

# プリエディット手法としての産業日本語に関する一考察

A Consideration for Technical Japanese as a Pre-Editing Technique



株式会社日本取引所グループ 総合企画部フィンテック推進室

**土井 惟成**

2015年株式会社日本取引所グループに入社。東京証券取引所 IT 開発部などを経て、2018年より現職。

✉ n-doi@jpx.co.jp

## 1 はじめに

東京証券取引所（以下、東証）は 3,700 社を超える企業（2020 年 9 月時点）が上場している世界最大の証券市場の一つである。東証上場会社は、投資家の投資判断に影響を与えうる情報を、東証が運営する Web システム（以下、TDnet）を通じて開示することが義務付けられている。上場会社が開示する会社情報は時々刻々と発生することから、その規模は膨大であり、2018 年に TDnet で開示された日本語の上場会社開示資料（以下、開示資料）は約 7.9 万文書、総ページ数は約 71 万ページに及んだ。

東証における海外投資家のプレゼンスは年々高まっており、2019 年には、日本の上場会社の株式に対する外国法人の保有比率（時価総額に対する金額ベース）は約 30% に及んでいる。しかしながら、2018 年における TDnet 上の開示資料では、英語による開示資料の占める割合（件数ベース）は約 11% に留まっている。東証及び日本取引所グループ（以下、JPX）では、このような開示資料の英訳を巡る状況の改善を目的として、機械翻訳に係る調査研究を継続的に行っている。

JPX における調査研究対象の一つとして、機械翻訳しやすい日本語への言い換え（プリエディット）が挙げられる。先行研究では、開示資料の機械翻訳の翻訳品質向上の手段として、産業日本語への言い換への有効性の検証を行った[1]。この時、言い換えルールとして特許

ライティングマニュアル第 2 版<sup>1</sup> を利用し、複数の作業員に対して開示資料中のテキストの言い換えを依頼した。検証の結果、開示資料においても特許ライティングマニュアルが準用できる可能性を見出した一方で、本マニュアルだけでは不足している点について示唆を得ることができた。

本稿では、JPX における機械翻訳に係る取り組みの概略を紹介のうえ、文献[1]における実験手順及び実験結果に触れつつ、プリエディット手法としての産業日本語に関する追加の考察等を述べる。

## 2 JPX における機械翻訳の取り組み

JPX における機械翻訳の調査研究には、文献[1]のプリエディットの検討のほか、開示資料に特化したニューラル機械翻訳モデル[2]（以下、NMT モデル）の検討や、大規模対訳コーパスの構築が挙げられる。

まず、開示資料に特化した NMT モデルに係る調査研究について説明する。一般に、汎用的な NMT モデルでは、専門用語の多い特定分野の文の翻訳品質は劣化する。この問題を解決する方法として、汎用的な学習データと特定分野の学習データを組み合わせることで、当該分野の機械翻訳の品質を高めるといった、カスタム翻訳モデルの利用が挙げられる。先行研究[3]では、開示資料におけるカスタム翻訳モデルの有用性を評価するため、開示資料の一つであるコーポレート・ガバナンス報告書(以

<sup>1</sup> <https://www.tech-jpn.jp/tokkyo-writing-manual/>

下、CG 報告書)を対象に、対訳コーパスの作成及びカスタム翻訳モデルを作成した。実験の結果、特定分野に即した専門用語や固有名詞の訳出によって翻訳品質の向上を確認した。ただし、一部に固有名詞の誤訳が残ったほか、文脈から判断すると不適切な翻訳といった、NMT モデルに一般的な誤訳は依然として散見された。

