コミュニケーションの過程で、母語話者により、会話中に重要な部分をハイライト処理し、そして、通常の会話パターンと比較している。実験中に、対話の質を定量化するために、四つの指標が設定されている。それぞれはリコールできる共通の部分(RC)、会話の快適さ(GCQ-C1)、会話明瞭度(GCQ-C2)とワークロード(WL)である。実験は英語を共通言語とする。被験者は母語話者28名、平均年齢40.89歳、非母語話者14名(平均TOEIC成績806.25)、平均年齢24.5歳である。実験結果によると、非母語話者は会話内容の理解度が高くなる、ワークロードも少なく、会話の快適さも高くなる代わりに、母語話者のワークロードが増加する。全般的に、この会話パターンは多言語コミュニケーションに役に立つ。したがって、本研究では、実験の部分も、ハイライト処理した日本語文を実験対照として、本研究で提案した処理方式の有効性を検証する。

しかしながら、日本語非母語話者の母語への翻訳の支援アプローチでは、非母語話者が現在のコミュニケーション環境で使用している共通言語を避けるだけであり、非母語話者の言語能力向上に寄与しない。また、日本語の平易化は読みやすさの向上によりある程度の読解の負荷軽減に繋がるものである、文量が削減されないため、文量が多いと依然として読解負荷が高く、リアルタイムのコミュニケーションにおいて即答が難しいである。

本研究では、非母語話者が日本語によるリアルタイムのコミュニケーション環境で、日本語長文を効率的に理解し、即時に応答できるようにすることを目的とする。そこで、本研究では、使用言語を日本語と想定し、非母語話者が日本語長文を理解できるようにするために、自然言語処理の自動要約技術を用いて、日本語長文を短縮して、非母語話者にも理解しやすい形に変換して提示する。

2.2 テキスト要約と文圧縮

本研究で使われた二種類の日本語自動要約技術は、どちらも自然言語処理分野の技術である。アルゴリズムは異なるが、処理の流れはほぼ同じである。自然言語処理には二種類のアプローチがある。

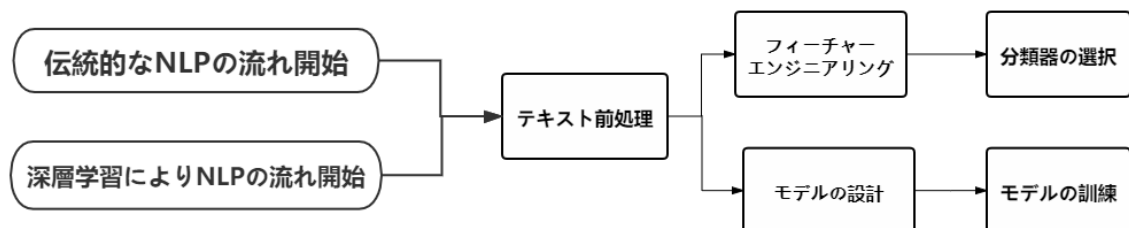


図 1. NLP の流れ

一番目は伝統的な方式である。本研究で採用した要約ツールはこの方式で開発したものである。まずは、用意されたテキストに前処理を行う。例えば、文を単語に分割する、意味がない単語を削除する、単語の品詞情報を分析するなどを前処理として行う。次はフィーチャーエンジニアリングである。特徴抽出とも呼ばれる。各文の長さや単語の出現頻度などを集計するステップである。次のステップは分類器を選択し、目的の結果を算出する。分類器とは、TFIDF や CNN など、文章の重要度を計算するアルゴリズムのことである。二番目は深層学習による自然言語処理である。この方式の処理フローは図 1 に示すように、テキストを

処理し、モデルの設計と訓練を行う。現時点では、多くの言語処理モデルが公開されている。しかしながら、この方法は十分なハードウェアサポートと時間が必要である。結果的にもこの方式がより良いである。

第3章 自動文書要約を用いた日本語対話支援システム

本章では、提案した自動文書要約を用いた日本語対話支援システムの詳細について説明する。

3.1 システム概要

本システムは多様な言語能力を有する非母語話者を対象として、相手が発信した日本語メッセージを要約する、または圧縮することで、文の難易度を下げて、非母語話者の負担軽減を目的とする。

3.1.1 システムワークフロー

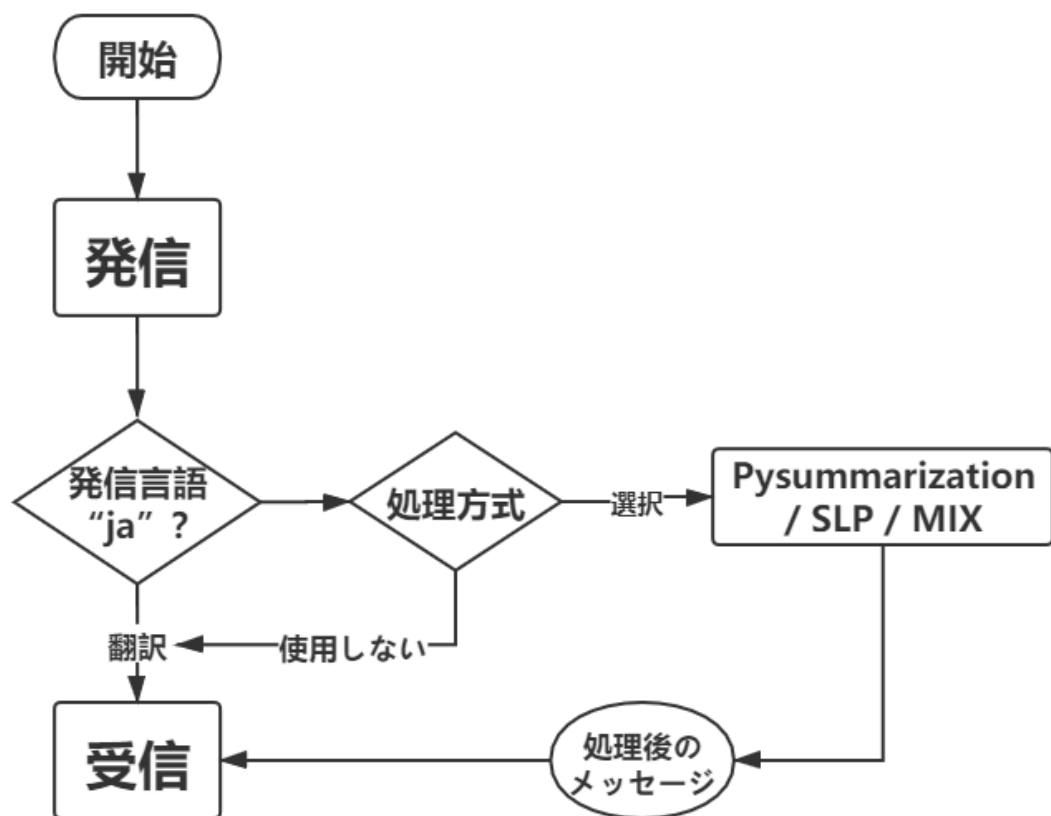


图 2. システムワークフロー

本システムのワークフローは図 2 の通りである。まず、ユーザーが登録する。

登録の際に、システムはユーザーに使用したい言語の選択を求める。ユーザーには、日本語母語話者、日本語勉強中の日本語非母語話者、他言語母語話者がいると想定する。ログイン後、ユーザー同士の間コミュニケーションを始める、メッセージを送信する際に、システムはユーザーが最初に選択した言語により次のアクションを判断する、発信者が日本語選択しない場合は、そのメッセージを受信者側の言語に直接翻訳されて送信する。発信者が日本語選択した場合は、受信者がメッセージの要約結果を見たい場合、システムは受信者が選択した処理方式に従ってメッセージを要約し、その結果を受信者側に送信する。受信者が要約結果を見たくないなら、システムが受信者選択した言語に応じて対応の翻訳結果を送信する、

