

2022年度（令和4年度） 卒業論文

JParaCrawlを用いた  
単文翻訳におけるドメイン適応

指導教員  
村上仁一

鳥取大学工学部 電気情報系学科  
自然言語処理研究室  
B19T2076C 名村 太一

## 概要

NMTにおいて様々なドメイン適応 (domain adaptation) の手法が研究されている [1]. また, ドメイン適応のためには大規模なコーパスが必要である. そこで大規模コーパスとして JParaCrawl[2] が開発された. しかし, まだ人手の翻訳精度に及ばない. 本研究ではファインチューニングに用いるコーパスが小さいことが原因だと考える. そこで JParaCrawl からテスト文に類似する対訳を抽出し, コーパスを拡張しファインチューニングを行う. 実験の結果, 対訳を追加しなかった場合と比較して, BLEU 値が 1.7 向上した.

# 目次

第1章	はじめに	1
第2章	背景	2
2.1	NMT	2
2.1.1	Transformer	2
2.2	JParaCrawl	3
2.3	ドメイン適応	3
2.3.1	データ選択 (類似文検索)	5
2.3.2	ファインチューニング	5
第3章	提案手法	7
第4章	実験	8
4.1	実験データ	8
4.2	実験設定	8
4.3	実験手順	9
4.4	BaseLine	9
第5章	実験結果	11
5.1	自動評価	11
5.2	人手評価	11
5.3	出力例	12
5.3.1	Proposed の良い例	12
5.3.2	Proposed の悪い例	13
第6章	考察	15
6.1	抽出するデータ量を変更	15
6.1.1	自動評価結果	16

6.1.2	Proposed と top1 の出力結果の違い . . . . .	17
6.2	ランダムに追加 . . . . .	17
6.3	抽出した対訳のみでファインチューニング . . . . .	19
6.4	JParaCrawl+単文対訳コーパスから抽出 . . . . .	19
6.5	Google 翻訳との比較 . . . . .	20
<b>第 7 章</b>	<b>先行研究</b>	<b>23</b>
<b>第 8 章</b>	<b>おわりに</b>	<b>24</b>
付 録 A	ファインチューニングの shell	25
付 録 B	TF-IDF を用いた類似文検索	26
付 録 C	BaseLine-Proposed の対比較評価結果	29
付 録 D	google-BaseLine の対比較評価結果	34
付 録 E	google-Proposed の対比較評価結果	39

# 目 次

2.1	Transformer model のアーキテクチャ Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).[3] より抜粋 . . . . .	4
2.2	ファインチューニングの図解 . . . . .	6
3.1	提案手法の図解 . . . . .	7
4.1	BaseLine の図解 . . . . .	10
6.1	ランダムに追加された特殊な例 . . . . .	18

# 表 目 次

4.1.1 実験データ	8
4.3.1 抽出された対訳の例	9
5.1.1 提案手法と BaseLine の BLEU 値	11
5.2.1 Proposed と BaseLine の対比較評価 (4 名)	11
5.3.1 Proposed の良い翻訳例と, 追加された文 1	12
5.3.2 Proposed の良い翻訳例と, 追加された文 2	12
5.3.3 Proposed の良い翻訳例と, 追加された文 3	13
5.3.4 Proposed の悪い翻訳例と, 追加された文 1	13
5.3.5 Proposed の悪い翻訳例と, 追加された文 2	14
5.3.6 Proposed の悪い翻訳例と, 追加された文 3	14
6.1.1 抽出された対訳の例とその順位	16
6.1.2 BaseLine と top1, top100 の BLEU 値	16
6.1.3 top1, Proposed, BaseLine の翻訳を比較	17
6.2.1 ランダムに抽出された対訳の例	18
6.2.2 BaseLine と random に 1 万, 10 万, 100 万を追加して学習した BLEU 値	18
6.3.1 JParaCrawl からテスト文に類似する文を抽出したコーパスのみでファインチューニングを行ったモデルの BLEU 値	19
6.4.1 単文対訳コーパス+JParaCrawl からテスト文に類似する文を抽出したコーパスでファインチューニングを行ったモデルの BLEU 値	19
6.5.1 Google 翻訳と BLEU 値で比較	20
6.5.2 Proposed と Google 翻訳の対比較評価 (4 名)	20
6.5.3 BaseLine と Google 翻訳の対比較評価 (4 名)	20
6.5.4 Proposed と Google 翻訳を比較して, Proposed が良かった例	21
6.5.5 Proposed と Google 翻訳を比較して, Google が良かった例	21

7.0.1 先行研究との違い . . . . .	23
--------------------------	----

# 第1章 はじめに

NMTにおいてドメイン適応についての様々な研究がされてきた [1]. また, ドメイン適応を行うためには十分なコーパスが必要である. しかし, 日英対訳コーパスにおいては一般的なドメインの大規模コーパスが無かった. そこでウェブから大規模かつ特定のドメインに特化しない日英対訳コーパスとして JParaCrawl[2] が開発された. しかし, まだ人手の翻訳精度に及ばない.

本研究ではファインチューニングに用いるコーパスが小さいことが原因だと考える. そこで, ファインチューニングに用いるコーパスを拡張することで翻訳精度の向上を図る. まず, 目的ドメインのテスト文に類似する対訳を JParaCrawl から抽出する. 次に, 抽出したコーパスと目的ドメインのコーパスを合わせた新たなコーパスを作成する. そして新たなコーパスを用いて JParaCrawl の事前学習済みモデルのファインチューニングを行う.