次に、大規模対訳コーパスの構築に係る調査研究について説明する。前述[3]の結果より、開示資料に特化した対訳コーパスの有用性が確認できたことから、JPX では開示資料に特化した対訳コーパスを構築している。この例として、TDDC (Timely Disclosure Documents Corpus) [4]と東証適時開示コーパス<sup>2</sup>が挙げられる。TDDC は、2016 年 1 月から 2018 年 6 月までの 30 ヶ月において TDnet で開示された資料 (PDF 形式) から手作業で構築された、約 140 万文対の対訳コーパスである。TDDC は第 6 回アジア翻訳ワークショップ (WAT 2019)<sup>3</sup>の公式コーパスとして提供されており、本ワークショップ参加者の評価結果は Web サイト上で公開されている。一方、東証適時開示コーパスは、TDnet 上の開示資料から構築したコーパスであり、その一つに大規模対訳コーパス (適時開示資料対訳コーパス (日英))がある。この対訳コーパスが TDDC と異なる点として、PDF 形式の開示資料からのテキスト抽出及び文対のアライメントが、手作業ではなく機械的な手法によって行われたことが挙げられる。いずれの対訳コーパスも、機械翻訳モデルの精度向上に寄与しており、学習データとしては十分な規模であることが示唆されている。しかしながら、数値及び固有名詞の誤訳、訳語の一貫性の不備、訳抜け等は改善できず、問題点として依然指摘されている。

これらの調査研究では、開示資料における全体的な機械翻訳の品質向上を確認したものの、固有名詞や数値をはじめとする、投資家にとって極めて重要な情報については、誤訳の大幅な低減は見受けられなかった。そこで JPX では、これらの誤訳を回避する方法として、NMT モデルの外部におけるテキストの制御に係る調査研究を行っている。

2 <https://www.jpx.co.jp/markets/paid-info-equities/misc/O4.html>

3 <http://lotus.kuee.kyoto-u.ac.jp/WAT/WAT2019/index.html>

### 3 プリエディットと産業日本語

近年、産業翻訳における機械翻訳の利用においては、機械翻訳そのものの精度向上と併せて、機械翻訳を前提としたワークフローの検討が進んでいる。その具体例として、ポストエディットやプリエディットの活用が挙げられる。ポストエディットとは機械翻訳モデルの出力文を人手によって後編集を行うことであり、プリエディットとは機械翻訳前に入力文を翻訳しやすいように言い換える作業である。開示資料中のテキストは長文かつ文構造が複雑になりやすい傾向にあることから、プリエディットによる機械翻訳への品質向上は特に期待できると考えられる。

プリエディットに密接した概念として、制限言語が存在する[5]。制限言語とは、人間による読みやすさや機械での処理しやすさを向上させるために、自然言語の構文や語彙等に制限を設けた言語のことである。制限言語の代表例として、ブレイン・イングリッシュ (Plain English)、やさしい日本語、産業日本語が挙げられる。産業日本語とは、「産業・技術情報を人に理解しやすく、かつ、コンピュータ (機械) にも処理しやすく表現するための日本語」と産業日本語研究会によって定義されている。開示資料のプリエディットにおいては、人間 (上場会社及び投資家) と機械 (機械翻訳) の双方にとっての可読性が求められることから、産業日本語はプリエディット後のテキストの指針として適していると考えられる。

そこで、文献[1]の研究では、開示資料においては産業日本語へのプリエディットによる翻訳品質の向上が期待できると考え、実験によってこの有効性を評価した。本研究では、開示資料の一つである CG 報告書中のテキストを対象に、産業日本語への言い換えをクラウドワーカー等に依頼のうえ、NMT モデルへの入出力結果を比較することで、プリエディットの有効性を検討した。この概要を【図 1】に示す。以下では、本研究の実験内容等とプリエディットに関する追加の考察等を述べる。