3.1.2 ユーザーインターフェース

要約ツールの種類は 3 つあるので、要約ツールの種類を自由に選べる機能も必要である。図 3 は要約ツールを選択する時のユーザーインターフェースである。

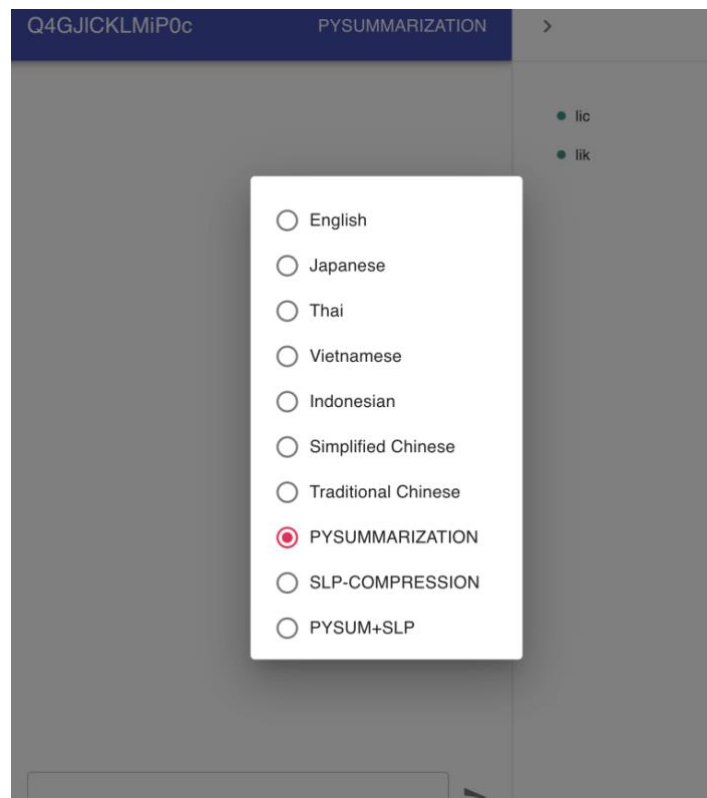


図 3 ユーザーインターフェース

3.2 システム構成

これからは本システムで使われているメッセージの要約方式について紹介する。

3.2.1 pysummarization¹

pysummarization は、自動要約、文書抽出、テキストフィルタリングのための Python3 ライブラリである。本ライブラリは、自然言語処理とニューラルネットワーク言語モデルを用いた自動要約機能が備えている。そして、このライブラリは、テキストクラスタリングによってフィルタリングされた文書やウェブスクレイピングされたテキストの主要なポイントをまとめた要約を作成することができる。また、Accel-brain-base を用いて、LSTM をベースとした Encoder/Decoder を実装し、Sequence-to-Sequence (Seq2Seq) 学習により、要約の精度も向上される。

本システムでは、主に SimilarityFilter 機能を使う。SimilarityFilter 機能は、文章内にある各単文に対し類似性の尺度を用いて計算し、冗長な文章などを短縮することができる。類似性の尺度を用いた計算とは、文章中の同じ意味を表す文や構造がよく似ている文の類似性を要約前に計算し、似ている文を削除することで、より正確な要約結果を得る手法である。本ライブラリでは、類似性を計算するために、Dice 係数、Jaccard 係数、Simpson 係数、TFIDF-COS 係数の4種類の係数が用意されている。本システムでは TFIDF-COS 係数を用いている。

自動要約の具体的な流れは、まずユーザーが入力したテキストを日本語のトークナイザーで文リストを作成する。そして、類似度を計算し、類似度閾値以上の類似度を持つ文を削除する。次にライブラリの自動文書オブジェクトを呼び出し、残った文を要約する。図4は pysummarization の入力文と出力の比較である。受信したメッセージの「..」部分はフィルタリングまたは要約のマークである。

¹ Accel-brain,<https://code.accel-brain.com/Automatic-Summarization/>

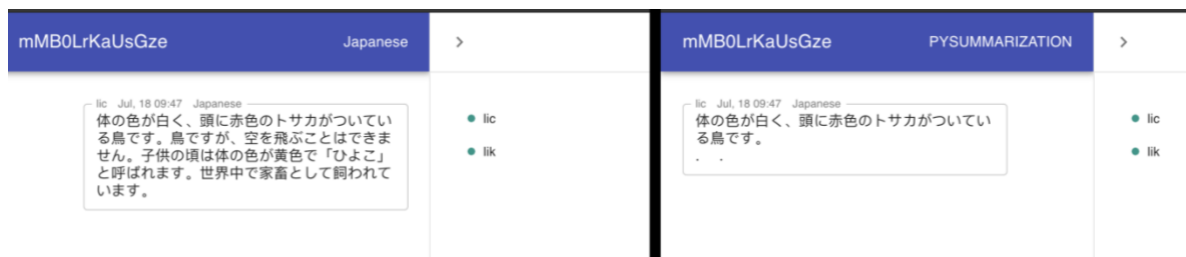


図 4 pysummarization の入力文と要約結果の比較

3.2.1 SLP-sentence-compression[5]

SLP は日本語長文を日本語単文に分割して、得られた各日本語単文を短縮する方式である。具体的な流れは、まず分割して得られた各単文に対して、自然言語処理ライブラリ「spacy」に基づいて、単語分割、品詞情報分析、係り受け解析などの事前処理を行い、特徴量を抽出する。この特徴に基づいて文の短縮を行う。

伝統的な文短縮方式である SLP は、単文中どの単語を保留し、どの単語が削除すべきかを判定する。対象単語を中心に前後 3 つの 6 単語の品詞情報と、係り受け解析で得られた、対象単語の修飾語を中心に前後 3 つの 6 単語の品詞情報を引数として保留可能性算出関数に渡す。例えば、「人の意見に共感や同意しないこと」という文で対象単語が「意見」であれば、図 5 の係り受け解析結果により、判定の必要がある関係は「意見」と「人」「の」「に」「共感」「や」の関係、そして、「意見」の修飾語「同意」と「に」「共感」「や」「し」「ない」「こと」の関係である。この後、訓練したモデルに基づいて、各単語が削除される可能性と削除されない可能性を算出する。ここでは、閾値を 0.2 に設定し、有効性が 0.2 未満の単語は削除する。もう一つのルールは、文中の動詞とルートは判定不要というもので、すべての結果には原文の動詞とルートが含まれる。つまり、図 5 の例では、要約結果には「し」と「こと」が含まれる。

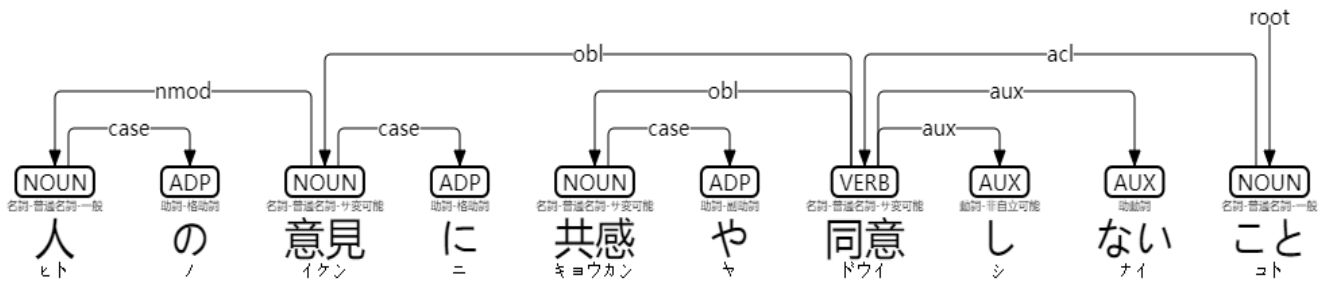


图 5 係り受け解析例

```

groups = {list: 10} [(人, 'del_1'), (人, 'not_del')], [(の, 'not_del')], [(意見, 'del_1'), (意見, 'not_del')], [(に, 'de...
> 00 = {list: 2} [(人, 'del_1'), (人, 'not_del')]
> 01 = {list: 1} [(の, 'not_del')]
> 02 = {list: 2} [(意見, 'del_1'), (意見, 'not_del')]
> 03 = {list: 2} [(に, 'del_1'), (に, 'not_del')]
> 04 = {list: 2} [(共感, 'del_1'), (共感, 'not_del')]
> 05 = {list: 1} [(や, 'not_del')]
> 06 = {list: 2} [(同意, 'del_1'), (同意, 'not_del')]
> 07 = {list: 1} [(し, 'not_del')]
> 08 = {list: 1} [(ない, 'not_del')]
> 09 = {list: 1} [(こと, 'not_del')]
__len__ = (int) 10

```

图 6 有効性算出

有効性が導き出された後、すべての有効性は図 6 のように示す。これらの有効性をもとに、要約結果をまとめる。したがって、SLP では、単文に対して複数の要約結果が出力される。そこで、信頼度を計算できるモデルを事前に用意し、システムが要約結果を信頼値によりランク付けし、信頼値の一番高い要約結果を非母語話者に送信する。