また, 目的ドメインのコーパスのみを用いて JParaCrawl の事前学習済みモデルをファインチューニングしたものを BaseLine とする. BaseLine と提案手法の比較を, BLEU 値による自動評価と人手による対比較評価で行う.



# 第2章 背景

## 2.1 NMT

NMT(Neural Machine Translation)とはニューラルネットワークを用いた機械翻訳手法である。LSTM, CNN, Transformer など様々なアーキテクチャがある。本研究ではTransformerを用いる。

### 2.1.1 Transformer

Vasawaniらによって開発されたNMTアーキテクチャである[3]。概略としては、左側のエンコーダで入力文の意味を学習し、右側のデコーダで意味を受け取り出力文を生成する。

#### Attention

Attentionは以下の数式で計算される[3]。

$$\begin{aligned} \text{Attention}(Q, K, V) &= \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \\ &= \begin{cases} \text{softmax}\left(\frac{\mathbf{q}_1 K^T}{\sqrt{d_k}}\right)V \\ \vdots \\ \text{softmax}\left(\frac{\mathbf{q}_n K^T}{\sqrt{d_k}}\right)V \end{cases} \end{aligned}$$

$Q$ はquery,  $K$ はkey,  $V$ はValueを表し、それぞれ行列である。また $d_k$ はスケールリング因子である。

入力文の各単語の横ベクトルを $\mathbf{q}$ で表し、入力文を行列 $Q$ で表す $\mathbf{q}$ と行列 $K$ との内積を取ることで、各keyとの類似度を計算する。softmax関数で、1までの確率に変換し、その確率に見合ったValueを得る。

## Multi Head Attention

Multi Head Attention は以下の数式で計算される [3].

$$\begin{aligned} \text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1 \cdots \text{head}_h)W^O \\ \text{where head}_i &= \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \end{aligned}$$

Transformer では隠れ状態状態を 512 次元で設定されている. それをヘッド数 (8 に設定されている) に分割して Attention を計算して, concat で結合している.  $W$  を更新することで, 学習している.

## Encoder

入力文を Embedding し, Positional Encoding することで単語の位置を加味した単語ベクトルの列を生成する. Multi-Head Attention 層に同じ入力を渡すことで (Self-Attention), 入力文の単語同士の関係性などを学習する.

## Decoder

Encoder から  $K$  と  $V$  を受け取り, 直前まで出力した文  $Q$  を Multi-Head Attention 層に入力することで, 次の単語を予測する.

## 2.2 JParaCrawl

ParaCrawl[4] というヨーロッパ言語と英語の大規模対訳コーパスをウェブから継続的に作成するプロジェクトがある. 最も多い独英の対訳では 2 億 7 千万以上が用意されている.

NTT コミュニケーション科学基礎研究所によって作成された日英対訳コーパスである. 対訳文は約 2000 万文が用意されている.

## 2.3 ドメイン適応

ドメイン特有の言い回しや表現がある. そのため NMT において, 汎用的なシステムよりそのドメインに適応したシステムの方が翻訳精度が向上する [5]. ドメイン適応するために様々な手法が研究されているが, コーパス中心とした適応 (data centric) と, モデ