### 4 産業日本語への言い換え

#### 4.1 言い換え作業

実験対象のテキストデータは、2019 年 7 月までに

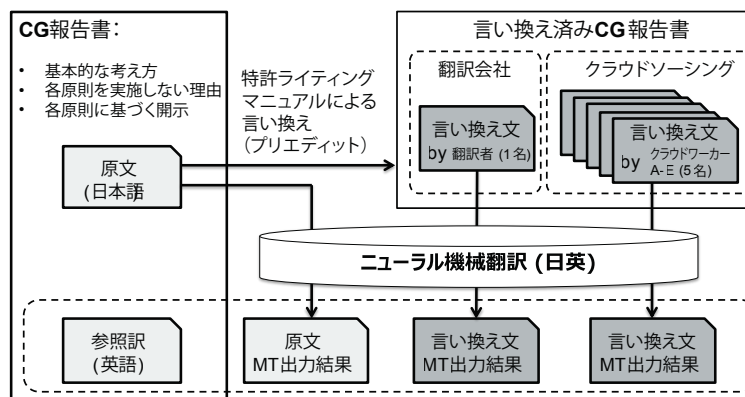


図1 文献[1]における実験の概要

日本語及び英語で開示されたCG報告書からパラグラフ単位で抽出した、全591文のテキストとした。パラグラフの選定に当たっては、1つのパラグラフ中に100文字以上の文が最低1つ含まれるようにした。この文量等を【表1】に示す。

本研究では、リエディットの言い換えルールとして特許ライティングマニュアル第2版（以下、本マニュアル）を利用した。この理由として、本マニュアルは言い換えルールが用例と併せてまとまっていること、また、清藤の発表[6]において、特許文書における機械翻訳では産業日本語へのリエディットの有用性が示唆されていたことが挙げられる。本マニュアルにおける言い換えルール<sup>4</sup>の概要を【表2】に示す。また、100文字以上の文にはルール1の短文化を行うよう、各作業者には明示的に指示した。

本研究では、これらの言い換え作業を、クラウドワーカー5名と翻訳者1名に依頼した。この目的の一つとして、日英翻訳の経験の有無による、リエディット結果への影響の有無の検証が挙げられる。

## 4.2 作業者による言い換え結果

5名のクラウドワーカー（A-E）及び翻訳者におけるリエディット時の適用ルール数の分布を【表3】に示す。【表3】より、ルール1の短文化以外の適用数は作業者による差異が大きいと言える。この要因として、ルール1には文字数という定量的な基準が設けられており、それ以外のルールには適用すべきかどうかの定量的な基

4 特許ライティングマニュアル第2版では、27の言い換えルールを7つのカテゴリに区分しているが、本稿ではこの各カテゴリに属する言い換えルールのそれぞれを総称してルール1-7と記載する。

表1 対象データの規模

項目	内容
パラグラフ数	218パラグラフ
総文数	591文（1パラグラフ当たり平均2.71文）
総文字数	63,896字（1文当たり平均108.1字）

表2 本研究で利用した言い換えルール

ルール	名称	概要
1	短文化	一文を短くして複雑な係り受けを無くす。
2	主語・目的語等の明示	主語や目的語等の省略された要素を明示して、文意を明瞭にする。
3	文構造の明確化	文の構造を工夫し、理解しやすい文にする。
4	並列要素の整理	要素を対等に並べることで、係り受けを明確にする。
5	読点の付与	適切な位置に読点を付与することで、係り受けを明確にする。
6	簡潔化	不要、冗長、難解な表現を避け、文意が変わらない簡潔な表現を用いる。
7	単語の言い換え	多義的な表現や曖昧な表現を明確・具体的な表現に改める。

表3 リエディットにおける適用ルール数（「ルールX」を「R.X」として表記）（総文数：591文）

対象	R.1	R.2	R.3	R.4	R.5	R.6	R.7	N/A
A	342	90	62	36	108	20	67	160
B	369	238	31	2	45	44	292	47
C	333	45	32	11	82	13	64	147
D	440	498	287	86	216	229	309	6
E	379	22	91	22	88	11	15	165
翻訳者	348	430	76	273	217	50	189	53

準が無いことが考えられる。また、同一の言い換えルールを適用していたとしても、作業者間において言い換え結果の差が生じている事例が散見された。以下では、不適切な言い換えの事例や、言い換えの可否判断が難しい