図 7 は SLP のワークフロー図である。

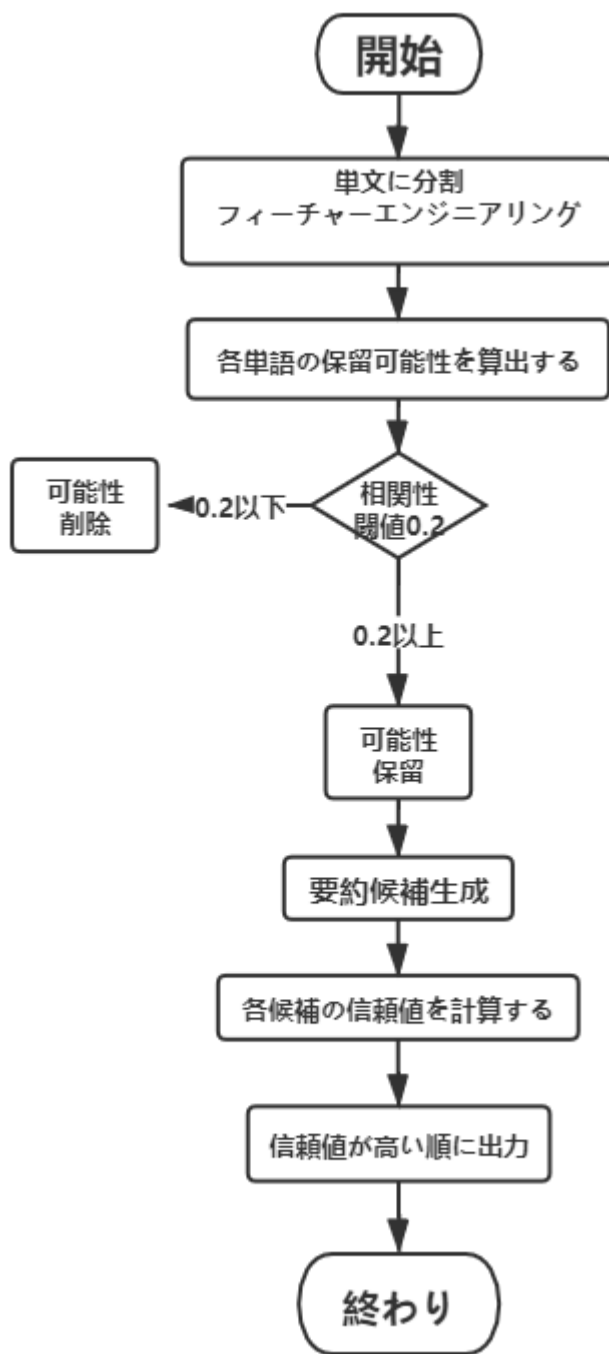


図 7 SLP のワークフロー図

図 8 は SLP の各関数の実行時のシーケンス図である。ja_spiliter() は複数文を単文に分割する関数である。nlp() は単文の特徴量抽出を行う関数である。

nbor() は単文中に単語の隣接する単語を返す. pos() は単語の品詞情報を返す.

3.2.2 psummarization と SLP の統合

psummarization と SLP 統合というのは, ユーザーから送られたメッセージは 2 つの要約ツールで処理される. 使用する順番によって 2 種類の組み合わせがある. 一番目はまず psummarization で処理し, 次に SLP で処理する. 処理の結果により, 要約結果が最も短くなった. 二番目はまず SLP で処理し, 次に psummarization で処理する. SLP では各単文に対して複数の要約結果が出るため, その要約結果が psummarization のノイズとなり, psummarization の処理結果の精度を低下させる. したがって, 前者の統合方式を採用する.

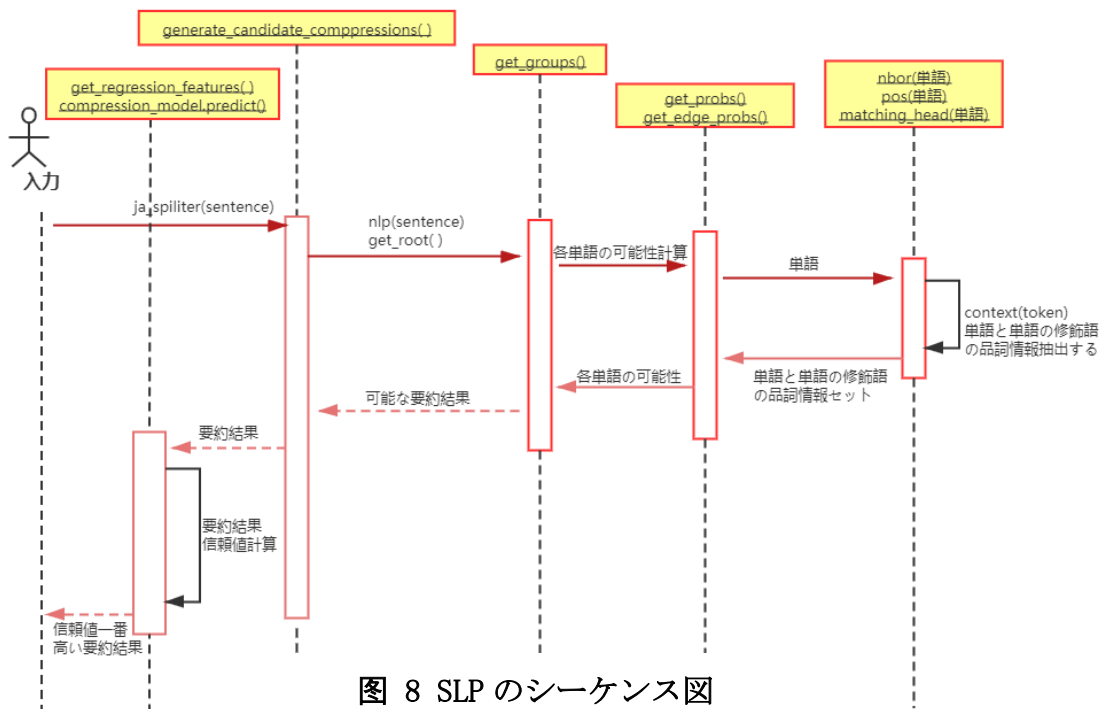


图 8 SLP のシーケンス図

第4章 実験

本章では、自動要約ツールを組み込んだ日本語対話支援システムを用いて行った実験について説明する。

4.1 実験目的

本実験の目的は、自動要約ツールを組み込んだ日本語対話支援システムの有効性を検証することである。そして、日本語能力レベルが異なる非母語話者に対して、異なる自動要約ツールでコミュニケーションをする際に、どのぐらいの違いがあるのかも検証する。

4.2 実験説明

実験ではタブーゲームの形式を採用し、母語話者と非母語話者のコミュニケーションを行う。タブーゲームとは、司会者が参加者にキーワードといくつかの禁止ワードを与え、一名の参加者が禁止ワードを使わずにキーワードを説明し、もう一名の参加者が前者の説明により、キーワードを推測するゲームである。実験の過程では、被験者の日本語能力レベル、キーワード推測の正解数、実験を完成までに要する時間を記録する。

4.2.1 実験フロー

まずは、タブーゲームのキーワードライブラリのウェブサイトで適切なキーワードと禁止ワードの単語セット 20 個を選択する¹。実際の日本語コミュニケーションの会話環境を模擬するために、日本語母語話者にキーワードの説明文を書いてもらう。キーワードの説明文を得た後、これらの説明文に要約処理、ハイライト処理を行い、それぞれは原文、PUSUMMARIZATION, SLP, PUSUMMARIZATION と SLP 統合、ハイライトの 5 種類の説明文を用意する。SLP は複数の要約結果があるため、被験者の負荷にならないように、最良の結果を抽出して、組み合わせで実験用の説明文を作成する。一方、キーワードごとに 5 種類の説明文があるため、対応する実験環境は 5 つある。実験結果の正確性を確保するために、各被験者は 5 種類の実験環境すべてを実験する。したがって、実験問題には、20 個の

¹ <https://www.pinterest.jp/sunnybracco/taboo-games/>

説明文を5組に分け、各組を異なる方式で処理し、表1に示すように実験問題セットを5セット作成する。これで、各実験問題セットに20問あり、その中には5種類の異なる処理結果が含まれる。これにより、キーワードや被験者日本語能力レベルの違いによるの誤差を最小限に抑える。

