国立国語研究所学術情報リポジトリ

# アジア／ヨーロッパ言語母語の日本語学習者による品詞使用量の差異：I-JASの2つの作文タスクを用いた調査



| メタデータ | 言語: Japanese<br>出版者:<br>公開日: 2022-01-07<br>キーワード (Ja):<br>キーワード (En): International Corpus of Japanese as a Second Language (I-JAS)<br>作成者: 西島, 光洋, NISHIJIMA, Mitsuhiro<br>メールアドレス:<br>所属: |
|---|---|
| URL | https://doi.org/10.15084/00003480 |

# アジア／ヨーロッパ言語母語の日本語学習者による品詞使用量の差異ーI-JAS の 2 つの作文タスクを用いた調査ー

西島光洋 (清華大学大学院／東京工業大学大学院) *

# Differences in Part-of-Speech Usage by Japanese Learners of Asian and European Languages —Investigation Using Two Writing Tasks of I-JAS—

Mitsuhiro Nishijima (Tsinghua University / Tokyo Institute of Technology)

**要旨**

本研究では、アジア言語母語およびヨーロッパ言語母語の中級日本語学習者（以下それぞれアジア／ヨーロッパ言語母語話者）による各品詞の使用量の差異を調査した。計 11 母語の日本語学習者それぞれに対して、I-JAS のストーリーライティング（SW）タスクとエッセイ（E）タスクそれぞれにおける、各品詞（大分類・細分類）のトークン数とタイプ数の頻度を計算した。その結果、対象とする母語数を増やすと、先行研究で指摘されていたアジア／ヨーロッパ言語母語話者間の差異が確認されなくなる場合があることが分かった。また、タスクによって、アジア／ヨーロッパ言語母語話者間の差異が確認できる品詞は異なることも分かった。特に、ヨーロッパ言語母語話者はアジア言語母語話者と比べて、SW タスクでは終助詞を多用する一方で、E タスクでは口語的な助詞を豊富に使用することが判明した。この結果を基に、ヨーロッパ言語母語話者が書く文書には、文書のジャンルに依らない、文体上の共通点が存在する可能性を指摘した。

## 1. はじめに

私たちが第 2 言語（L2）で文書を書く際、自身の母語はプラスにもマイナスにも影響を与え、その文書には母語の痕跡が多かれ少なかれ残っているとされる (迫田 2020)。近年では、L2 文書の書き手の母語を機械学習手法で識別する「母語識別」という研究分野も誕生し、L2 日本語文書からの母語識別を含めて、これまで盛んに研究が行われている (Malmasi 2016, 西島他 2021)。さて、その母語識別では、書き手の母語を 100% 正確には識別できず、一般に誤識別が起こりうる。その際、フランス語とイタリア語といった言語系統的な類縁性を有する言語、ないし韓国語と中国語といった地理的な類縁性を有する言語（すなわち、言語系統的な類縁性を有するとは限らないが、使用される地域が近接している言語）が互いに誤識別されやすいという現象がこれまでに確認されている (Gebre et al. 2013, Li and Zou 2017)。つまり、L2 文書には、母語の類縁性の痕跡も残っていると推察される。

日本語の L2 習得研究でも、アジア言語母語対ヨーロッパ言語母語という母語の地理的な

* nishijima.m.ae［at］m.titech.ac.jp

類縁性の差異に着目した研究がいくつか存在する。例えば、田中他 (2016) は「読み手意識」、Ishikawa (2017) および石川 (2020a) は品詞使用量の観点から、アジア言語母語の日本語学習者とヨーロッパ言語母語の日本語学習者[(1)]が産出する日本語の差異を指摘している。

これらの研究に触発され、本稿では書き言葉における品詞使用量の観点から、アジア言語母語話者とヨーロッパ言語母語話者が産出する日本語の差異を明らかにする。その際、2 つの異なる作文タスクで産出された文書を分析に用いることで、両者の差異が確認できる品詞は文書のジャンルによって異なることを示す。ただし、これらの結果を基に文体的な観点からの考察を行い、ヨーロッパ言語母語話者が書く文書には、文書のジャンルに依らない、文体上の共通点が存在する可能性を指摘する。

本稿の構成は次の通りである。2 節では、アジア言語母語話者とヨーロッパ言語母語話者が産出する日本語の差異に言及した先行研究を述べる。3 節で本稿で行う調査の概要を述べ、4 節で調査の結果と考察を行う。5 節で本稿のまとめを述べる。

## 2. 先行研究

いくつかの研究は、アジア言語母語話者とヨーロッパ言語母語話者の差異が日本語文書に表出することを指摘している。田中他 (2016) は、ヨーロッパ言語（トルコ語、ハンガリー語、ドイツ語、フランス語、ドイツ語）母語話者が書いた、I-JAS に含まれるエッセイ（E）文書を、中国語、英語母語話者および日本語母語話者が同じテーマで書いた E 文書と比較している。その結果、「〜ましょう」といった呼びかけ表現や終助詞などの指標で測定される「読み手意識」は、中国語母語話者および日本語母語話者と比べて、ヨーロッパ言語母語話者の方が相当高いと指摘している[(2)]。

Ishikawa（石川）によってなされた 2 つの研究は、書き言葉ないし話し言葉における品詞の使用状況を各母語話者ごとに調査し、アジア言語母語話者とヨーロッパ言語母語話者による品詞使用量の差異を指摘している。Ishikawa (2017) は韓国語、中国語、英語、トルコ語母語話者および日本語母語話者によってそれぞれ書かれた I-JAS のストーリーライティング（SW）文書を用いて、助詞、名詞、動詞、助動詞、副詞、接続詞、連体詞、形容詞の使用量をそれぞれ調査している。その調査の結果は、日本語習熟度が初級から中級レベルにかけての韓国語、中国語母語話者による連体詞の使用量は、英語、トルコ語母語話者および日本語母語話者による連体詞の使用量よりも相当多いことを示している。石川 (2020a) は、I-JAS の対話データを用いて、日本語習熟度が中級後半レベルの計 12 種類の母語話者による副詞のトークン数およびタイプ数を調査している。その結果、ヨーロッパ言語（ハンガリー語、ロシア語、ドイツ語、英語、フランス語、スペイン語、トルコ語）母語話者による副詞のトークン数が、アジア言語（韓国語、中国語、ベトナム語、タイ語、インドネシア語）母語話者よりも多く、ヨーロッパ言

(1) 簡単のため、本稿では以下それぞれ、アジア言語母語話者、ヨーロッパ言語母語話者と呼ぶ。また、例えば中国語など具体的な言語名の場合も、中国語母語の日本語学習者のことを中国語母語話者などと呼ぶ。