事例等を紹介する。

#### 4.2.1 短文化に伴う補足漏れ

ルール1の短文化の誤った適用事例として、短文化後の後半の文において、主語や目的語、接続詞等が漏れている例が散見された。この例を【表4】の例1に示す。この要因として、短文化の実施後、全体を通した読みやすさの確認や他のルールの検討を、各作業者が失念していたことが考えられる。この結果、各作業員においてルール2の主語・目的語等の明示における適用の判断にばらつきが生じ、【表3】におけるルール2の適用数が作業員間で大きく異なると推察する。

#### 4.2.2 主語及び目的語等の補足の誤り

ルール2の主語・目的語等の明示において、前後の文脈を誤解して、誤った主語及び目的語の補足を行っている事例が散見された。【表4】の例2では、前後の文脈から、ある企業が取締役会の実効性評価の結果について述べており、その主語は「当社」が適切であるところ、作業員は「社外取締役」がそのような評価を行ったと誤認した。このような主語等の補足の誤りは、特定の作業員に限らず生じており、翻訳者においても同様に誤った補足を行っている事例を確認した。本実験対象のテキストデータでは、各文の主語及び目的語の候補が複数存在しており、補足する語句の誤りを誘発しやすかったと考えられる。ただし、この誤りは、文の作成者ではない第三者が主語及び目的語を補足したことに起因することから、開示資料の作成者自身がプリエディットを行う際には生じにくいと推察する。

#### 4.2.3 言い換えルールの過剰な適用

ある作業員においては、過剰に言い換えルールを適用

する傾向があり、言い換え後の文がかえって読みにくくなってしまうという事象を確認した。この例を【表4】の例3に示す。この原因として、当該作業員は、言い換え作業をより多く実施することが目的となってしまう、文としての読みやすさの改善を軽視してしまったことが考えられる。

#### 4.2.4 明確化すべき係り受けの判断の難しさ

産業日本語では文節の係り受けを明確にすることが求められており、そのための言い換えルールとして、ルール1の短文化、ルール3の文構造の明確化、ルール4の並列要素の整理、ルール5の読点の付与が挙げられる。この中でも、ルール3とルール5は、適用可否の判断における作業員への依存が大きく、結果として作業員間の作業結果に差が生じやすくなっていると考えられる。

ルール3の文構造の明確化とルール5の読点の付与は、全体の文構造を把握した上で、文構造が明確になるよう、語順の変更等を検討する必要がある。この検討の際、作業員は各文節の係り受けを把握し、それらの距離の妥当性の評価が求められる。そのため、本マニュアルを参考にしたとしても、これらの言い換えルールの適用可否の判断には、作業員自身の能力への依存が大きいと推察する。

#### 4.2.5 簡潔化すべき表現の判断の難しさ

ルール6の簡潔化とルール7の単語の言い換えは、概ね適切に行われていた。ただし、それぞれの適用可否の判断は作業員にとって困難であり、その結果として【表3】の適用数に差が生じたと推察する。

産業日本語では不要な表現や多義的な表現を避け、簡潔な表現が求められる。これを実現するための言い換えルールとして、本マニュアルではルール6の簡潔化と

表4 不適切な言い換え例（下線部は言い換え箇所を示す）

例	原文	不適切な言い換え文	適用ルール
1	当行の取締役会は、社外取締役がその半数を占める構成となっており、客観性と透明性を確保できる体制となっております。	当行の取締役会は、社外取締役がその半数を占める構成です。 <u>客観性と透明性を確保できる体制となっております。</u>	1
2	一方、今後の課題として、取締役会の多様性の確保や社外取締役の比率の検討等が必要であることを確認いたしました。	一方、今後の課題として、取締役会の多様性の確保や社外取締役の比率の検討等が必要であることを <u>社外取締役が確認</u> いたしました。	2
3	一方、ジェンダーにおける多様性については課題と認識しており、実効性および適正規模の観点から女性取締役の登用を継続的に検討してまいります。	一方、 <u>当社は、ジェンダーの多様性を、課題と認識しています。そこで、実効性および適正規模の観点から、当社は、女性取締役の登用を、引き続き検討します。</u>	1, 2, 3, 6, 7