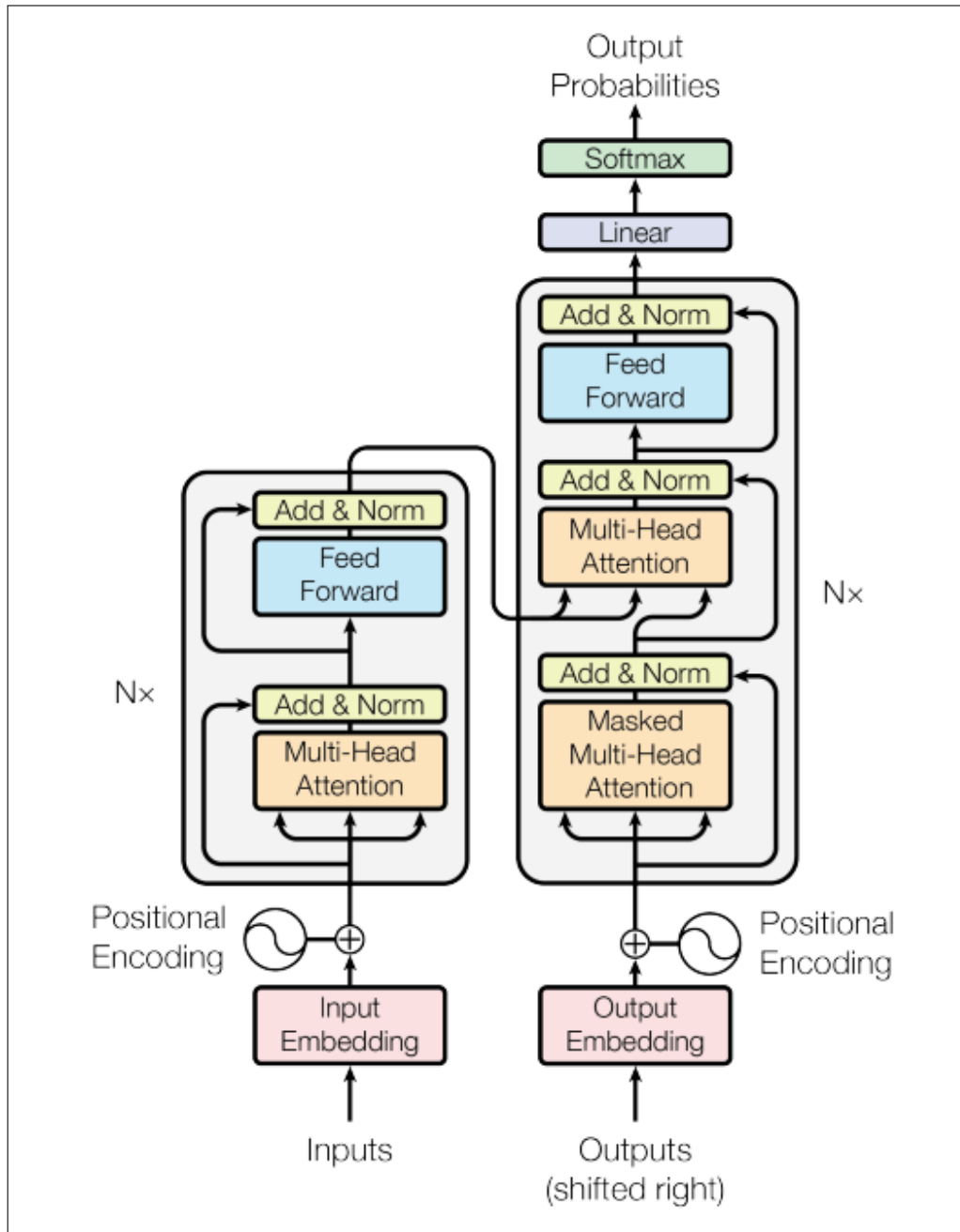


図 2.1: Transformer model のアーキテクチャ Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).[3] より抜粋

ル中心とした適応 (model centric) に大きく分けられる [1]. 本研究ではこれらを組み合わせてドメイン適応を行う.

### 2.3.1 データ選択 (類似文検索)

データ選択 (data selection) とはコーパスによるドメイン適応の手法である [1]. 目的ドメインのコーパスに類似する文を他のコーパスから抽出し, コーパスを拡張する. 様々なアルゴリズムが考案されているが [6], 本研究では TF-IDF を用いる.

#### 2.3.1.1 TF-IDF

TF-IDF とはある文書に含まれる単語がどのくらい重要かを表す. TF(Term Frequency) は「ある単語」が「ある文書内」でどのくらいの頻度で出現するかを表す. IDF(Inverse DocumentFrequency) は「全文書中」で「ある単語を含む文書」がどれくらい存在するかを表す. そしてこれらの積が TF-IDF で, 低頻出の単語と, ある単語がある文章に大量に含まれるとき, その単語を重要だと表す.

scikit-learn[7] を用いて計算し, それは以下の式で計算され,  $t$  が単語,  $d$  がある文書,  $n$  が全文書数,  $df(t)$  が単語  $t$  が含まれる文書数を表す.

$$\begin{aligned}tf-idf(t, d) &= tf(t, d) \times idf(t) \\tf(t, d) &= f_{t,d} \\idf(t) &= \log \frac{1+n}{1+df(t)} + 1\end{aligned}$$

TF-IDF で単語を数値化し, 固定次元 (次元=辞書の数) の文ベクトルを作成する. その文ベクトル同士のコサイン類似度を求めることで, 類似文を求める.

### 2.3.2 ファインチューニング

ファインチューニング (fine-tuning) とはドメイン適応の手法である [8][1]. 事前学習済みモデルを目的ドメインのコーパスを用いてパラメータを微調整することである. これによってモデルをドメインに適応させる. 本研究では JParaCrawl のファインチューニングの方法に従って行う.

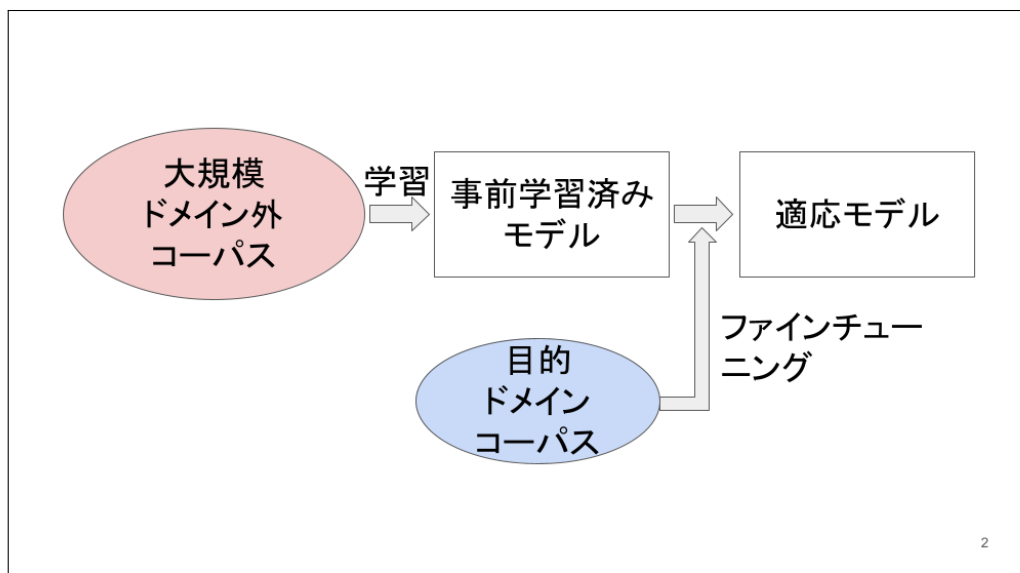


図 2.2: ファインチューニングの図解

## 第3章 提案手法

本研究では、以下の手順を提案する。

1. TF-IDF でテスト文に類似する対訳を JParaCrawl から抽出
2. 抽出した対訳を目的ドメインのコーパスに付け加えて、新たなコーパスを作成
3. 新たなコーパスを用いて、JParaCrawl で学習された事前学習済みモデルをファインチューニング