(2) 英語母語話者が書くエッセイの傾向は、田中他 (2016) の指すヨーロッパ言語母語話者が書くエッセイの傾向とは異なることも指摘している。田中他 (2016) は、その原因として、英語でのエッセイ・ライティング教育の影響もあるのではないかと推察している。

語母語話者がアジア言語母語話者よりも副詞を多用する傾向が見られた。その一方で、タイプ数については、アジア言語母語話者とヨーロッパ言語母語話者とで明確な差は認められなかったとしている。

これらの研究に対して、いくつかの問題点が指摘できる。まず、対象とする母語数の問題である。田中他 (2016) では、ヨーロッパ言語母語話者の母語として、英語を含めて計 6 言語を対象としているのに対して、アジア言語母語話者の母語は、（母語としての日本語を除いて）中国語しか対象としていない。Ishikawa (2017) でも、学習者の母語として、韓国語、中国語、英語、トルコ語しか対象としていない。そのため、これらの結果がアジア言語母語話者とヨーロッパ言語母語話者の差異を捉えているのか、それとも特定の母語話者に特異的な現象を捉えているのかについては疑問が残る。石川 (2020a) は I-JAS に含まれる 12 種類の母語すべてを対象としているため、対象とする母語数の点では問題ない。しかし、石川 (2020a) では学習者の話し言葉を対象としているのに対して、本稿は学習者の書き言葉を対象とする点で異なる。また、石川 (2020a) は数ある品詞のうち副詞のみを対象としており、他の品詞でもアジア言語母語話者とヨーロッパ言語母語話者の差異が確認されるのかについては未知である(3)。

そして、これらの研究では、同一の基準で異なるタスク間の比較検討を行なっておらず、あるタスクで確認された品詞使用量の差異が他のタスクでも確認されるのかについてはまったく議論されていない(4)。実際、タスクによって、アジア言語母語話者とヨーロッパ言語母語話者の差異が確認できる品詞は異なる。例えば、対話タスクでは、ヨーロッパ言語母語話者はアジア言語母語話者と比べて、副詞を多用する傾向があることは先に述べた通りである。しかし、Ishikawa (2017) の SW タスクを用いた分析の結果からは、日本語習熟度が中級後半レベルの場合、中国語、英語、韓国語母語話者の順に副詞の使用量が多く、ヨーロッパ言語（英語）母語話者による副詞の使用量がアジア言語（中国語、韓国語）母語話者よりも多いという事実は認められない。

そこで、本研究では、できるだけ多くの母語話者を対象として、アジア言語母語話者とヨーロッパ言語母語話者によるあらゆる品詞の使用量の差異を、2 つの異なる作文タスクを用いて調査する。

## 3. 調査概要

### 3.1 調査する言語データ

本研究では、『多言語母語の日本語学習者横断コーパス』（International Corpus of Japanese as a Second Language, I-JAS）を用いた。I-JAS とは、計 1000 人の日本語学習者が産出した日本語の発話データおよび作文データを、日本語能力テストの点数などの背景情報とともに収

(3) その後、石川 (2020b) は、韓国語、中国語、英語母語話者および日本語母語話者に対象を絞り、発話における形容詞の使用状況を調査している。しかし、韓国語、中国語母語話者と英語母語話者との間で、形容詞使用量の差は確認できなかった。

(4) 確かに田中他 (2016) では、2 つの異なる作文タスクを用いて、ヨーロッパ言語母語話者の読み手意識の高さを調査している。しかし、その調査の方法はそれぞれのタスクで大きく異なっており、同一の基準で異なるタスク間の比較検討をしているとは言えない。

録したコーパスである (迫田他 2020)。I-JAS は、日本語学習者の母語として、韓国語（韓）、中国語（中）、ベトナム語（越）、タイ語（泰）、インドネシア語（尼）、ハンガリー語（洪）、ロシア語（露）、ドイツ語（独）、英語（英）、フランス語（仏）、スペイン語（西）、トルコ語（土）の計 12 言語を対象としている。そのうち本研究では、トルコ語を除く 11 言語を調査の対象とした。確かに先行研究において、トルコ語はヨーロッパ言語として分析されることがある (田中他 2016, 石川 2020a)。しかし、トルコ語はアルタイ語族に属するため、言語系統的な観点で見ると、典型的なヨーロッパ言語であるインド・ヨーロッパ語とは異なる。加えて、トルコ[(5)]はアジアとヨーロッパの境目に位置するため、地理的な観点からも、トルコ語はアジア言語ともヨーロッパ言語とも言えない。そのため、今回はトルコ語を除外して議論することにした。本研究では、韓国語、中国語、ベトナム語、タイ語、インドネシア語をアジア言語、ハンガリー語、ロシア語、ドイツ語、英語、フランス語、スペイン語をヨーロッパ言語と定義する。

本研究では、J-CAT の点数が 101 点から 250 点まで（つまり日本語習熟度が中級レベル）で、かつ海外教室環境で学ぶ日本語学習者が書いた SW 文書および E 文書を対象とした。SW タスクは必須調査である一方で、E タスクは任意調査であり、一部の被験者のみから E 文書が収集されている。E 文書を書いた人数を母語別に確認したところ、フランス語母語話者の数が最も少なく 15 人だった。そこで、各母語ごとの被験者数を揃えるために、E 文書を書いた被験者を各母語ごとに 15 人ずつ選択した。具体的には、フランス語母語話者は前述の 15 人をすべて選択した。その他の母語話者については、フランス語母語話者 15 人の J-CAT の得点分布となるべく一致するように 15 人をそれぞれ選択した。なお、中国語、ドイツ語、英語母語話者のデータはそれぞれ複数の国および地域で収集されているが、被験者の選択にあたっては、データの収集国・地域は特に考慮していない。ここで選択した被験者らが書いた SW 文書と E 文書を以降の調査で用いた。

SW タスクは、各被験者が 2 種類のコマ割り漫画を見てそれぞれのストーリーを記述するタスクであり、各被験者ごとに 2 つの文書がある。そこで本研究では、その 2 つの文書を繋ぎ合わせることで 1 つの文書とした。その SW 文書に対して、コーパス作成者が書き込んだ記号などを削除する前処理を行った。

一方で、E タスクは、ファーストフードと家庭料理を比較しながら、食生活についての考えを記述するというタスクである。E 文書に対しても、エッセイの本文と関係ないと考えられる記号などを削除する前処理を行った。

前処理を終えたこれらの文書に対して、日本語の自然言語処理ライブラリ GiNZA（バージョンは 4.0.0）で形態素解析を行い、各形態素に対して品詞情報を付加した。