ルール7の単語の言い換えが定められている。本マニュアルでは、簡潔化すべき表現の実例を幅広く記載することで、簡潔化すべき表現を外延的に定義している。しかしながら、本マニュアルに記載されている実例は特許文書で頻出な語句であり、開示資料には必ずしも即しておらず、作業者の判断に困難が生じたと考えられる。ただし、簡潔化の作業そのものは本マニュアルが参考となることから、不適切な簡潔化は生じにくかったと推察する。

### 4.3 作業者の感想

本作業の実施後、各作業者に対して、言い換えが困難だった事例、適用可否の判断が困難だった言い換えルール等についてアンケートを実施した。本節では本アンケート結果の概要を述べる。

まず、言い換えが困難だった事例としては、ほぼ全員の作業者が長文の言い換えを挙げた。ルール1の短文化の適用可否は作業者にとって明らかであったものの、短文化作業そのものは、文構造全体の把握が難しかった旨の回答が大部分を占めた。特に、150文字以上の文は、途中で主語が変わることが多く、係り受けの判断に労力を要したという所感が寄せられた。

適用可否の判断が困難だった言い換えルールとしては、明確化すべき係り受けの判断に関連するルール3（文構造の明確化）とルール5（読点付与）、表現の簡潔化に関連するルール6（簡潔化）とルール7（単語の言い換え）が挙げられた。これらは、4.2.4項及び4.2.5項の記載を支持すると考える。また、適用可否の判断を助けるための施策として、複数の作業者より例文の充実化やFAQの作成の提案があった。

全体を通じた所感としては、「日本人であるのにも関わらず、難しいと感じました」、「母語であるはずの日本語をきちんと理解できていないことが分かりました」といった、本作業の難しさに係る感想が多数寄せられた。また、「いい経験ができました」、「このような言い換えは、言語の構造を考えるのに有効だと思います」といった、本作業に対する前向きな感想も寄せられた。

### 4.4 小括

本研究では、言い換えルールが用例と併せてまとまっていること等を踏まえ、作業者の言い換え作業の仕様書として本マニュアルを採択した。本作業の結果、開示資

料のテキストの場合、本マニュアルだけでは作業者間において言い換え結果の差が生じることが判明した。これを解決する方法として、本マニュアルに加えて、開示資料に即した例文の充実化やFAQの作成が挙げられる。

## 5 翻訳品質への影響

### 5.1 実験設定

本研究では、前述のプリエディットの後、原文と各言い換え文を汎用的なNMTサービスに入力し、プリエディットされた文の翻訳品質を評価した。汎用的なNMTサービスにはCloud Translation API<sup>5</sup>を採択した。翻訳品質の評価には、BLEU[7]とNIST[8]による自動評価と、特許庁が定める「特許文献機械翻訳の品質評価手順」<sup>6</sup>中の「内容の伝達レベル」による人手評価を採択した。この人手評価では、文単位で2名の翻訳者による5段階の主観評価を行い、それぞれの平均値を算出した。また、各言い換えルールの有効性の評価のため、クラウドワーカー及び翻訳者の人手評価結果に対して重回帰分析を行った。

### 5.2 実験結果及び考察

自動評価の結果を【表5】、人手評価の結果を【表6】、重回帰分析の結果を【表7】、言い換え文及び翻訳文のサンプルを【表8】に示す。各表におけるp値は両側検定によって算出した。