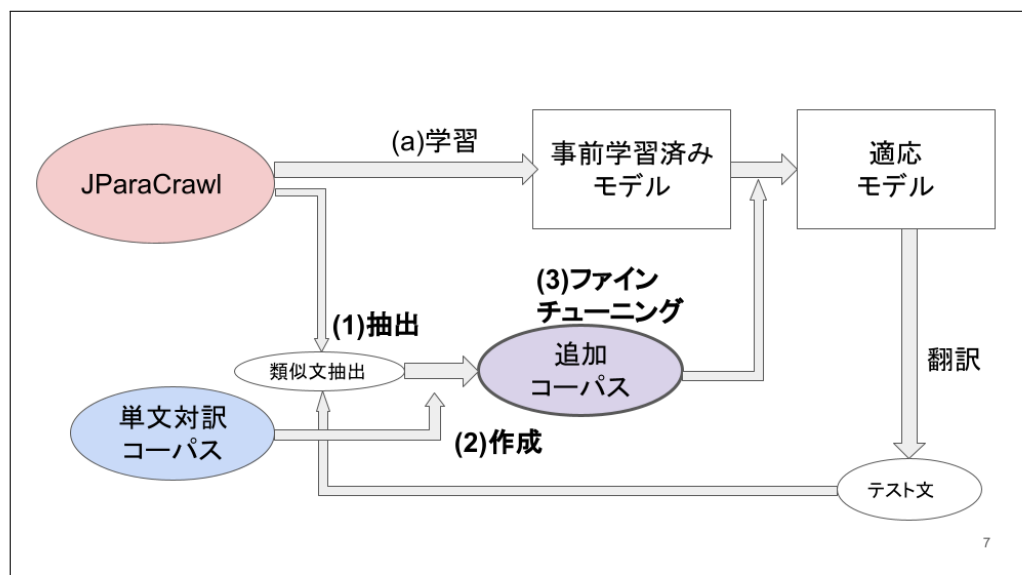


図 3.1: 提案手法の図解

## 第4章 実験

### 4.1 実験データ

実験に用いるデータを表 4.1.1 に表す. 単文対訳コーパスとテスト文は電子辞書を中心に採取された単文を用いる [9]. また, 本実験では JParaCrawl の対訳の内, Bicleaner の値を 0.7 以上の対訳のみを使用する.

表 4.1.1: 実験データ

JParaCrawl ver 3.0 (Bicleaner 0.7 以上)	18,450,971
単文対訳コーパス	163,188
テスト文	16,328

### 4.2 実験設定

前処理には sentencepiece[10] を用いて, 語彙数 32,000 とし, サブワード単位で分割する. NMT の学習には FAIRSEQ[11] を用いて実験を行う. アーキテクチャは Transformer[3] である. 学習のパラメータは [2] に従って, 同様の設定で行う. 実験に用いたものをソースコード A.1 に示す. 具体的には, restore-file に事前学習済みモデルのパスを指定する. 最適化アルゴリズムには adam を用いる. 学習率スケジュールは inverse square root アルゴリズムを用いて, `-warmup-init-lr` から始まり, `-lr` まで線形で学習率を増加し, `-warmup-updates` ステップで `-stop-min-lr` に到達するように減少させる. 勾配は 1.0 を超えないようにクリップする. 過学習を防ぐために, `-weight-decay` と `-dropout` を設定する. 損失関数は label smoothing ありの cross entropy で計算する. ミニバッチは 5,000 トークンまでを扱う. 事前学習済みモデルが 24,000 ステップまで学習しているので, そこから 4,000 ステップでファインチューニングを行う. 100 バッチごとにモデルを保存し, 最後の 8 つのモデルのパラメータを平均した値を使用する.

## 4.3 実験手順

1. TF-IDF でテスト文に類似する文を JParaCrawl から抽出する
2. 単文対訳コーパスと抽出した文を合わせたコーパスを用いて, JParaCrawl で学習された事前学習済みモデルをファインチューニングする
3. テスト文の翻訳を行い, 評価を行う

抽出する量を, テスト文 1 文に対して類似した上位 10 文を Proposed として実験を行う. 抽出するプログラムをソースコード B.1 にしめす. 抽出したデータの例を表 4.3.1 に示す. 入力文の日本語文を使用してこのような結果が得られた. この結果から, あまり入力文に類似してないことがわかる.