表 1 実験問題構成

実験問題種類	種類 1	種類 2	種類 3	種類 4	種類 5
「①②③④⑤」 単語の組を表す。 「High1」： ハイライト処理 「mix」： pysummarization とSLP統合処理	①. 原文	①. mix	①. SLP	①. pysum	①. High1
	②. High1	②. 原文	②. mix	②. SLP	②. pysum
	③. pysum	③. high1	③. 原文	③. mix	③. SLP
	④. SLP	④. pysum	④. high1	④. 原文	④. mix
	⑤. mix	⑤. SLP	⑤. pysum	⑤. High1	⑤. 原文

4.2.2 被験者の種類

日本語能力試験 JLPT では、日本語能力を N1 から N5 までの 5 段階に分けており、N1 が最も高いレベルである。日本へ仕事や留学をする前に、ほとんどの日本語非母語話者がこの試験を実験する。提案したシステムがどのレベルの日本語非母語話者に有効なのかを検証するために、被験者を日本語母語話者、日本語レベル N1, N2, N3 合わせて約 30 人用意した。N1 レベルに到達する条件は、10000 単語が分かり、900 時間以上の学習時間が必要である。N2 に到達する条件は、6000 単語が分かり、600 時間以上の学習時間が必要である。N3 に到達する条件は、3000 単語が分かり、450 時間以上の学習時間が必要である。

第5章 評価

本章では，4章の実験データを洗い出し，自動文章要約を用いた日本語対話支援システムの有効性について分析を行う．その後，実験から得られた問題点について考察を行う．

5.1 評価方法

本実験のデータから定量分析を行う．自動要約文章を用いた日本語対話支援システムの有効性を定量分析により評価するため，説明変数として，各日本語能力レベルと処理方式，目的変数として，各日本語能力レベルの被験者が問題完了までに要した平均時間，各日本語能力レベルの被験者の平均正答数，各日本語能力レベルの被験者の解答効率を用いた．

5.2 実験結果

表2は各日本語能力レベルの被験者が各処理方式での平均解答時間を示す，表3は各日本語能力レベルの被験者が各処理方式での平均正答数を示す．統合処理はPYSUMMARIZATIONとSLPの統合処理である．

表2 平均解答時間

処理方式	平均解答時間 単位: 秒	母語話者	日本語能力 N1	日本語能力 N2	日本語能力 N3
原文		122.7	225.4	239.9	673.5
PYSUMMARIZATION		216.1	203.1	245.1	427.0
SLP		232.6	245.9	241.8	352.7
統合処理		201.9	208.2	307.4	720.8
ハイライト処理		182.7	235.4	291.6	802.0

表3 平均正答数

処理方式	平均正答数	母語話者	日本語能力 N1	日本語能力 N2	日本語能力 N3
原文		4	3.5	2.4	3
PYSUMMARIZATION		3.5	3.1	3.0	2.25
SLP		3.5	3.4	3.2	3.0
統合処理		3.1	3.1	2.3	2.8
ハイライト処理		3.8	3.5	3.1	1.7

5.3 分析

図9と図11によると、日本語母語話者にとって、自動文書要約を用いた日本語対話支援システムはネガティブな効果がある、解答時間と正答率が下げる、文書要約ツールは説明文の一貫性を破壊し、文の可読性を下げる。

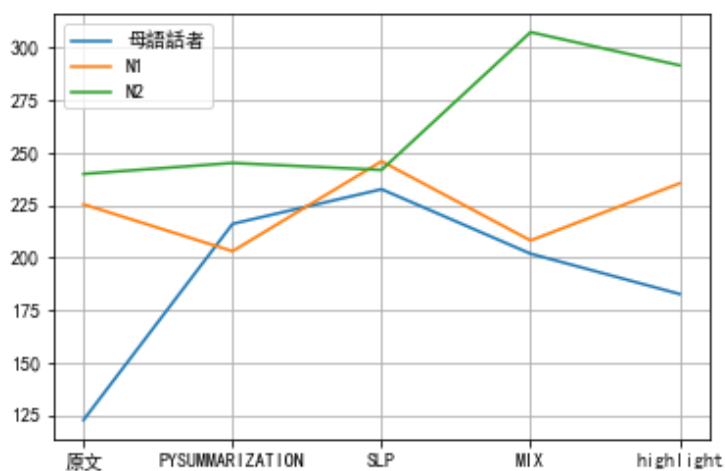


图 9 平均解答時間

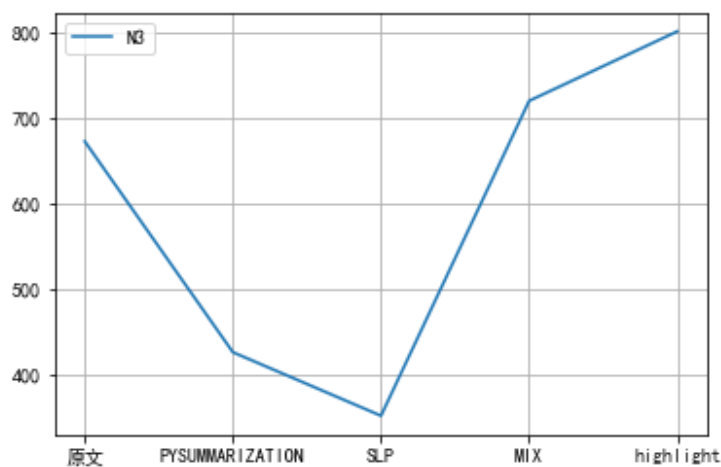


图 10 N3 レベル被験者平均解答時間

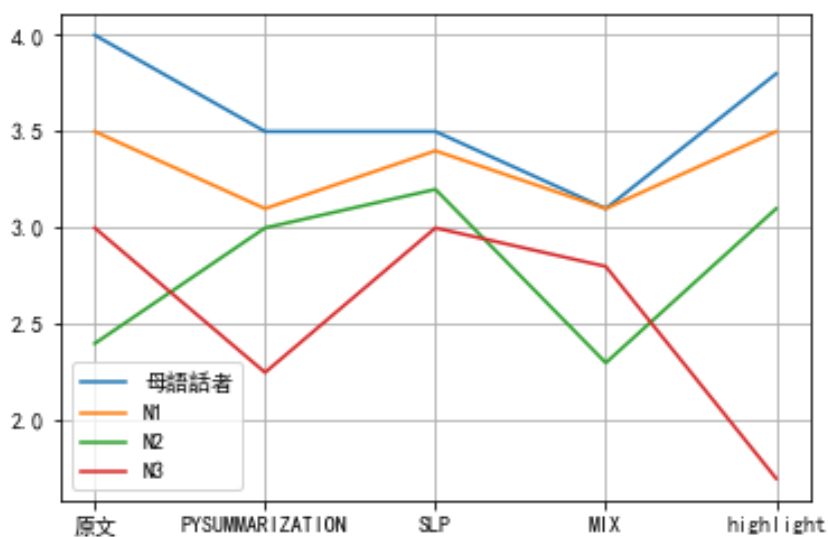


图 11 平均正答数

N1 レベルの被験者のデータから見ると、解答時間と正解数の数値変化の傾向は同じである。SLP 処理や統合処理の実験環境での解答時間や正答数は、日本語母語話者の値に近い。つまり、短い文を理解することに関しては、N1 レベルと母語話者の間にはほとんど差がない。文書要約ツールは説明文の一貫性を破壊し、文の可読性を下げる。母語話者を混乱させる可能性もある。一方、

原文とハイライトの実験環境での回答時間は、母語話者とは大きな差がある。つまり、母語話者と比べて、N1 レベルの人は、長文の理解に時間を要する。

N2 レベルの被験者は PYSUMMARIZATION と SLP 処理の実験環境では、解答時間がほぼ変わってないが、正答数が上がる。つまり、要約により被験者が説明文の内容をより良く理解でき、要約文がポジティブな効果を与えた。また、二つの中では SLP 処理の方が良い。

N3 レベルの被験者は PYSUMMARIZATION と SLP 処理の実験環境で、解答時間が大幅に減る。原文環境で正答数が N2 レベルの被験者より多かったのは、N3 レベルの被験者は翻訳ツールを使って質問を翻訳していた可能性がある。

図 12, 図 13, 図 14 は被験者解答効率の比較である。解答効率は以下の式で定義する。

$$\text{効率} = \frac{\text{平均正答数}}{\text{平均解答時間} * 60}$$

効率は被験者の一分間の正答数を示す。要約処理は日本語母語話者、N1 レベルの被験者の効率を下げた。これは彼らにとって、大切な情報を漏れた可能性がある。一方、N2, N3 レベルの被験者の効率は上昇し、本システムは N2, N3 レベルの人に有効であることを示す。PYSUMMARIZATION と SLP の統合処理は短縮しすぎたために、どのレベルでも実験結果は悪い。ハイライト処理の実験結果では期待通りの結果が得られなかった。被験者からのコメントによると、テキストのハイライト箇所は邪魔になることがあった。また、ハイライトに依存してしまい、思考力を低下させ、集中できないというコメントもあった。さらに、間違った所がハイライトされて、解答効率が悪くなる場合もある。