【表5】より、一部の作業者ではBLEUスコアの有意な向上は見られなかったものの、NISTスコアは全ての作業者において有意差が確認できた。NISTスコアの方がより有意に向上していることから、内容語の出力が特に改善したと考えられる。また、【表6】より、クラウドワーカーと翻訳者の両方で人手評価の有意な改善が確認できた。これらの結果は、本マニュアルによるプリエディットが、機械翻訳の出力文の品質向上に寄与することを支持している。

【表7】より、人手評価結果の向上に特に寄与した言い換えルールは、ルール1の短文化だと言える。この要

5 <https://cloud.google.com/translate/?hl=ja>

6 [https://www.jpo.go.jp/system/laws/sesaku/kikaihonyaku/document/tokkyohonyaku\\_hyouka/O1.pdf](https://www.jpo.go.jp/system/laws/sesaku/kikaihonyaku/document/tokkyohonyaku_hyouka/O1.pdf)

因として、短文化による係り受けの明確化や訳抜けの抑制が考えられる。【表8】の例1のように、短文化では文節の係り受けが明確になることで、NMTモデルにとって文意が把握しやすくなることが期待される。また、【表8】の例2のように、NMTモデルでは長文における訳抜けの発生が知られているが、短文化によってその抑制が期待される。

【表7】によると、他の言い換えルールでは翻訳精度の大幅な向上は見受けられなかった。この理由としては、4.2節に挙げたような、各作業における不適切な言い換えに起因すると考えられ、必ずしもこれらの言い換えルールが翻訳品質の向上に寄与しないとは言えない。各言い換えルールの有効性の評価のためには、言い換え作業時における各言い換えルールの難しさ等の評価や、適

表5 自動評価結果

対象	BLEU			NIST		
	スコア	差分	p値	スコア	差分	p値
原文	22.98	-	-	2.66	-	-
A	24.13	1.14	<0.01	2.76	0.10	<0.01
B	23.7	0.72	<0.05	2.78	0.11	<0.01
C	23.93	0.94	<0.01	2.76	0.09	<0.01
D	23.37	0.38	0.36	2.76	0.10	<0.01
E	23.55	0.56	<0.05	2.75	0.08	<0.01
翻訳者	23.86	0.88	<0.05	2.77	0.10	<0.01

表6 人手評価結果

対象	スコア	差分	p値
原文	3.78	-	-
クラウドワーカー	4.13	0.35	<0.01
翻訳者	4.23	0.45	<0.01

表7 重回帰分析の結果 (R2 = 0.20、文数 = 1182)

ルール	名称	係数	標準誤差	t値	p値
1	短文化	0.73	0.05	15.8	<0.01
2	主語・目的語等の明示	0.00	0.05	-0.07	0.94
3	文構造の明確化	0.00	0.06	0.03	0.98
4	並列要素の整理	0.11	0.05	2.13	<0.05
5	読点の付与	0.02	0.05	0.46	0.64
6	簡潔化	0.07	0.07	0.96	0.34
7	単語の言い換え	0.01	0.05	0.11	0.91

表8 言い換え文及びMT文のサンプル (括弧中の文字は言い換え作業、数値は人手評価結果を示す)