### 3.2 調査方法

本研究では、以下の 4 つの観点から、アジア言語母語話者とヨーロッパ言語母語話者の品詞使用量の差異を調査した。

(a) 各品詞（大分類）に該当する形態素のトークン数
(b) 各品詞（細分類）に該当する形態素のトークン数

[(5)] I-JAS は、海外教室環境で学ぶトルコ語母語話者のデータをトルコで収集している。

(c) 各品詞（大分類）に該当する形態素のタイプ数
(d) 各品詞（細分類）に該当する形態素のタイプ数

具体的にはまず、(a) から (d) までの 4 つの観点から、SW 文書と E 文書それぞれにおいて、各品詞に該当する形態素の絶対頻度を各母語話者ごとに計算した。タイプ数を計算するにあたっては、形態素の書字形出現形[(6)]が完全に一致するものをまとめて 1 語とみなした。以上の操作を行うと、各観点、各文書、各母語話者に対して、各品詞の絶対頻度が得られる。観点 (a) および (b) のトークン数については、各品詞の絶対頻度を相対頻度に変換した。一方、タイプ数は文書長に比例して増加するとは限らず、本研究において各母語ごとの人数は 15 人と一定である。そこで、観点 (c) および (d) のタイプ数については、そのまま絶対頻度を用いた。以降、絶対頻度と相対頻度をまとめて頻度と呼ぶ。得られる結果をより安定にするために、各観点、各文書に対して、11 の母語話者群すべてでそれぞれ 1 回以上使用されている品詞のみを調査の対象とした。

各観点、各文書および各品詞 $i$ に対して、以下の式 (1) で定める $v(i)$ を計算する：

$$v(i) := \frac{5(\bar{r}_{i,\mathrm{A}} - \bar{r}_i)^2 + 6(\bar{r}_{i,\mathrm{E}} - \bar{r}_i)^2}{\sum_{j=1}^{5}(r_{i,\mathrm{A},j} - \bar{r}_{i,\mathrm{A}})^2 + \sum_{j=1}^{6}(r_{i,\mathrm{E},j} - \bar{r}_{i,\mathrm{E}})^2}. \tag{1}$$

ここで、上式で用いている記号の定義はそれぞれ以下の通りである。

- $r_{i,\mathrm{A},j}$ $(j = 1, \ldots, 5)$：5 種類のアジア言語母語話者による品詞 $i$ の頻度
- $r_{i,\mathrm{E},j}$ $(j = 1, \ldots, 6)$：6 種類のヨーロッパ言語母語話者による品詞 $i$ の頻度
- $\bar{r}_{i,\mathrm{A}}$：$r_{i,\mathrm{A},j}$ $(j = 1, \ldots, 5)$ の平均値
- $\bar{r}_{i,\mathrm{E}}$：$r_{i,\mathrm{E},j}$ $(j = 1, \ldots, 6)$ の平均値
- $\bar{r}_i$：$r_{i,\mathrm{A},j}$ $(j = 1, \ldots, 5)$ および $r_{i,\mathrm{E},j}$ $(j = 1, \ldots, 6)$ の平均値

この $v(i)$ という指標は、判別分析 (cf. 平井 2012) におけるクラス間変動とクラス内変動の比の最大化というアイデアに基づく。アジア言語母語話者とヨーロッパ言語母語話者による品詞 $i$ の頻度の差（$v(i)$ の分子）が大きければ大きいほど $v(i)$ の値が大きくなる。また、アジア言語母語話者およびヨーロッパ言語母語話者内部における品詞 $i$ の頻度のばらつき（$v(i)$ の分母）が小さければ小さいほど、$v(i)$ の値が大きくなる。すなわち、$v(i)$ の値が大きければ大きいほど、アジア言語母語話者およびヨーロッパ言語母語話者内部における品詞 $i$ の使用が安定的であり、かつ品詞 $i$ の使用量がアジア言語母語話者とヨーロッパ言語母語話者とで大きく異なることを意味している。

また、ある閾値を用いて、アジア言語母語話者とヨーロッパ言語母語話者による品詞 $i$ の頻度が二分できる、すなわち以下の 2 つの式

$$\begin{aligned} \max_{1 \le j \le 5} r_{i,\mathrm{A},j} &\le \min_{1 \le j \le 6} r_{i,\mathrm{E},j}, \\ \max_{1 \le j \le 6} r_{i,\mathrm{E},j} &\le \min_{1 \le j \le 5} r_{i,\mathrm{A},j} \end{aligned} \tag{2}$$

のいずれかが成り立つとき、アジア言語母語話者とヨーロッパ言語母語話者が品詞 $i$ に関して分離可能であると呼ぶことにする。

(6) GiNZA の出力における “orth_”のこと。

表 1　SW タスクにおける各母語話者による連体詞（大分類）のトークン数の相対頻度（%）

| 韓 | 中 | 越 | 泰 | 尼 | 分離可能性 |
|---|---|---|---|---|---|
| 1.69 | 1.19 | 1.02 | 0.82 | 1.76 | No |
| 洪 | 露 | 独 | 英 | 仏 | 西 |
| 0.60 | 1.23 | 1.06 | 0.73 | 0.84 | 0.65 |

表 2　E タスクにおける各母語話者による連体詞（大分類）のトークン数の相対頻度（%）

| 韓 | 中 | 越 | 泰 | 尼 | 分離可能性 |
|---|---|---|---|---|---|
| 0.87 | 0.50 | 0.77 | 0.31 | 0.80 | No |
| 洪 | 露 | 独 | 英 | 仏 | 西 |
| 0.81 | 1.18 | 0.50 | 0.60 | 0.64 | 0.42 |

## 4.　結果と考察

### 4.1　先行研究との比較

本項では、本研究の結果を Ishikawa (2017) および石川 (2020a) の結果と比較し、タスクが異なっても同様にアジア言語母語話者とヨーロッパ言語母語話者の差異は確認されるのか、そして母語数を増やしても両者の差異は確認できるのかを検証する。

(i) Ishikawa (2017) が指摘している連体詞使用量の差異について

観点 (a) を用いて、SW タスク、E タスクそれぞれにおける連体詞のトークン数の相対頻度を計算した結果を表 1 および表 2 にそれぞれ示す。表 1 から、SW タスクでは韓国語および中国語母語話者による連体詞の使用量が英語母語話者よりも多いことが観察できる。これは、Ishikawa (2017) と同様の結果である。