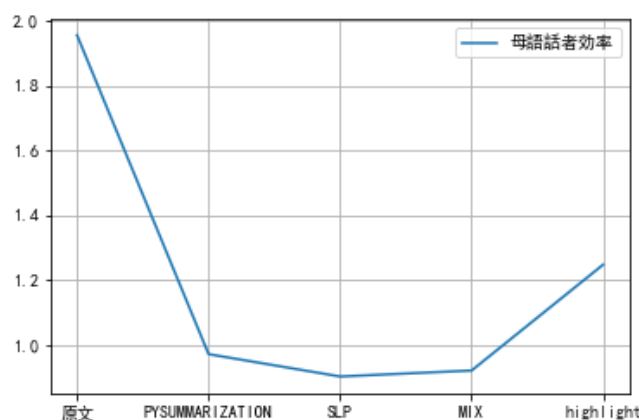


图 12 日本語母語話者解答効率変化図

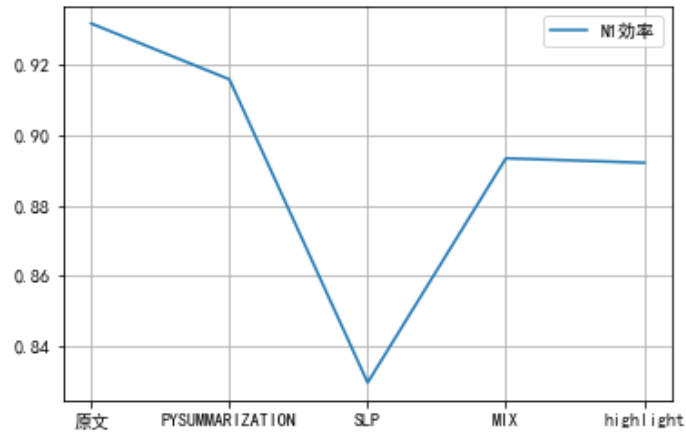


图 13 N1 レベル被験者解答効率変化図

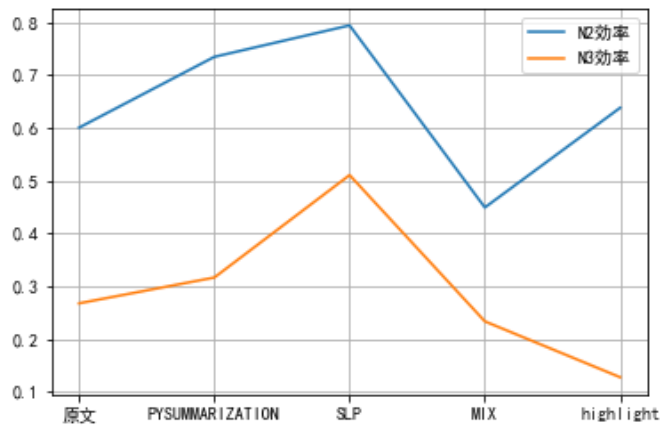


图 14 N2 と N3 レベルの被験者解答効率変化図

5.4 考察

5.4.1 有効性の検証

自動文書要約を用いた日本語対話支援システムは対話支援システムとして、N1 レベルの人にはあまり効果がないが、N1 レベルであっても、日本語長文を理解する能力は日本語母語話者とは大きな差があることを発見した。一方で、提案システムは N2 と N3 レベルの人の対話効率を上げることが可能であり、解答効率において、N2 レベルの被験者が 0.2 上昇した、0.8 になった、N1 レベルの被

験者原文環境での効率 α は 0.93 である, そして, N1 と N2 の間に, 300 時間の学習時間差がある, SLP を使うことにより, 約 200 時間の程度の日本語能力を上昇した, 自動文書要約を用いた日本語対話支援システムの有効性を示した. さらに, 要約方式では, 複数文から重要文をフィルタリングする要約方式よりも, 各単文を短縮する要約方式の方が解答効率を向上させた.

実験結果の有意差を検定するために, T 検定を行う. 有意水準 α を 0.05 とする, 表 4 と表 5 は実験データの T 検定の結果である. 表 4 と表 5 によると, 今回のデータ量では有意に差があることまでは検証できなかった.

5.4.2 問題点

本実験では, ハイライト処理は対話効率に寄与しなかった. ハイライトにより多言語コミュニケーションの改善の関連研究では, 被験者の平均 TOEIC 点数は 990 点中 806.25 点であり, 日本語 N1 は 180 点満点のため, 換算すれば, N1 は 150 点の被験者が必要であった. しかしながら, 本実験では, ほとんどの被験者がこのレベルに達しておらず, 被験者の日本語レベルが足りなかったと考えられる.

5.4.3 これからの課題

本実験の結果により, システムはまだ不完全な部分がある. 特に改良や実験が必要な部分は以下の 3 点である.

- 1) PYSUMMARIZATION のフィルタリング閾値調整.
- 2) 深層学習ベースの自動文書要約ツールの利用
- 3) 発信者による自動ハイライト処理機能の実装.

本実験では, PYSUMMARIZATION 処理と統合処理は良い結果が得られなかった. 最大の原因はあまりにも多くの内容を省略することで, 重要な情報を失ってしまうことである. PYSUMMARIZATION のフィルタリング閾値を上げて, フィルタリングする内容を削減することで, PYSUMMARIZATION 処理と統合処理はより良い結果を生み出す可能性がある.

本システムで用いている要約ツールは伝統的な手法に基づいて開発されている. 一方で, 近年, 深層学習ベースの要約ツールによる精度向上が見られる. したがって, 深層学習ベースの要約ツールがあれば, 要約結果はより正確になり, 実験の結果もより良くなる可能性がある. さらに, 統合処理方式も, 試せる組み合わせが多様になる.

発信者による自動ハイライト処理機能は現時点では実装していない. 実験結

果によると、このシステムは N2 と N3 レベルの人に有用である。この機能があれば、日本語レベルが高いユーザーにも有用になることが期待できる。

表 4 N2 レベル実験データの T 検定結果

自由度8, $t_{0.05}(8) = 1.860$	時間	正答数	効率
統計量t (原文と pysummarization)	0.039	0.958	0.473
統計量t (原文と SLP)	-0.025	1.214	0.794

表 5 N3 レベル実験データの T 検定結果

自由度2, $t_{0.05}(2) = 2,920$	時間	正答数	効率
統計量t (原文と pysummarization)	-0.412	-1.015	-0.1
統計量t (原文と SLP)	1.138	0.999	1.044

第6章 おわりに

日本語会話支援のために、自動文書要約を用いた日本語対話支援システムを提案した。自動文書要約を用いた日本語対話支援システムの有効性を検証するため、タブーゲーム形式の実験を行った。その結果は自動文書要約を用いた日本語対話支援システムは N2 と N3 レベルの人に対して、対話支援システムとして有効性があることがわかった。

本研究の貢献は以下の2点である。

要約ツールとチャットツールの結合

自動文書要約を用いた日本語対話支援システムを提案した。文書要約ツールとチャットツールの結合により、日本語非母語話者が日本語長文の意味を理解できるようにする。日本語非母語話者の日本語コミュニケーションの効率を上げる。

システムの有効性の検証

本研究では、タブーゲーム形式の日本語対話模擬実験を行った。実験データにより、自動文書要約を用いた日本語対話支援システムは N2 と N3 レベルの人のコミュニケーション効率を上げることがわかった。N2 レベルの人にとって、約200時間の学習の程度を上昇した。また、複数文から重要文をフィルタリングする要約方式よりも、各単文を短縮する要約方式の方がより解答効率を向上させた。今後は、さらにコミュニケーション効率を向上させるために、深層学習ベースの要約ツールを本システムに組み込む予定である。