表 4.3.1: 抽出された対訳の例

	日本語文	英語文
入力文	「コノサー」誌の編集者である。	She is an editor at Connoisseur magazine .
類似文	サーバーの設定をサーバーの編集アイコンをクリックして編集することができます。 第 3 条 (学術誌編集委員会) 学術誌の編集と発行の業務を遂行するために学術誌編集委員会を置く。  しかし、それ以来「コノン」が人気になった以来、それはもっと正確になるでしょう。	You can edit a server’s settings by clicking the server’s Edit icon.  Article 3 (Editorial Committee) The Editorial Committee shall be installed to perform duties of editing and publication of the journals.  But it will be more accurate to say that since then ”Konon” began to become popular.

評価には SacreBLEU[12] を用いて, BLEU[13] を計算する. また, 人手による対比較評価も行う.

## 4.4 BaseLine

本実験では, BaseLine を単文対訳コーパスのみを用いて事前学習済みモデルをファインチューニングした結果とする. また, 以下に BaseLine の図を図 4.1 に示す.



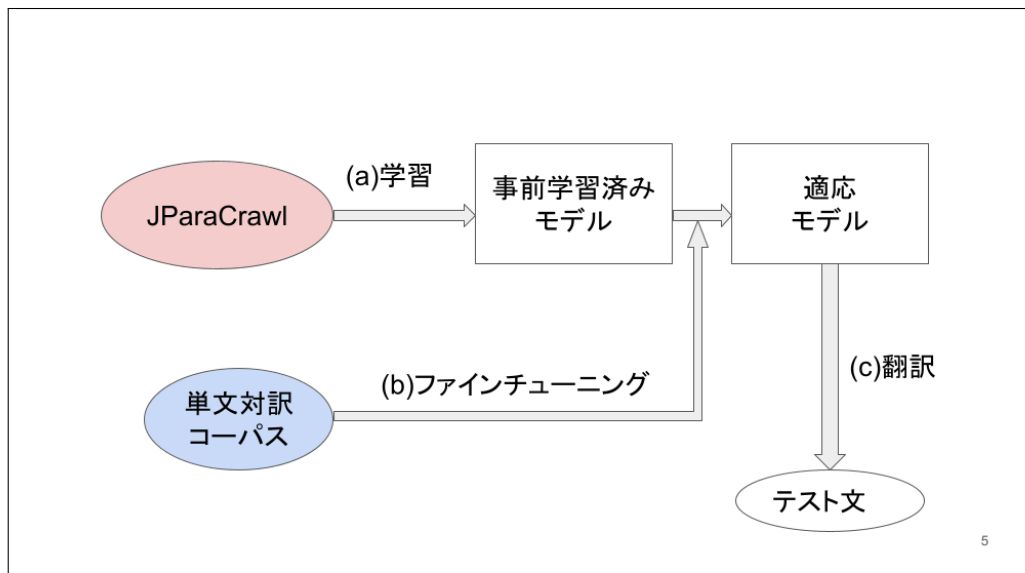


図 4.1: BaseLine の図解

## 第5章 実験結果

### 5.1 自動評価

表 5.1.1 に BLEU 値の評価結果を示す.

表 5.1.1: 提案手法と BaseLine の BLEU 値

model	コーパス量	BLEU
Proposed	324,233	<b>26.2</b>
BaseLine	163,188	24.5

表 5.1.1 より, BLEU 値において Proposed は BaseLine より向上することが確認できた.

### 5.2 人手評価

人手評価は同研究室の学生 4 名で行った. そしてその平均値を人手評価とする.

Proposed と BaseLine の対比較評価を行う. テスト文の中からランダム 100 文に対して, Proposed の方が良い, BaseLine の方が良い, ほとんど差がないの 3 つの分類で対比較評価を行った. その結果を表 5.2.1 に示す.

表 5.2.1: Proposed と BaseLine の対比較評価 (4 名)

評価	割合 (%)
Proposed	28.5
BaseLine	16.8
差がない	54.8

その結果, BaseLine の方が良いが 16.8% に対して, Proposed の方が良いが 28.5% で上回った. 人手評価でも Proposed が BaseLine を上回る結果になった.

## 5.3 出力例

### 5.3.1 Proposed の良い例

BaseLine の翻訳結果と比較して Proposed の方が良い文と追加された対訳を表 5.3.1, 表 5.3.2, 表 5.3.3 に示す.

表 5.3.1: Proposed の良い翻訳例と, 追加された文 1  
文

入力文	知床では自然破壊が進んでいる.
Proposed	Natural destruction is progressing in Shiretoko .
BaseLine	In the intellectual world , destruction of nature is increasing .
参照文	The destruction of the natural environment around Shiretoko has become serious .
類似文 1 日本語文	知床自然の宝庫知床は, 生物と自然が共有した世界を見ることができます.
類似文 1 英語文	Shiretoko Treasury of nature Shiretoko can see the world shared by living things and nature.
類似文 2 日本語文	冬の知床世界自然遺産として知られる知床.
類似文 2 英語文	Winter's Shiretoko Shiretoko known as a natural world heritage site.

表 5.3.2: Proposed の良い翻訳例と, 追加された文 2  
文

入力文	私は電車事故で足留めを食った。
Proposed	I got stuck in a train accident .
BaseLine	I was kept indoors in a train accident .
参照文	I was stranded as a result of the train accident .
類似文 1 日本語文	負傷者は病院へ運ばれ、警察が聞き取り調書をとっているの でこの駅 (Bardawan?) で 3 時間も足留めされていたのだ。
類似文 1 英語文	The wounded were taken to the hospital & the police were fil- ing their report, thus we were delayed at that station ('Bar- dawan'?) for three hours.
類似文 2 日本語文	しかし、その代償として、学生や船員などの在外キリバス人 のほとんどが足留めを食らい、帰国できなくなった。
類似文 2 英語文	But the price has been that most of their own nationals stranded overseas could not get home, including students and sailors.