Ishikawa (2017) では、学習者の母語として、トルコ語を除いて、韓国語、中国語、英語の 3 つしか対象としていない。それでは、対象とする母語数を増やした場合に、アジア言語母語話者とヨーロッパ言語母語話者による連体詞の使用量の差異は確認できるだろうか。簡潔性のために、ここでは連体詞以外の品詞に対する結果は示さないが、式 (1) で定義した $v(i)$ の値は、SW 文書に出現する全 12 種類の品詞（大分類）の中で、$i$ が連体詞のとき最も大きかった。しかし、表 1 から分かるように、アジア言語母語話者とヨーロッパ言語母語話者は連体詞に関して分離不可能である。つまり、対象とする母語数を増やしたとき、アジア言語母語話者とヨーロッパ言語母語話者の差異が必ずしも確認できないことが分かる。

また、E タスクでは、母語数を絞ったとしても、SW タスクにおける連体詞使用量の傾向とは異なる。表 2 から、確かに E タスクでも韓国語母語話者による連体詞の使用量は英語母語話者よりも多いことが分かる。しかし、E タスクにおける中国語母語話者による連体詞の使用量は、英語母語話者、より広くスペイン語を除くヨーロッパ言語母語話者よりも少なく、中国語母語話者が連体詞を多く使用する現象を E タスクでは観察できな

表 3　SW タスクにおける各母語話者による副詞（大分類）のトークン数の相対頻度（%）

| 韓 | 中 | 越 | 泰 | 尼 | 分離可能性 |
|---|---|---|---|---|---|
| 0.71 | 1.70 | 1.21 | 0.61 | 1.79 | No |
| 洪 | 露 | 独 | 英 | 仏 | 西 |
| 1.28 | 1.09 | 1.40 | 1.16 | 1.65 | 0.89 |

表 4　E タスクにおける各母語話者による副詞（大分類）のトークン数の相対頻度（%）

| 韓 | 中 | 越 | 泰 | 尼 | 分離可能性 |
|---|---|---|---|---|---|
| 2.52 | 2.38 | 1.70 | 2.20 | 2.71 | No |
| 洪 | 露 | 独 | 英 | 仏 | 西 |
| 1.76 | 2.31 | 2.45 | 2.96 | 3.26 | 2.47 |

い。よって、文書のジャンルは、品詞使用量の差異に影響を与えうることが分かる。

(ii) 石川 (2020a) で指摘されている副詞使用量の差異について

SW タスクにおける副詞の使用量は、多い方から順に、インドネシア語、中国語、フランス語、ドイツ語、ハンガリー語、ベトナム語、英語、ロシア語、スペイン語、韓国語、タイ語母語話者という順序で並んでることが表 3 から確認できる。特に、韓国語、中国語、英語母語話者に対象を絞ってみると、Ishikawa (2017) における日本語習熟度が中級後半の学習者を対象とした場合の順序とまったく同じになった。一方、E タスクにおける副詞の使用量は、多い方から順に、フランス語、英語、インドネシア語、韓国語、スペイン語、ドイツ語、中国語、ロシア語、タイ語、ハンガリー語、ベトナム語母語話者という順序で並んでいることが表 4 から確認できる。この結果を SW タスクにおける結果と比較してみると、SW タスクにおける副詞使用量の多さは、E タスクにおける副詞使用量の多さと直結しないことが分かる。例えば、韓国語、中国語、英語母語話者のみに注目した場合、SW タスクでは中国語母語話者の副詞使用量が 3 者の中で最も多いのに対して、E タスクでは中国語母語話者の副詞使用量が 3 者の中で一番少ない。すなわち、連体詞の場合と同様に、文書のジャンルは、品詞使用量の差異に影響を与えうることが分かる。

### 4.2　SW タスクに対する結果

観点 (a) から (d) のそれぞれに基づいて SW タスクにおける各母語話者の品詞使用量を調べたとき、分離可能性が確認された品詞は、観点 (b) の「助詞-終助詞」（以下終助詞）および観点 (d) の「名詞-固有名詞-地名-国」（以下国名）および「助詞-係助詞」（以下係助詞）の 3 つであった。

このうち、国名と品詞付けされた形態素はすべて、漫画に登場する女性の主人公の名前「マリ」である [(7)]。被験者 ID が SES14 のスペイン語母語話者が 1 箇所半角文字で「ﾏﾘ」と入力しているため、スペイン語母語話者による国名のタイプ数が 2 で、それ以外の母語話者による

(7) GiNZA は「マリ」を一貫して国名と誤解析する。

国名に関するタイプ数が 1 であった。そのため、国名の分離可能性が確認されたが、これはスペイン語母語話者、ひいてはヨーロッパ言語母語話者に特有の現象とは言えない。また係助詞のタイプ数は、すべての母語話者に対して 2 であった。このとき、分離可能性の定義を満たすものの、各母語話者同士の差異を確認することはできない。

以下では、終助詞に関する分離可能性に注目する。表 5 は、観点 (b) を用いて、SW 文書に出現した 29 品詞のトークン数の相対頻度を各母語話者ごとに計算し、式 (1) で定義した $v(i)$ の値が大きい上位 20 品詞を示したものである。まず、表 5 から、アジア言語母語話者とヨーロッパ言語母語話者は終助詞に関して確かに分離可能であり、ヨーロッパ言語母語話者による終助詞のトークン数の相対頻度はアジア言語母語話者よりも多いことが分かる。また表 5 において、終助詞に対する $v(i)$ の値も一番大きい。そのため、他の品詞と比較しても、終助詞の使用習慣はアジア言語母語話者およびヨーロッパ言語母語話者内部でそれぞれ比較的安定しており、かつアジア言語母語話者とヨーロッパ言語母語話者とで大きく異なっていることが分かる。

終助詞は「聞き手に対する働きかけや発話時における話し手の気持ちの動きを表すモダリティ表現」(井上 2014) であり、会話でしばしば用いられる。SW 文書において終助詞と判定された形態素を含む文脈を確認したところ、その多くが登場人物の発話内容や思考内容を表す部分で用いられていることが分かった(8)。また、分離可能性こそ確認できなかったものの、表 5 において「補助記号-括弧閉」「補助記号-括弧開」（以降 2 つを合わせて括弧と呼ぶ。）に対する $v(i)$ の値は他の品詞と比べてかなり大きな値を取っている。そして、相対頻度を見ても、ヨーロッパ言語母語話者による括弧の使用量は、アジア言語母語話者よりも多い傾向にあることが分かる。以上の結果を総括すると、アジア言語母語話者と比較して、ヨーロッパ言語母語話者は物語の描写において、登場人物による発話内容や思考内容をそのままの形で伝達する直接話法をより多用する傾向があるのではないかと推察される。そこで本研究では、以下 2 つ基準

- 括弧を伴い、その括弧で包まれた引用文がある登場人物の発話内容または思考内容を表すと判断できる場合
- 括弧を伴わないが、遠藤 (1982) が提唱する話法の分類のうち、「完全直接話法」「一般直接話法」にあたると判断できる場合(9)