謝辞

本研究を行うにあたり，ご指導していただいた村上陽平准教授に深く感謝を申し上げます。また，お世話になっている社会知能研究室の皆様に感謝を申し上げます。

参考文献

- [1] 福島拓, 吉野孝, 喜多千草: 対面討論における非母語話者支援システム Panelive の開発, ワークショップ 2008 (GN Workshop 2008), Vol. 2008, pp. 37-42 (2008).
- [2] 田中英輝, 熊野正, 後藤功雄, & 美野秀弥. (2018). やさしい日本語ニュースの制作支援システム. *自然言語処理*, 25(1), 81-117.
- [3] Pan, M. H., Yamashita, N., & Wang, H. C. (2017, February). Task Rebalancing: Improving Multilingual Communication with Native Speakers-Generated Highlights on Automated Transcripts. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 310-321).
- [4] 奥野陽, グラム・ニュービグ, 萩原正人. 自然言語処理の基本と技術[M]. 翔泳社, 2016.
- [5] Mianuddin, marisaaquilina, <https://github.com/marisaaquilina/slp-sentence-compression>

付録：卒業論文の付録の付け方

A.1 付実験問題集

1 キーワード	鶏	砂漠	船	銀行
禁止ワード1	卵	沙	海	建物
禁止ワード2	食べ物	サハラ砂漠	旅行	お金
禁止ワード3	KFC	ドライ	航海	ATM

2 キーワード	バスケットボール	汚染	奇術(magic)	馬
禁止ワード1	NBA	空気	偽物	動物
禁止ワード2	Micheal Jordan	都市	素晴らしい	速い
禁止ワード3	球	工業		乗る

3 キーワード	リンゴ	学ぶ	退屈	虹
禁止ワード1	日常	勉強	授業	赤
禁止ワード2	安い	学校	寝る	綺麗
禁止ワード3		つまらない	興奮	雨

4 キーワード	財布	恋愛	寿司	圧倒的
禁止ワード1	動物	友達	握る	強い
禁止ワード2	速い	関係	日本	弱い
禁止ワード3	乗る		美味しい	競争

5 キーワード	うるさい	歴史的	蛇	反対
禁止ワード 1	大声	歴史	動物	賛成
禁止ワード2	雑音	古い	毒	会議
禁止ワード3	お隣さん	古跡		議論

1 キーワード:鶏

体の色が白く、頭に赤色のトサカがついている鳥です。鳥ですが、空を飛ぶことはできません。子供の頃は体の色が黄色で「ひよこ」と呼ばれます。世界中で家畜として飼われています。

Highlight:

体の色が白く、頭に赤色のトサカがついている鳥です。鳥ですが、空を飛ぶことはできません。子供の頃は体の色が黄色で「ひよこ」と呼ばれます。世界中で家畜として飼われています。

PYSUMMARIZATION:

体の色が白く、頭に赤色のトサカがついている鳥です。

. . .

SLP-COMPRESSION:

番号 短縮結果

体の色が白く、トサカがついている鳥。

@

鳥ですが、空を飛ぶことはできません。

@

子供の頃体黄色ひよこ呼ばれます。

@

家畜として飼われています。

@

PYSUM+SLP MIX:

番号 短縮結果

体の色が白く、トサカがついている鳥。

@

2 キーワード:砂漠

降雨が極端に少なく、主にアフリカ大陸に存在します。見晴らしがとてもいいです。所々に「オアシス」と呼ばれる、水のある場所が存在します。とても暑くて、こまめな水分補給をしなければいけません。

Highlight:

降雨が極端に少なく、主にアフリカ大陸に存在します。見晴らしがとてもいいです。所々に「オアシス」と呼ばれる、水のある場所が存在します。とても暑くて、こまめな水分補給をしなければいけません。

PYSUMMARIZATION:

降雨が極端に少なく、主にアフリカ大陸に存在します。

- . . 見晴らしがとてもいいです。
- . . 所々に「オアシス」と呼ばれる、水のある場所が存在します。
- . . とても暑くて、こまめな水分補給をしなければいけません。
- . .

SLP-COMPRESSION:

番号 短縮結果

降雨 が 少なく アフリカ 大陸 に 存在 し ます。

@

見晴らし いい です。

@

所々 オアシス と 呼ばれる 、 場所 が 存在 し ます。

@

こまめ な 水分 補給 を し なけれ いけ ませ ん。

@

PYSUM+SLP MIX:

番号 短縮結果

降雨 が 少なく アフリカ 大陸 に 存在 し ます。

@

見晴らし いい です。

@

所々 オアシス と 呼ばれる 、 場所 が 存在 し ます。

@

こまめ な 水分 補給 を し なけれ いけ ませ ん。

@

3 キーワード : 船

水の上に浮くもので、小さいものは木で、大きいものは鉄でできていることが多い。重いものを大量に輸送する時によく使われます。動力源はオールで漕いだり、エンジンを使ったりします。

Highlight:

水の上に浮くもので、小さいものは木で、大きいものは鉄でできてることが多い。
重いものを大量に輸送する時によく使われます。動力源はオールで漕いだり、エンジンを使ったりします。

PYSUMMARIZATION:

水の上に浮くもので、小さいものは木で、大きいものは鉄でできてることが多い。
． ． 動力源はオールで漕いだり、エンジンを使ったりします。

． ．

SLP-COMPRESSION:

番号 短縮結果

水の上 浮くもので、小さいものは木ものできてること が。

@

大量に輸送する時によく使われます。

@

動力源はオールで漕い、エンジンを使っ します。

@

PYSUM+SLP MIX:

番号 短縮結果

水の上 浮くもので、小さいものは木ものできてること が。

@

動力源はオールで漕い、エンジンを使っ します。

@

4 キーワード:銀行

NG: 建築物, 钱 , ATM

世界のどこにでもあり、国や人々にとって非常に重要な場所であり、あなたの財

産を守るのにも役立ちます。その安全性が非常に高く、人生では、誰もがその場所に行く必要があります。毎日多くの人々がビジネスをするためにその場所に行きたいと思っています。

Highlight:

世界のどこにでもあり、国や人々にとって非常に重要な場所であり、あなたの財産を守るのにも役立ちます。その安全性が非常に高く、人生では、誰もがその場所に行く必要があります。毎日多くの人々がビジネスをするためにその場所に行きたいと思っています。

PYSUMMARIZATION:

世界のどこにでもあり、国や人々にとって非常に重要な場所であり、あなたの財産を守るのにも役立ちます。

. その安全性が非常に高く、人生では、誰もがその場所に行く必要があります。

.

SLP-COMPRESSION:

番号 短縮結果

世界のどこにでもあり重要な場所であり財産を守るのにも役立ちます。

@

行く必要があります。

@

毎日多くの人々がビジネスをする行きたい思っています。

@

PYSUM+SLP MIX:

番号 短縮結果

世界のどこにでもあり重要な場所であり財産を守るのも役立ちます。

@

行く必要があります。

@

5 キーワード:バスケットボール NG; NBA, Micheal Jordan, 球

アメリカ人発明された、5人対5人の2チームが、一つのボールを手で扱う運動である。運動量が多く、レベルの高いプレイをするには身体的能力が求められることなどが挙げられる。

Highlight:

アメリカ人発明された、5人対5人の2チームが、一つのボールを手で扱う運動である。運動量が多く、レベルの高いプレイをするには身体的能力が求められることなどが挙げられる。

PYSUMMARIZATION:

アメリカ人発明された、5人対5人の2チームが、一つのボールを手で扱う運動である。

. . .

SLP-COMPRESSION:

番号 短縮結果

アメリカ発明された、5人対5人の2チーム、一つのボールを扱う運動ある。

@

レベル 高い プレイ を する に は 身 体 的 能 力 が 求 め ら れ る こ と 等
が 挙 げ ら れ る。

@

PYSUM+SLP MIX:

番号 短縮結果

2 アメリカ 発 明 さ れ た , 5 人 対 5 人 の 2 チーム、一つの ボール
を 扱 う 運 動 ある。

@

6 キーワード:汚染

汚れに染まることです。混入物が自然環境に入り、良くない変化を引き起こすこと
と である。化学物質または雑音や熱、光といったエネルギーの形をとりうる。現
在、人々はこの問題に注意を払っている、それが起こらないように取り組んでい
ます。

Highlight:

汚れに染まることです。混入物が自然環境に入り、良くない変化を引き起こすこ
と である。化学物質または雑音や熱、光といったエネルギーの形をとりうる。現
在、人々はこの問題に注意を払っている、それが起こらないように取り組んでい
ます。

PYSUMMARIZATION:

汚れに染まることです。

- ・ 混入物が自然環境に入り、良くない変化を引き起こすことである。
- ・ 現在、人々はこの問題に注意を払っている、それが起こらないように取り
組んでいます。

・

SLP-COMPRESSION:

番号 短縮結果

汚れに染まること。

@

混入物が自然環境に入り、良くない変化を起こすことである。

@

雑音や熱エネルギーの形をとりうる。

@

現在、人々この問題に注意を払っている、それが起こらないように組んでいます

@

PYSUM+SLP MIX:

番号 短縮結果

汚れに染まること。

@

混入物が自然環境に入り、良くない変化をことである。

@

現在、人々この問題に注意を払っている、それが起こらないように組んでいます。

@

7キーワード:奇術 魔术 NG: 偽物 素晴らしい

人間の錯覚や思い込みを利用し、実際には合理的な原理を用いてあたかも「実現不可能なこと」が起きているかのように見せかける芸能。通常、観客に見せることを前提としてそのための発展を遂げてきたものをいう。

Highlight:

人間の錯覚や思い込みを利用し、実際には合理的な原理を用いてあたかも「実現不可能なこと」が起きているかのように見せかける芸能。通常、観客に見せることを前提としてそのための発展を遂げてきたものをいう。

PYSUMMARIZATION:

人間の錯覚や思い込みを利用し、実際には合理的な原理を用いて。

- . . . あたかも「実現不可能なこと」が起きているかのように見せかける芸能。
- . . . 通常、観客に見せることを前提としてそのための発展を遂げてきたものをいう。
- . . .