例1	原文	なお、当社は、女性役員として監査役1名を選任しており、 <b>当該監査役は</b> 取締役会へ出席し、議論へ参加しております。
	参照訳	The Company has appointed one female officer as a corporate auditor, and <b>the corporate auditor attended</b> meetings of the Board of Directors and participated in discussion.
	原文 MT 文 (4.0)	The Company has appointed one Audit & Supervisory Board Member as <b>a female officer who attends</b> the Board of Directors meetings and participates in discussions.
	言い換え文 (翻訳者)	なお、当社は、女性役員として監査役1名を選任しております。 <b>当該監査役は</b> 取締役会へ出席し、議論へ参加しております。
	言い換え MT 文 (5.0)	The Company has appointed one corporate auditor as a female director. <b>The corporate auditor attends</b> the board of directors and participates in discussions.
例2	原文	当社は、任意の指名・報酬委員会などの独立した諮問委員会を設置していませんが、取締役候補の選任や取締役の報酬については、 <b>取締役会の決議に先立ち</b> 、独立社外取締役及び親会社に対し説明を行い、適切な助言を得ております。
	参照訳	Although NTT DATA has not set up an independent advisory committee such as a voluntary nomination committee or remuneration committee, <b>in advance of the resolution by the Board of Directors</b> , we provide independent directors and the parent companies with explanations about the nomination of candidates for directors and remuneration for directors and receive appropriate advice from them.
	原文 MT 文 (2.0)	The Company has not established an independent advisory committee such as a voluntary nomination and remuneration committee. We give explanations and obtain appropriate advice.
	言い換え文 (D)	当社は、任意の指名・報酬委員会などの独立した諮問委員会を設置していません。ですが、当社は、取締役候補の選任や取締役の報酬については、 <b>取締役会の決議に先立ち</b> 、独立社外取締役及び親会社に対し説明を行い、適切な助言を得ています。
	言い換え MT 文 (5.0)	The Company has not established an independent advisory committee such as a voluntary nomination and compensation committee. <b>However, prior to the resolution of the Board of Directors</b> , the Company explains the selection of director candidates and director remuneration to independent outside directors and the parent company to obtain appropriate advice.

切なプリエディットを前提とした場合の翻訳品質への寄与等の評価が必要と考える。

## 6 おわりに

本研究では、開示資料の機械翻訳におけるプリエディットの有効性を検討した。実験の結果、CG 報告書を対象にしたプリエディットにおいて、特許ライティングマニュアルによる産業日本語への言い換えルールが準用可能であること、また、短文化が特に有効であることを示した。一方で、開示資料のテキストの場合、本マニュアルだけでは作業員間において言い換えの結果に差が生じることが判明した。そのため、開示資料における適切なプリエディットを助けるための施策や、言い換えルールごとの更なる分析が望まれると考える。

以上を踏まえ、今後の課題としては次が挙げられる。

- ・ 開示資料に即したプリエディットマニュアルの検討（言い換え例の拡充、特許ライティングマニュアル以外の言い換えルール等）
- ・ プリエディットの難しさに係る更なる分析
- ・ 他の開示資料におけるプリエディットの有効性の検証
- ・ ポストエディット及びプリエディットを前提とした開示資料の産業翻訳ワークフローに向けた検討

## 参考文献

- [1] 土井惟成, 大西恒彰, 百石弘澄, 高頭俊, 山藤敦史. 上場企業開示資料の機械翻訳におけるプリエディットの検討. 言語処理学会第 26 回年次大会 (NLP2020), pp. 525-929, 3 2020.
- [2] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS' 14, pp. 3104-3112, Cambridge, MA, USA, 2014. MIT Press.
- [3] 土井惟成, 近藤真史, 山藤敦史. コーポレート・ガバナンス報告書における機械翻訳の検討. 言語処理学会第 25 回年次大会 (NLP2019), pp. 926-929, 3 2019.
- [4] Nobushige Doi, Yusuke Oda, and Toshiaki Nakazawa. TDDC: Timely Disclosure Documents Corpus. In Proceedings of The 12th Language Resources and Evaluation Conference (LREC 2020), pp. 3719-3726, Marseille, France, 5 2020.
- [5] 宮田玲, 藤田篤. 機械翻訳向けプリエディットの有効性と多様性の調査. 通訳翻訳研究への招待, pp. 53-72, 2017.
- [6] 清藤弘晃. 産業日本語の考え方と NMT. 第 29 回 JTF 翻訳祭, 10 2019.
- [7] Kishore Papineni, Salim Roukos, Todd Ward, and Wei jing Zhu. Bleu: a method for automatic evaluation of machine translation. pp. 311-318, 2002.
- [8] G. DODDINGTON. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. Proceedings of the HLT Conference, San Diego, California, 2002, pp. 138-145, 2002.



5

産業日本語関連