のいずれかを満たす場合を直接話法であると認定して、各母語話者ごとに直接話法の出現回数に関する統計を取った。その結果を表 6 に示す。表 6 から、フランス語母語話者が使用した直接話法の回数が若干少ないという例外は見られるものの、ヨーロッパ言語母語話者による直接話法の使用回数がアジア言語母語話者よりも多いという傾向が確認できる。

(8) その他、被験者 ID が CCH28 の学習者が産出した「かわいそう一日ですね」（下線は筆者による。）のように、伝達者（文書を書いた被験者）が語りの総括で用いているケースが、今回対象としたデータの中で 10 件程度見られた。しかしこの用法は、登場人物の発話内容や思考内容を表す部分で用いられる場合と比べると多くはないので、今回は考察の対象外とする。

(9) 遠藤 (1982) は、日本語の話法を、元の発話者（登場人物）の視点がより残っている順に、「完全直接話法」「一般直接話法」「修正直接話法」「一般間接話法」「拡大間接話法」の 5 つに分類している。修正直接話法では、元の発話者の視点は引用文中にまだ残っているとされる。その一方で、一般直接話法では許容されていた、会話でしばしば脱落する「は」「を」のような助詞の付加や、引用文末の丁寧体を普通体へ変換することなどが義務的になるとされる。

表 5　SW タスクで各母語話者が使用した品詞（細分類）のうち、$v(i)$ の値が大きい上位 20 品詞のトークン数の相対頻度（%）

| 品詞 | $v(i)$ | 韓 | 中 | 越 | 泰 | 尼 | 洪 | 露 | 独 | 英 | 仏 | 西 | 分離可能性 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 助詞-終助詞 | 2.04 | 0.26 | 0.17 | 0.10 | 0.11 | 0.21 | 0.42 | 0.29 | 0.31 | 0.37 | 0.32 | 0.27 | Yes |
| 名詞-普通名詞-副詞可能 | 1.38 | 4.63 | 3.68 | 3.85 | 3.83 | 4.11 | 3.12 | 2.91 | 3.89 | 3.06 | 3.37 | 2.73 | No |
| 補助記号-括弧閉 | 1.19 | 0.11 | 0.17 | 0.22 | 0.14 | 0.35 | 0.73 | 0.56 | 0.72 | 0.70 | 0.18 | 0.41 | No |
| 補助記号-括弧開 | 1.16 | 0.08 | 0.14 | 0.22 | 0.11 | 0.35 | 0.73 | 0.47 | 0.72 | 0.70 | 0.18 | 0.41 | No |
| 名詞-数詞 | 0.92 | 0.34 | 0.24 | 0.57 | 0.43 | 0.25 | 0.29 | 0.15 | 0.17 | 0.03 | 0.25 | 0.03 | No |
| 連体詞 | 0.55 | 1.69 | 1.19 | 1.02 | 0.82 | 1.76 | 0.60 | 1.23 | 1.06 | 0.73 | 0.84 | 0.65 | No |
| 代名詞 | 0.53 | 1.84 | 1.94 | 1.05 | 1.29 | 1.26 | 1.33 | 1.14 | 1.30 | 0.70 | 0.88 | 0.99 | No |
| 名詞-普通名詞-助数詞可能 | 0.40 | 0.30 | 0.27 | 0.19 | 0.57 | 0.25 | 0.29 | 0.26 | 0.20 | 0.07 | 0.18 | 0.07 | No |
| 名詞-普通名詞-サ変可能 | 0.28 | 2.71 | 2.76 | 2.74 | 2.58 | 2.74 | 2.99 | 2.32 | 2.60 | 2.53 | 2.28 | 2.32 | No |
| 名詞-固有名詞-地名-国 | 0.19 | 2.07 | 1.50 | 0.64 | 1.22 | 1.09 | 1.38 | 1.67 | 2.05 | 1.23 | 1.76 | 1.70 | No |
| 形状詞-一般 | 0.19 | 0.11 | 0.34 | 0.13 | 0.11 | 0.28 | 0.39 | 0.29 | 0.68 | 0.23 | 0.25 | 0.10 | No |
| 名詞-固有名詞-人名-名 | 0.17 | 2.41 | 1.81 | 2.93 | 2.65 | 2.74 | 2.37 | 2.44 | 2.25 | 3.06 | 4.25 | 3.71 | No |
| 補助記号-読点 | 0.10 | 3.46 | 6.88 | 5.96 | 6.05 | 5.62 | 4.82 | 4.93 | 4.68 | 6.12 | 5.30 | 4.50 | No |
| 名詞-普通名詞-一般 | 0.09 | 9.97 | 10.84 | 10.19 | 10.71 | 10.15 | 10.33 | 9.72 | 10.86 | 10.41 | 9.48 | 9.91 | No |
| 形容詞-非自立可能 | 0.08 | 0.23 | 0.55 | 0.10 | 0.07 | 0.21 | 0.26 | 0.32 | 0.27 | 0.60 | 0.18 | 0.27 | No |
| 補助記号-句点 | 0.08 | 3.69 | 1.81 | 6.37 | 2.72 | 0.63 | 6.53 | 7.34 | 0.27 | 2.73 | 2.25 | 7.12 | No |
| 接続詞 | 0.06 | 0.34 | 0.37 | 0.16 | 0.50 | 0.77 | 0.34 | 0.59 | 0.27 | 0.40 | 0.25 | 0.27 | No |
| 助動詞 | 0.04 | 16.13 | 17.08 | 16.66 | 17.66 | 17.52 | 15.69 | 17.64 | 18.92 | 16.99 | 16.93 | 17.75 | No |
| 助詞-係助詞 | 0.03 | 4.81 | 6.51 | 6.11 | 6.16 | 6.14 | 5.39 | 5.58 | 5.02 | 5.05 | 6.81 | 6.58 | No |
| 動詞-非自立可能 | 0.02 | 7.71 | 6.48 | 6.11 | 6.34 | 6.85 | 6.74 | 6.28 | 6.25 | 6.62 | 6.88 | 6.68 | No |

表 6　SW タスクにおいて、各母語話者が使用した直接話法の回数

| | 韓 | 中 | 越 | 泰 | 尼 | 洪 | 露 | 独 | 英 | 仏 | 西 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 発話 | 4 | 4 | 5 | 4 | 11 | 28 | 14 | 23 | 18 | 6 | 11 |
| 思考 | 1 | 1 | 2 | 2 | 0 | 2 | 5 | 0 | 3 | 2 | 4 |
| 合計 | 5 | 5 | 7 | 6 | 11 | 30 | 19 | 23 | 21 | 8 | 15 |