SLP-COMPRESSION:

番号 短縮結果

人間の 錯覚 や 思い込み を 利用 し 、 合理的 な 原理 を 用い。

@

実現 不 可能 な こと 起 き て い る の よう に 見 せ 芸 能。

@

通常 観 客 に 見 せ る こ と を 前 提 と し そ の た め の 発 展 を 遂 げ き た も の を い う。

@

PYSUM+SLP MIX:

番号 短縮結果

人間の 錯覚 や 思い込み を 利用 し 、 合理的 な 原理 を 用い。

@

不可能なことを起きるかのよう見せ芸能。

@

通常観客に見せることを前提としそのための発展を遂げきたものをいう。

@

8 キーワード:馬

NG: 騎, 動物, 速い

首と頭が長く、長い四肢をもつ。角はない。よく発達した蹄をもち、硬い土の上を走ることが出来る。尾と、頭から首の上部にかけての鬣だけは長い、全身の毛は短い。非常に賢くてカッコいいです、道も方向もよく知っていて。車がない時代では、人の乗り物として、数千年の歴史があります、古代の戦争にも登場しました。

Highlight:

首と頭が長く、長い四肢をもつ。角はない。よく発達した蹄をもち、硬い土の上を走ることが出来る。尾と、頭から首の上部にかけての鬣だけは長いが、全身の毛は短い。非常に賢くてカッコいいです、道も方向もよく知っていて。車がない時代では、人の乗り物として、数千年の歴史があります、古代の戦争にも登場しました。

PYSUMMARIZATION:

首と頭が長く、長い四肢をもつ。

・ 非常に賢くてカッコいいです、道も方向もよく知っていて。

・

SLP-COMPRESSION:

番号 短縮結果

頭が長く長い四肢をもつ。

@

0 角はない。

@

蹄をもち土の上を走ることが出来る。

@

上部にかけの鬣だけは長いが毛短い。

@

賢くてカッコいいです、方向も知っています。

@

車がない時代では、乗り物として歴史があります、古代の戦争も登場しました。

@

PYSUM+SLP MIX:

番号 短縮結果

頭が長く長い四肢をもつ。

@

賢くカッコいいです道も方向もよく知っています。

@

9 キーワード:リンゴ

木の実です。甘くて美味しいです。秋から冬にかけての季節が旬です。外皮の色は赤や黄緑または黄色をしている。果肉は淡黄色から白色の品種が多い。外皮近くなど果肉が赤からピンク色になる赤肉系の品種もある。

Highlight:

木の実です。甘くて美味しいです。秋から冬にかけての季節が旬です。外皮の色は赤や黄緑または黄色をしている。果肉は淡黄色から白色の品種が多い。外皮近くなど果肉が赤からピンク色になる赤肉系の品種もある。

PYSUMMARIZATION:

木の実です。

- . . . 甘くて美味しいです。
- . . . 果肉は淡黄色から白色の品種が多い。
- . . .

SLP-COMPRESSION:

番号 短縮結果

木の実。

@

甘くて美味しいです。

@

秋から冬にかけての季節が旬です。

@

色は赤や黄緑色している。

@

果肉は淡黄色から白色の品種多い。

@

果肉が赤からなる赤肉系の品種もある。

@

PYSUM+SLP MIX:

番号 短縮結果

木の実。

@

甘くて美味しいです。

@

果肉は淡黄色から白色の品種多い。

@

10 キーワード：学ぶ

教えを受けて知識やスキルを新しく習得すること、教えてもらって覚えることです。主に、学生が日々行なっています。既に習得した知識を修正することを指すこともあります。

Highlight:

教えを受けて知識やスキルを新しく習得すること、教えてもらって覚えることです。主に、学生が日々行なっています。既に習得した知識を修正することを指すこともあります。

PYSUMMARIZATION:

教えを受けて知識やスキルを新しく習得すること、教えてもらって覚えることです。

- ． ． 主に、学生が日々行なっています。
- ． ． 既に習得した知識を修正することを指すこともあります。
- ． ．

SLP-COMPRESSION:

番号 短縮結果

教えを受けスキル習得すること、教えてもらって覚えること。

@

主に、学生が行なっています。

@

習得した知識を修正することを指すこともあります。

@

PYSUM+SLP MIX:

番号 短縮結果

教えを受けて習得すること、教えもらって覚えることです。

@

主に、学生が行なっています。

@

習得した知識を修正することを指すこともあります。

@

11 キーワード:退屈

なすべきことがなくて時間をもてあましその状況に嫌気がさしている様。もしくは実行中の事柄について関心を失い飽きている様、及びその感情である。教科書をただ棒読みするだけの先生の授業や、会社での単調な作業はひどく苦痛であるの気持ち。

Highlight:

なすべきことがなくて時間をもてあましその状況に嫌気がさしている様。もしくは実行中の事柄について関心を失い飽きている様、及びその感情である。教科書をただ棒読みするだけの先生の授業や、会社での単調な作業はひどく苦痛であるの気持ち。

PYSUMMARIZATION:

なすべきことがなくて時間をもてあましその状況に嫌気がさしている様。

．．．教科書をただ棒読みするだけの先生の授業や、会社での単調な作業はひどく苦痛であるの気持ち。

．．．

SLP-COMPRESSION:

番号 短縮結果

なす べき こと が なく て あまし 嫌気 さし いる 様。

@

もしくは 事柄 に て 関心 を 失い 飽き ている 様 、 その 感情 である。

@

単調 な 作業 は ひどく 苦痛 ある の 気持ち。

@

PYSUM+SLP MIX:

番号 短縮結果

なす べき こと が なく 時間 を あまし 嫌気 さし ている 様。

@

単調 な 作業 ひどく 苦痛 ある の 気持ち。

@

12 キーワード: 彩虹 にじ NG 赤 綺麗 雨

大気中に浮遊する水滴の中を光が通過する際に、分散することで特徴的な模様が見られる大気光学現象である。観察者から見れば地平線の下にある、円弧状の光の帯であり、帯の中には様々な色の光の束が並んでいるように見える。

Highlight:

大気中に浮遊する水滴の中を光が通過する際に、分散することで特徴的な模様が見られる大気光学現象である。観察者から見れば地平線の下にある、円弧状の光の帯であり、帯の中には様々な色の光の束が並んでいるように見える。

PYSUMMARIZATION:

大気中に浮遊する水滴の中を光が通過する際に、分散することで特徴的な模様が見られる大気光学現象である。

. . .