表 5.3.3: Proposed の良い翻訳例と、追加された文 3

文	
入力文	彼女は思いがけない質問にまごついたようだった。
Proposed	She seemed to be confused by the unexpected question .
BaseLine	She seemed to have gotten confused with the unexpected question .
参照文	She seemed to be embarrassed at the unexpected question .
類似文 1 日本語文	実に奇妙でしたよ」最初の数週間は、前の仕事と同じ作業環境で次の仕事をすることにまごついたとポールは語る。
類似文 1 英語文	Being in the same working environment in his next job was a bit disorienting for the first few weeks, Paul says.
類似文 2 日本語文	ですから、イエスがヨルダン川に来て、洗礼者ヨハネからバプテスマを受けようとしたとき、ヨハネはまごついたのです。
類似文 2 英語文	This is why John the Baptist was so perplexed when Jesus came to the Jordan River, asking to be baptized by him.

表 5.3.1 より、BaseLine では「知床」が翻訳されなかったのに対して、Proposed では「知床」が含まれる文が追加されることによって、正しく翻訳されるようになった。

表 5.3.2, 表 5.3.3 の結果では、直接対応している対訳は無かったが良い翻訳が出力された。

### 5.3.2 Proposed の悪い例

BaseLine の翻訳結果と比較して Proposed の方が悪かった翻訳と、追加された文を表 5.3.4, 表 5.3.5, 表 5.3.6 に示す。

表 5.3.4: Proposed の悪い翻訳例と、追加された文 1

文	
入力文	その事件はうやむやに葬られた。
Proposed	The case was buried peacefully .
BaseLine	The matter was hushed up .
参照文	The matter has been hushed up .
類似文 1 日本語文	遺体は後に、裸にされて傷つけられた状態で発見され、ブロードフットの遺体とともにラグマーニーにある私の兄の要塞の近くに葬られた。
類似文 1 英語文	The body was afterwards found, naked and mutilated, and, with Broadfoot's, was buried near my brother's fort Lughmaunee.

表 5.3.5: Proposed の悪い翻訳例と、追加された文 2

文	
入力文	彼女は我を通した。
Proposed	She passed herself through .
BaseLine	She had her own way .
参照文	She had her own way .
類似文 1 日本語文	そして事業を通した社会活動を旨としています。
類似文 1 英語文	And the purpose is social activities through business.
類似文 2 日本語文	その他報道発表, インタビュー, 社説及び新聞記事等を利用した, マスメディアを通したコミュニケーションも可能である。
類似文 2 英語文	It is also possible to 2937 communicate through the media using press releases, interviews, editorials and articles.

表 5.3.6: Proposed の悪い翻訳例と、追加された文 3

文	
入力文	しばしば寒暖計が零度を下回る。
Proposed	The mercury often falls below zero .
BaseLine	The thermometer is often below zero .
参照文	The thermometer often goes below zero .
類似文 1 日本語文	水銀温度計よりも安全であることに加えて、デジタル温度計はデジタル技術を使用して正確な読み取りを提供しました。
類似文 1 英語文	Apart from being safer than mercury thermometer, digital thermometers provided accurate reading using digital technology.
類似文 2 日本語文	これを行うには、容器内に温水をダイヤルし、従来の水銀温度計によって最初に温度を測定し、次に電子を測定する必要があります。
類似文 2 英語文	To do this, you need to dial warm water into the container and measure its temperature first by a conventional mercury thermometer, and then electronic.

表 5.3.4 より、「葬られた」の翻訳は「hushed up」が正しいが、「葬られた」と「buried」を含む対訳を学習したため、間違った翻訳をしたと考えられる。

また、表 table:bad-Proposed2, 表 table:bad-Proposed3 の結果でも、誤った対応が学習されたため、間違った翻訳をしたと考えられる。

## 第6章 考察

### 6.1 抽出するデータ量を変更

この節では JParaCrawl から抽出するデータ量を変更する。Proposed ではテスト文 1 文に対して類似する上位 10 文を抽出していた。ここではテスト文 1 文に対して類似する上位 1 文と 100 文としたものを top1 と top100 として実験を行う。類似文として抽出された、例として上位何文目かを添えて表 6.1.1 に示す。

表 6.1.1: 抽出された対訳の例とその順位

	日本語文	英語文
入力文	100万円を資本に商売を始めた。	He started his business with a capital of a million yen .
第1位	資本金10百万円6.	Capital : 10 million yen 6.
第10位	私はこの商売を始めたばかりだから、良いものを作らないと商売は長く続かない。」	I just started in this business, and I have to make a good product.
第50位	給与所得控除を一律で10万円引き下げ、さらに控除額の上限もいまの「年収1千万円以上で年220万円」から「800万円以上で年190万円」に引き下げますが、すべての納税者が受けられる基礎控除は、38万円から48万円に引き上げられます。商売心でそうしたでしょうか?(違います)。	Reduce the salary income deduction by 100,000 yen uniformly and further reduce the upper limit of the deduction amount now to "2.2 million yen a year with annual income of 10 million yen or more" to "190 million yen / year over 8 million yen", but all The basic deduction that taxpayers can receive will be raised from 380,000 yen to 480,000 yen. (No) For what did I do so?
第100位	(規定では明文化されなくなりましたが、実際には500万円以上は判断基準として残っています)500万円以上の資本金を用意したとしても、「その資本金をどのように拠出したか」を証明しなければいけません。	(Although it is no longer clearly stated in the regulations, in reality, 5 million yen or more remains as a criterion) Even if you prepare a capital of 5 million yen or more, you must prove "how you contributed that capital" .