### 4.3　E タスクに対する結果

観点 (a) から (d) のそれぞれに基づいて E タスクにおける各母語話者の品詞使用量を調べたとき、分離可能性が確認された品詞は、観点 (c) の「助詞」、観点 (d) の「補助記号-読点」「助詞-係助詞」「助詞-準体助詞」「名詞-助動詞語幹」の 5 つであった。そのうち、「補助記号-読点」「助詞-係助詞」「助詞-準体助詞」「名詞-助動詞語幹」について、筆者が詳しく調べたところ、4.2 項で分離可能性が確認された国名および係助詞と同様に、あるほんの一部の学習者による特異的な使用に止まっていたり、タイプ数がすべての母語話者に対して等しいために分離可能性が検出されていた。そのため、これら 4 つの品詞は、アジア言語母語話者とヨーロッパ言語母語話者の差異を反映しているとは言えない。

以下では、助詞に関する分離可能性に注目する。表 7 は、観点 (c) を用いて、E タスクにおける各品詞のタイプ数の絶対頻度を各母語話者ごとに計算したものである。まず、表 7 から、アジア言語母語話者とヨーロッパ言語母語話者は助詞に関して確かに分離可能であり、ヨーロッパ言語母語話者による助詞のタイプ数の絶対頻度はアジア言語母語話者よりも多いことが分かる。また、助詞に対する $v(i)$ の値も表 7 において一番大きい。そのため、他の品詞と比較しても、助詞の使用習慣はアジア言語母語話者およびヨーロッパ言語母語話者内部でそれぞれ比較的安定しており、かつ両者の間で大きく異なっていることが分かる。

ヨーロッパ言語母語話者がどのような助詞を豊富に使用したことで両者の分離可能性が確認されたのかを調べるために、ヨーロッパ言語母語話者の中でも、助詞のタイプ数が特に多いドイツ語、英語、スペイン語母語話者に特に焦点を当てる。表 8 は、表 7 における助詞の行を、助詞の各下位分類ごとに分けて示したものである。表 8 から、アジア言語母語話者と比べて、ドイツ語母語話者は終助詞を、英語母語話者は副助詞を、スペイン語母語話者は接続助詞と副助詞を豊富に使用する傾向があることが分かる。そこで、次に示す 4 つの基準のいずれかに当てはまる形態素をすべて抽出した。

- ドイツ語母語話者が使用した終助詞のうち、高々 1 種のアジア言語母語話者にしか使用されていない終助詞
- 英語母語話者が使用した副助詞のうち、高々 1 種のアジア言語母語話者にしか使用されていない副助詞
- スペイン語母語話者が使用した接続助詞および副助詞のうち、高々 1 種のアジア言語母語話者にしか使用されていない接続助詞および副助詞
- 6 種のヨーロッパ言語母語話者すべてで使用されているが、高々 3 種のアジア言語母語話者にしか使用されていない助詞

表 7　E タスクで各母語話者が使用した品詞（大分類）のタイプ数の絶対頻度

| 品詞 | $v(i)$ | 韓 | 中 | 越 | 泰 | 尼 | 洪 | 露 | 独 | 英 | 仏 | 西 | 分離可能性 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 助詞 | 2.20 | 34 | 36 | 36 | 35 | 36 | 38 | 37 | 41 | 41 | 39 | 42 | Yes |
| 形容詞 | 0.44 | 49 | 48 | 56 | 44 | 48 | 62 | 58 | 52 | 56 | 65 | 45 | No |
| 副詞 | 0.25 | 46 | 39 | 33 | 37 | 39 | 36 | 43 | 39 | 44 | 49 | 49 | No |
| 連体詞 | 0.23 | 10 | 4 | 5 | 7 | 6 | 9 | 10 | 11 | 8 | 7 | 5 | No |
| 代名詞 | 0.21 | 12 | 14 | 14 | 14 | 13 | 13 | 10 | 13 | 13 | 7 | 14 | No |
| 動詞 | 0.20 | 181 | 179 | 194 | 173 | 190 | 195 | 161 | 176 | 180 | 164 | 172 | No |
| 接尾辞 | 0.12 | 18 | 18 | 19 | 22 | 26 | 23 | 17 | 19 | 19 | 13 | 20 | No |
| 接続詞 | 0.07 | 5 | 7 | 7 | 9 | 9 | 12 | 8 | 8 | 7 | 6 | 9 | No |
| 接頭辞 | 0.07 | 5 | 3 | 5 | 3 | 4 | 4 | 5 | 3 | 3 | 5 | 10 | No |
| 感動詞 | 0.04 | 2 | 1 | 2 | 1 | 4 | 2 | 2 | 3 | 1 | 2 | 5 | No |
| 名詞 | 0.03 | 384 | 433 | 461 | 300 | 373 | 487 | 432 | 423 | 372 | 313 | 424 | No |
| 助動詞 | 0.02 | 31 | 33 | 35 | 31 | 32 | 31 | 28 | 31 | 36 | 33 | 32 | No |
| 形状詞 | 0.01 | 31 | 33 | 32 | 29 | 31 | 36 | 33 | 31 | 27 | 31 | 26 | No |
| 補助記号 | 0.01 | 15 | 10 | 11 | 7 | 17 | 16 | 12 | 11 | 7 | 11 | 12 | No |

表 8　各母語話者による、助詞の各下位分類のタイプ数

| 品詞 | 韓 | 中 | 越 | 泰 | 尼 | 洪 | 露 | 独 | 英 | 仏 | 西 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 接続助詞 | 9 | 9 | 9 | 10 | 9 | 10 | 10 | 10 | 9 | 10 | 11 |
| 副助詞 | 11 | 9 | 9 | 8 | 11 | 11 | 9 | 11 | 14 | 12 | 12 |
| 係助詞 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 3 |
| 終助詞 | 2 | 4 | 4 | 5 | 3 | 3 | 5 | 6 | 4 | 4 | 4 |
| 準体助詞 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 格助詞 | 8 | 10 | 10 | 9 | 9 | 10 | 9 | 9 | 10 | 9 | 10 |
| 合計 | 34 | 36 | 36 | 35 | 36 | 38 | 37 | 41 | 41 | 39 | 42 |