SLP-COMPRESSION:

番号 短縮結果

水滴の中を光が通過する分散することで特徴的な模様が見られる光学現象ある。

@

観察者から見れ線の下にある、光の帯あり、帯の中は束が並んでいるように見える。

@

PYSUM+SLP MIX:

番号 短縮結果

水滴の中を光が通過する分散することで特徴的な模様が見られる光学現象ある。

@

13 キーワード:財布

これらは、一般に「貴重品」と呼ばれる価値の高い物品を収納し携帯の便が良いように工夫されたもので。特に携帯性に主眼がおかれているが、その収納機能も

携帯に際して邪魔にならない形で。デザインによって「男性用」あるいは「女性用」として製品化されているものもある。

Highlight:

これらは、一般に「貴重品」と呼ばれる価値の高い物品を収納し携帯の便が良いように工夫されたもので。特に携帯性に主眼がおかれているが、その収納機能も携帯に際して邪魔にならない形で。デザインによって「男性用」あるいは「女性用」として製品化されているものもある。

PYSUMMARIZATION:

これらは一般に「貴重品」と呼ばれる価値の高い物品を収納し携帯の便が良いように工夫されたもので。

．．．デザインによって「男性用」あるいは「女性用」として製品化されているものもある。

．．．

SLP-COMPRESSION:

番号 短縮結果

らは一般に貴重品呼ばれる物品を収納し便が良いように工夫されたもの。

@

携帯性主眼おかれいるその機能携帯に際して邪魔にならない形で。

@

デザインによって「男性用」「女性用」化されいるものもある。

@

PYSUM+SLP MIX:

番号 短縮結果

らは一般に貴重品と呼ばれる物品を収納し便が良いように工夫されたもの。

@

デザインによって「男性用」「女性用」化されているものもある。

@

14 キーワード:恋愛

NG:心, 朋友, 関係

男女における, このことに対して表現が違います。私たち全員が経験したことです, 特に若い頃に, 私たちにとっても甘くて幸せを感じさせます。人を好きになって, 会いたい, いつまでもそばにいたいと思う, 満たされない気持ちです。

Highlight:

男女における, このことに対して表現が違います。私たち全員が経験したことです, 特に若い頃に, 私たちにとても甘くて幸せを感じさせます。人を好きになって, 会いたい, いつまでもそばにいたいと思う, 満たされない気持ちです。

PYSUMMARIZATION:

男女における, このことに対して表現が違います。

. . . 私たち全員が経験したことです, 特に若い頃に, 私たちにとっても甘くて幸せを感じさせます。

. . . そばにいたいと思う, 満たされない気持ちです。

. . .

SLP-COMPRESSION:

番号 短縮結果

男女におけるこのことに対して表現が違います。

@

幸せを感じさせます。

@

そばにいたいと思う、満たされない気持ちです。

@

PYSUM+SLP MIX:

番号 短縮結果

男女におけるこのことに対して表現が違います。

@

幸せを感じさせます。

@

そばにいたいと思う、満たされない気持ちです。

@

15 キーワード:寿司 NG: 握る 日本 美味しい

1 千年以上の歴史があり、一般に米飯などと主に魚介類を組み合わせた食べ物です。一般的に長方形または円形です。作り方も、食べる方もいくつかあります、アジアのある国でとても有名です。

Highlight:

1 千年以上の歴史があり、一般に**米飯**などと主に**魚介類を組み合わせた食べ物**です。一般的に**長方形または円形**です。作り方も、食べる方もいくつかあります、**アジア**のある国でとても有名です。

PYSUMMARIZATION:

1 千年以上の歴史があり、一般に米飯などと主に魚介類を組み合わせた食べ物です。

SLP-COMPRESSION:

番号 短縮結果

1 千年以上の歴史あり 米飯などと魚介類を組み合わせた食べ物。

@

一般に長方形または円形です。

@

方も、方もいくつかあります、アジアのある国でとても有名です。

@

PYSUM+SLP MIX:

番号 短縮結果

1 千年以上の歴史あり 米飯などと魚介類を組み合わせた食べ物。

@

16 キーワード:圧倒的

他のものより極端に能力などに差がある様子。比較にならないくらい、段違いの様子。程度が甚だしく、相手を圧倒するほどであるさま。力量や規模などが非常に多く、相手にならない、歯が立たないなどと思われるような様子。

Highlight:

他のものより極端に能力などに差がある様子。比較にならないくらい、段違いの様子。程度が甚だしく、相手を圧倒するほどであるさま。力量や規模などが非常に多く、相手にならない、歯が立たないなどと思われるような様子。

PYSUMMARIZATION:

他のものより極端に能力などに差がある様子。

・ 程度が甚だしく、相手を圧倒するほどであるさま。

・

SLP-COMPRESSION:

番号 短縮結果

極端に能力などに差ある様子。

@

比較にならない段違いの様子。

@

程度が甚だしく、相手を圧倒するほどであるさま。

@

力量や規模が非常に多く相手にならない、歯が立たないと思われるような様子

@

PYSUM+SLP MIX:

番号 短縮結果

極端に能力などに差ある様子。

@

程度が甚だしく、相手を圧倒するほどであるさま。

@

17 キーワード:うるさい

形容詞です。物音が大きすぎて耳障りである様子。大きな音を出して周りに不快

感を与えている状態。また、扱いに手間がかかるため面倒に感じられる状態。ポジティブな意味として、大変優れている状態を表すこともある

Highlight:

形容詞です。物音が大きすぎて耳障りである様子。大きな音を出して周りに不快感を与えている状態。また、扱いに手間がかかるため面倒に感じられる状態。ポジティブな意味として、大変優れている状態を表すこともある

PYSUMMARIZATION:

形容詞です。

- . . 物音が大きすぎて耳障りである様子。
- . . 大きな音を出して周りに不快感を与えている状態。
- . . ポジティブな意味として、大変優れている状態を表すこともある。
- . .

SLP-COMPRESSION:

番号 短縮結果

形容 詞 です。

@

物音が大きすぎて耳障りである様子。

@

音を出し不快感与えている状態。

@

扱い手間かかるため面倒に感じられる状態。

@

ポジティブな意味として大変優れている状態を表すこともある。

@

PYSUM+SLP MIX:

番号 短縮結果

形容詞です。

@

物音が大きすぎて耳障りである様子。

@

音を出し不快感与えている状態。

@

ポジティブな意味として大変優れている状態を表すこともある。

@

18 キーワード:歴史的

遠い過去から続いているさま。また、すでに過去のものであるさま。何かしらの事物が時間的に変遷したありさま、あるいはそれに関する文書や記録のことをいう。主に国家や文明など人間の社会を対象とする。

Highlight:

遠い過去から続いているさま。また、すでに過去のものであるさま。何かしらの事物が時間的に変遷したありさま、あるいはそれに関する文書や記録のことをいう。主に国家や文明など人間の社会を対象とする。

PYSUMMARIZATION:

遠い過去から続いているさま。

- ・ . また、すでに過去のものであるさま。
- ・ . 主に国家や文明など人間の社会を対象とする。

・ .

SLP-COMPRESSION:

番号 短縮結果

遠い過去から続いているさま。

@

すでに過去のものであるさま。

@

事物が時間的に変遷したありさまいう。

@

主に国家や文明など人間の社会対象とする。

@

PYSUM+SLP MIX:

番号 短縮結果

遠い過去から続いているさま。

@

すでに過去のものであるさま。

@

主に国家や文明など人間の社会対象とする。

@

19 キーワード: 蛇

NG: 毒, 動物

これは、誰もがそれについて話すと青ざめるものです。体が細長く、四肢は退化している、翼も足もありません、冷血であると言われます。人間に対してとても危険なものです、種類も多く、熱帯雨林によく見られます。色が明るいほどより危険です。

Highlight:

これは、誰もがそれについて話すと青ざめるものです。体が細長く、四肢は退化している、翼も足もありません、冷血であると言われます。人間に対してとても

危険なものです、種類も多く、熱帯雨林によく見られます。色が明るいほどより危険です。

PYSUMMARIZATION:

これは、誰もがそれについて話すと青ざめるものです。

. . . 体が細長く、四肢は退化している、翼も足もありません、冷血であると言われます。

. . . 色が明るいほどより危険です。

. . .

SLP-COMPRESSION:

番号 短縮結果

これは誰もが青ざめるものです。

@

四肢は退化している翼もありません冷血であると言われます。

@

人間に対しとても危険なものです多く、熱帯雨林によく見られます。

@

色が明るいほどより危険です

@

PYSUM+SLP MIX:

番号 短縮結果

これは誰もが青ざめるものです。

@

四肢は退化している翼もありません冷血であると言われます。

@

色が明るいほどより危険です。

@

20 キーワード:反対

人の意見に共感や同意しないこと。逆らうこと。もしくは、あるものに対して位置や向き、順序、ありかた、言動などが逆の関係にあること。

Highlight:

人の意見に共感や同意しないこと。逆らうこと。もしくは、あるものに対して位置や向き、順序、ありかた、言動などが逆の関係にあること。

PYSUMMARIZATION:

人の意見に共感や同意しないこと。

・ 逆らうこと。

・ もしくは、あるものに対して位置や向き、順序、ありかた、言動などが逆の関係にあること。

・

SLP-COMPRESSION:

番号 短縮結果

人の意見に同意しないこと。

@

逆らうこと。

@

言動逆の関係にあること。

@

PYSUM+SLP MIX:

番号 短縮結果

人の意見に同意しないこと。

@

逆らうこと。

@

言動逆の関係にあること。

@