### 6.1.1 自動評価結果

表 6.1.2: BaseLine と top1, top100 の BLEU 値

model	コーパス量	BLEU
top1	179,501	25.0
top100	1,559,702	24.0
BaseLine	163,188	24.5

表 6.1.2 より, top1 では BaseLine より BLEU 値が向上した. Proposed の方が top1 と比べて BLEU 値が高かった. また, top100 では BaseLine より BLEU 値が低下した.

## 6.1.2 Proposed と top1 の出力結果の違い

top1 の結果から BaseLine よりも BLEU 値が高いため, Proposed と比較したものを表 6.1.3 に示す.

表 6.1.3: top1, Proposed, BaseLine の翻訳を比較

文	
入力文	私は酒で頭がぼんやりしてしまった。
top1	My brain has been muddled by alcohol .
Proposed	My head went blank from drinking .
BaseLine	My brain has gotten muddled with alcohol .
参照文	The spirits muddled my brain .
Proposed 類似文 日本語	持続時間も長く、頭が真っ白になってしまうという声も。
Proposed 類似文 英語	Some people say that it lasts for a long time and that it makes them go blank.

この結果では, BaseLine より top1 の方がよくなった. しかし, JParaCrawl より抽出する量を増やした Proposed はかえって悪くなった. この原因として, top1 には無かったが Proposed には含まれていた類似文が影響したと考える.

## 6.2 ランダムに追加

この節では JParaCrawl からランダムに抽出した対訳を追加し, 学習を行う. ランダムに抽出した例を表 6.2.1 に示す. それぞれ 1 万, 10 万, 100 万の対訳を抽出し, 追加したコーパスで学習を行った. その結果を表 6.2.2 に示す. また, 追加された特殊な例を図 6.1 に示す.



表 6.2.1: ランダムに抽出された対訳の例

日本語文	英語文
『スーパープレミアムソフト W バニラリッチ』のマンハイムでの上演も、コンビニに象徴される現代日本のカリカチュアを、ヨーロッパの観客にアイロニカルにぶつけていた。	"Super Premium Soft W Vanilla Rich" performance in Mannheim also hit European audiences with the modern Japanese caricature symbolized by convenience stores.
私たちプラン・ジャパンとしても、パナソニックさんとの連携により、活動地域の状況をさらに改善していきたいと思しますので、何かあればまずご相談いただければと思います。	I Through our partnership with Panasonic, we at Plan Japan would like to make further improvements in areas where we are working.I hope that you'll contact us when anything comes up.
人々は指差し合って雪崩の複雑な原因について議論し、なぜもっと防御策が講じられていなかったのかと問い続けた。	People pointed fingers, arguing about the complex causes of the avalanche and asking why people hadn ' t been better-protected.

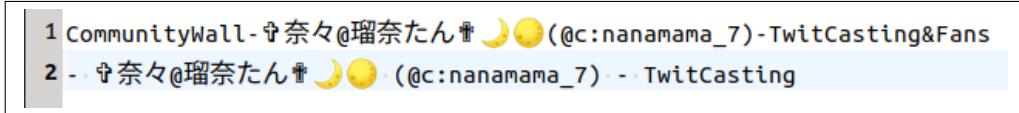


図 6.1: ランダムに追加された特殊な例

表 6.2.2: BaseLine と random に 1 万, 10 万, 100 万を追加して学習した BLEU 値

model	コーパス量	BLEU
random 10,000	173,188	24.7
random 100,000	263,188	25.7
random 1,000,000	1,163,188	23.6
BaseLine	163,188	24.5

random 100,000 が最も BLEU 値が高い結果だった。しかし、類似文を追加した Proposed の方が BLEU 値が向上した。

## 6.3 抽出した対訳のみでファインチューニング

この節では単文対訳コーパスを使用せずに，JParaCrawlより抽出したコーパスのみでファインチューニングを行う．抽出する量をテスト文1文に対して1文，10文，100文とし，それぞれJPara top1，JPara top10，JPara top100とする．その結果を表6.3.1に示す．

表 6.3.1: JParaCrawlからテスト文に類似する文を抽出したコーパスのみでファインチューニングを行ったモデルの BLEU 値

model	コーパス量	BLEU
JPara top1	16,313	14.6
JPara top10	161,045	17.8
JPara top100	1,396,514	20.5
BaseLine	163,188	24.5

表 6.3.1 より，全ての結果で BaseLine を下回ったが，類似する文を増やすと BLEU 値が向上した．一方表 6.1.2 では，top100 において，コーパスを大きくしすぎると BLEU 値が低下した．しかし，この実験では BLEU 値が向上した．

## 6.4 JParaCrawl+単文対訳コーパスから抽出

テスト文により類似したコーパスでファインチューニングを行えば翻訳精度が向上すると考える．そのため，この節では単文対訳コーパスと JParaCrawl を合わせたコーパスからテスト文に類似する文を抽出する．抽出する量は，テスト文1文に対して1文，10文，100文とし，それぞれを JPara+単文 top1，JPara+単文 top10，JPara+単文 top100 とする．その結果を表 6.4.1 に示す．