その結果、誤解析の結果生じた形態素を除いて、接続助詞「けど」「けれど」「とも」、副助詞「ぐらい」「しも」「って」「どころ」、終助詞「なあ」「ね」「の」の計 10 種の形態素が抽出された。参考として、各母語話者 15 人のうち何人がこれら 10 種の形態素を使用したのかに関する統計を取った結果を表 9 に示す。これら 10 種の形態素のうち、7 種の助詞「けど」「けれど」「ぐらい」「って」「なあ」「ね」「の」はすべて口語的な助詞と言える。実際に、上で挙げた 3 つの終助詞「の」「なあ」「ね」について、4.2 項で述べたように、終助詞自体がそもそも話し言葉で用いられやすいという性質がある[(10)] 。また副助詞「ぐらい」「って」も口語性が高いとされている (グループ・ジャマシイ 2001)。そして接続助詞「けれど」「けど」も、よりカジュアルな場で使用されるとされ、口語性が高いと言える (泉原 2007)。逆に、表 9 のような表を用いて、ヨーロッパ言語母語話者と比較して多くのアジア言語母語話者が使用した助詞を目視で確認してみたが、その中に口語的な助詞は存在しなかった。以上の結果から、ヨーロッパ言語母語話者はアジア言語母語話者と比べて、口語的な助詞を豊富に使用していると言える。

### 4.4　考察

4.2 項では、アジア言語母語話者と比較して、ヨーロッパ言語母語話者は SW タスクで終助詞を多用していることを明らかにした。そして、ヨーロッパ言語母語話者は括弧を多用する傾向も見られ、これらは直接話法を多用する傾向と関連があることも述べた。一方で、4.3 項では、アジア言語母語話者と比較して、ヨーロッパ言語母語話者は E タスクで口語的な助詞をより豊富に使用していることが明らかとなった。このように、日本語文書に現れるアジア言語母語話者とヨーロッパ言語母語話者の差異は、一見タスク、つまり文書のジャンルに依存するように思われる。しかし本稿では、SW タスクにおける両者の違いと E タスクにおける両者の違いは決して無関係ではなく、ヨーロッパ言語母語話者が書く文書には、文書のジャンルに依らず、文書の読み手に臨場感を感じさせる文体的効果が存在する、という説明を試みる。

まず、直接話法（ないし直接引用）をどう捉えるかは研究者によって相違が見られるものの、直接話法では伝達者の視点とは別に、元の発話者の視点を引用文に残し、元の発話者の立

(10) ただし、すべての終助詞が口語的な性格を帯びているわけではなく、疑問の意味を表す終助詞「か」は書き言葉でも一般に用いられる。ただし、これら 3 つの終助詞はすべて、口語的な性格が強いとされている (グループ・ジャマシイ 2001, 見坊他 2014)。

表 9　各 10 種の助詞を使用した各母語話者の人数

| | 韓 | 中 | 越 | 泰 | 尼 | 洪 | 露 | 独 | 英 | 仏 | 西 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| けど | 3 | 3 | 0 | 0 | 5 | 3 | 3 | 5 | 2 | 5 | 3 |
| けれど | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| とも | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| ぐらい | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 2 | 2 | 1 |
| しも | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| って | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 1 | 1 | 2 |
| どころ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| なあ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| ね | 0 | 3 | 1 | 2 | 0 | 1 | 4 | 2 | 1 | 2 | 1 |
| の | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

場からある出来事を描写するという基本的な見方は多くの研究者の間で一致が見られる (遠藤 1982, 藤田 1985, 鎌田 1988, 砂川 1989)。特に鎌田 (1988) は、場の二重性を作り出すことによって、伝達者の場に劇的効果をもたらすことができるとして、直接話法のもたらす文体的効果に言及している。

直接話法がもたらす劇的効果に加えて、終助詞の果たす役割も見逃せない。終助詞は話し手のモダリティを表現する一形式であるが、直接話法の引用文に現れる終助詞は、伝達者ではなく、元の発話者のモダリティを表現する。そのため、終助詞を引用文中に用いることで、元の発話者の内面をより細かに描写することを可能にしている。終助詞および直接話法を多用するヨーロッパ言語母語話者は伝達者として、元の発話者の内面に入り込み、その元の発話者の視点から出来事の描写を行うため、読み手に臨場感を感じさせる語りとなっていると言える。

また、口語的な文体は、文書の読み手に語りかける文体の形成と密接に関係している。小磯他 (2011) は、口語的だという印象の形成は、読み手への語りかけの程度を含む複数の観点が関わっている可能性を示唆している。加藤他 (2014) も、口語的な文体と語りかける文体とは全同ではないとしつつも、口語的なくだけた表現が語りかけるという印象と関係している可能性に言及している。そして、読み手意識と語りかける文体との関係も見逃せない。2 節ですでに紹介したように、田中他 (2016) は呼びかけ表現や終助詞などを指標として書き手の読み手意識を測定している。彼女らは、その呼びかけ表現や終助詞などが多く含まれているヨーロッパ言語母語話者の E 文書を「まるで読み手に語りかけるよう」だと評している。また加藤他 (2014) も、語りかける印象を与える表現が、読み手意識を感じさせる特徴を備えていると指摘している。換言すれば、田中他 (2016) の研究が指摘しているヨーロッパ言語母語話者の読み手意識の高さも、語りかける文体の形成に寄与していると言えよう。以上をまとめると、ヨーロッパ言語母語話者が書く文書の口語的な文体や読み手意識の高さが、語りかける文体を醸成しているのである。確かに、語りかける文体と臨場感を感じさせる文体との関係については今

後更なる議論が必要であるものの[11]、語りかける文体は、文書の読み手に書き手との距離をより近く感じさせ、書き手と正に対話しているかのような臨場感を感じさせる要素の 1 つと言えるだろう。

## 5. おわりに

本稿では、日本語習熟度が中級レベルのアジア言語母語話者とヨーロッパ言語母語話者による品詞使用量の差異を、I-JAS の 2 つの作文タスクを用いて調査した。その結果、アジア言語母語話者と比較して、ヨーロッパ言語母語話者は SW タスクで終助詞および括弧を多用する傾向が見られ、これらは直接話法を多用する傾向と関連があることも述べた。一方で、ヨーロッパ言語母語話者は E タスクで口語的な助詞をより豊富に使用していることが明らかとなった。このように、両者の違いが反映される具体的な品詞は、文書のジャンルによって異なる。しかし、文体に着目すると文書のジャンルに依らない共通点が見られ、ヨーロッパ言語母語話者が書く文書には、読み手に臨場感を感じさせる文体的効果が存在する可能性があることが分かった。

今後の課題として、次の 4 点が挙げられる。まず、本研究では作文タスクを用いてアジア言語母語話者とヨーロッパ言語母語話者による品詞使用量の差異を調査した。しかし、石川 (2020a) が行なっているように、発話タスクにおいて両者の差異はどの品詞に現れるのかに関する包括的な研究も必要である。

次に、本研究では品詞使用量に着目して、アジア言語母語話者とヨーロッパ言語母語話者が産出する日本語の差異を論じた。しかし、両者の差異が現れるのは、品詞使用量だけに止まらない。例えば、奥野 (2018) および奥野他 (2019) は、能動態・受動態の使用偏向には両者の差異が現れることを指摘している。したがって、品詞使用量および態の使用偏向以外に、両者の差異がどのような形で現れるのかについては更なる研究が必要である[12]。

また、ヨーロッパ言語母語話者がアジア言語母語話者と比べて SW タスクで終助詞を多用する現象は前述の通りだが、その原因については明らかでない。ここでは、アジア言語とヨーロッパ言語をそれぞれ代表して、筆者の解する中国語と英語を基に考えてみたい。英語には文末に置いて話し手のモダリティを表すような助詞は存在せず、文中の助動詞などを通して話し手のモダリティが表現される。その一方で、中国語には「語気助詞」と呼ばれる文末助詞が存在し、日本語の終助詞と同様に話し手のモダリティを表現し得る。そのため、対照分析的な観点から考えると、中国語母語話者は英語母語話者よりも終助詞を多用すると予測するのが自然であるが、本稿の結果はこの予測とは逆の結果を示している。したがって、どうして予測に反して、SW タスクでヨーロッパ言語母語話者がアジア言語母語話者よりも終助詞を多用したの

[11] 小磯他 (2011) は、テキストの印象を表す評価語対を構成するために、いくつかのテキストを用意して、複数の被験者にそれらのテキストから受ける印象を自由記述で回答させている。その結果、「読み手に語りかける—語りかけの少ない」とは別に、「臨場感のある—臨場感のない」という評価語対が得られている。しかし、両者が類似した概念かどうかについては言及されていない。

[12] 筆者は、本稿で示した結果以外にも、Universal Dependencies に基づく依存関係ラベルに対して、$v(i)$ の値および分離可能性を分析した。しかし、分離可能性が確認された依存関係ラベルは存在せず、$v(i)$ が突出して大きい値を取るような依存関係ラベルも存在しなかった。

かについては、今後検討の余地がある。

最後に、「臨場感を感じさせる文体」という言葉でヨーロッパ言語母語話者が書く文書に共通して存在する文体的効果をまとめあげること自体が適切なのかどうかについても一考の余地がある。本稿ではまずヨーロッパ言語母語話者が書く文書は「口語的な文体」を有することを指摘した上で、「口語的な文体」と「読み手意識」から「語りかける文体」を導き出した。その後、「語りかける文体」から「臨場感を感じさせる文体」を導き出した。しかし、「語りかける文体」を間に挟んで結びつけられた「口語的な文体」と「臨場感を感じさせる文体」とは一体どのような関係にあるのかについては慎重に吟味する必要がある。

## 文　献


迫田久美子 (2020)『改訂版 日本語教育に生かす第二言語習得研究』アルク.

Shervin Malmasi (2016) “Native Language Identification: Explorations and Applications.” Ph.D. dissertation, Macquarie University.

西島光洋・劉穎・中田和秀 (2021)「日本語学習者コーパス I-JAS を用いた母語識別」言語処理学会第 27 回年次大会発表論文集, pp. 1028–1032.

Binyam Gebrekidan Gebre, Marcos Zampieri, Peter Wittenburg, and Tom Heskes (2013) “Improving Native Language Identification with TF-IDF Weighting.” *Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 216–223.

Wen Li, and Liang Zou (2017) “Classifier Stacking for Native Language Identification.” *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 390–397.

田中真理・迫田久美子・野田尚史 (2016)「日本語学習者コーパスにおける対話―ロールプレイ、メール、エッセイの分析をとおして―」ヨーロッパ日本語教育, 20, pp. 102–119.

Shin’ichiro Ishikawa (2017)“Learners’ Acquisition and Use of L2 Japanese Vocabulary: Influence of L1 Backgrounds and L2 Proficiency Levels—A Learner Corpus-based Analysis—.” *Acquisition of Japanese as a Second Language*, 20, pp. 10–27.

石川慎一郎 (2020a)「発話における副詞の使用」『日本語学習者コーパス I-JAS 入門―研究・教育にどう使うか―』くろしお出版, pp. 167–184.

石川慎一郎 (2020b)「日本語・中国語・韓国語・英語母語話者の日本語発話における形容詞使用実態―I-JAS に基づく調査―」日本語学習者コーパス「I-JAS」完成記念シンポジウム予稿集, pp. 27–34.

迫田久美子・石川慎一郎・李在鎬 (編) (2020)『日本語学習者コーパス I-JAS 入門―研究・教育にどう使うか―』くろしお出版.

平井有三 (2012)『はじめてのパターン認識』森北出版.

井上優 (2014)「終助詞」『日本語文法事典』大修館書店, pp. 265–266.

遠藤裕子 (1982)「日本語の話法」言語, 11:3, pp. 86–94.

グループ・ジャマシイ (2001)『中文版日本語句型辞典: 日本語文型辞典中国語訳簡体字版』くろしお出版.

見坊豪紀・市川孝・飛田良文・山崎誠・飯間浩明・塩田雄大 (2014)『三省堂国語辞典』第七版 三省堂.

泉原省二 (2007)『日本語類義表現使い分け辞典』研究社.

藤田保幸 (1985)「「内的引用」における話法の転換について―話法転換の a 線―」語文, 46, pp. 14–21.

鎌田修 (1988)「日本語の伝達表現」日本語学, 7:9, pp. 59–72.

砂川有里子 (1989)「引用と話法」『講座日本語と日本語教育（4）日本語の文法・文体（上)』明治書院, pp. 355–387.

小磯花絵・田中弥生・小木曽智信・近藤明日子 (2011)「テキストの多様性をとらえる分類指標の構築を目指して」特定領域研究「日本語コーパス」平成 22 年度公開ワークショップ（研究成果報告会）予稿集, pp. 431–438.

加藤祥・柏野和佳子・立花幸子・丸山岳彦 (2014)「語りかける書きことばの表現」国立国語研究所論集, 8, pp. 85–108.

奥野由紀子 (2018)「日本語学習者に共通して見られる現象と母語による違い―I-JAS におけるストーリー描写課題の分析より―」日本語教育連絡会議論文集, 30, pp. 67–75.

奥野由紀子・呉佳穎・村田裕美子 (2019)「日本語学習者の能動態と受動態の使用傾向にみられる母語による違い―中国語とドイツ語での語りの比較から―」日本語研究, 39, pp. 79–93.

**関連 URL**

| | |
|---|---|
| 『多言語母語の日本語学習者横断コーパス』 | https://www2.ninjal.ac.jp/jll/lsaj/ |
| GiNZA | https://megagonlabs.github.io/ginza/ |