表 6.4.1: 単文対訳コーパス+JParaCrawlからテスト文に類似する文を抽出したコーパスでファインチューニングを行ったモデルの BLEU 値

model	コーパス量	BLEU
JPara+単文 top1	16,130	16.9
JPara+単文 top10	155,781	21.2
JPara+単文 top100	1,279,893	22.8
BaseLine	163188	24.5

表 6.4.1 より，全ての結果で BaseLine を下回ったが，類似する文を増やすと BLEU 値が向上した．また，表 6.3.1 と比較して全体的に BLEU 値が向上した．

この結果より，テスト文に類似していることより一定量の単文対訳コーパスが重要であると考えられる．単文対訳コーパスは単文の短い文で構成されている．それに対して JParaCrawl の多くは長い文で構成される．この違いにより表 5.1.1 の結果が最も良かったと考える．

## 6.5 Google 翻訳との比較

Google 翻訳との比較を行う．表 6.5.1 で Google 翻訳と BLEU 値で比較する．また，表 6.5.2 に，Proposed と Google 翻訳の対比較評価の結果，表 6.5.3 に BaseLine と Google 翻訳との対比較評価の結果を示す．

表 6.5.1: Google 翻訳と BLEU 値で比較

model	BLEU
Proposed	26.2
Google	26.9
BaseLine	24.5

表 6.5.2: Proposed と Google 翻訳の対比較評価 (4 名)

評価	割合 (%)
Proposed	13.3
Google	33.0
差がない	53.8

表 6.5.3: BaseLine と Google 翻訳の対比較評価 (4 名)

評価	割合 (%)
BaseLine	16.3
Google	39.8
差がない	44.0

表 6.5.4, 6.5.5 に Google 翻訳と Proposed の例を示す．

表 6.5.4: Proposed と Google 翻訳を比較して, Proposed が良かった例

文	
入力文	雨で試合が流れた。
Proposed	The game was rained out .
Google	The game was washed away by the rain.
参照文	The game was rained out .

文	
入力文	交渉は穏やかに行われた。
Proposed	The negotiations were conducted peacefully .
Google	Negotiations were peaceful.
参照文	The negotiations were conducted in a peaceful atmosphere .

文	
入力文	どうぞお楽にしてください。
Proposed	Please make yourself at home .
Google	Please enjoy yourself.
参照文	Please make yourself at home .

表 6.5.5: Proposed と Google 翻訳を比較して, Google が良かった例

文	
入力文	彼が得た報酬は微々たるものだった。
Proposed	His remuneration he received was substantive .
Google	His reward was meager.
参照文	He only got a small reward .

文	
入力文	信長勢は再び京へ上った。
Proposed	The Nobunaga went up to Kyoto again .
Google	Nobunaga's army went up to Kyoto again.
参照文	Nobunaga's army went up to Kyoto again .

文	
入力文	彼女は直感的に物事をとらえる。
Proposed	She catches things intuitively .
Google	She perceives things intuitively.
参照文	She perceives things intuitively .

表 6.5.1 より, BLEU 値において, Proposed の値は Google 翻訳の値に近かった. しかし, 表 6.5.2 の結果ではまだ及ばなかった.

## 第7章 先行研究

Poncelas らは、データ選択アルゴリズムとして TF-IDF, INR, FDA の3つのアルゴリズムを用いて、テスト文に類似する対訳を抽出しファインチューニングに用いている [6]。これらのアルゴリズムを用いて以下の3つの条件でファインチューニング用のデータを作成している。

1. 大規模コーパスから抽出したデータ
2. 目的ドメインコーパスから抽出したデータ
3. 大規模コーパスと目的ドメインコーパスを合わせたコーパスから抽出したデータ

本研究との違いを表 7.0.1 に示す。

表 7.0.1: 先行研究との違い

	先行研究	本研究
モデル	LSTM	Transformer
言語	独英	日英
ドメイン	健康, ニュース	単文
最も良かった結果	3の合わせたコーパスからテスト文に類似したデータでのファインチューニング	Proposed の目的ドメインコーパスと同量の類似文

先行研究では、3の結果が最も良かった。この実験は本研究の6.4節の実験設定と類似していると考えられる。しかし本研究ではいずれの結果も BaseLine より悪くなった。

この理由は、本研究のドメインが単文であることが原因だと考える。単文翻訳では文長が重要な情報であるが、JParaCrawl の多くは長い文で構成されている。そのため、テスト文に類似していることよりも一定量の単文対訳コーパスが含まれている Proposed の結果が最も良かったと考える。

## 第8章 おわりに

本研究では、TF-IDF を用いて JParaCrawl からテスト文に類似する対訳文を抽出した。その対訳文を目的ドメインのコーパスに追加することでコーパスを拡張した。拡張したコーパスを用いて事前学習済みモデルのファインチューニングを行うことで BaseLine と比較して BLEU 値と人手評価の両方で翻訳精度の向上が見られた。具体的には BLEU で 1.7 向上した。

コーパスを拡張しすぎると翻訳精度の低下が確認されたので、適切なコーパスの量や、他の対訳の抽出法を今後の課題にする。

# 付録A ファインチューニングのshell

ソースコード A.1: ファインチューニングの shell

---

```
1 CUDA_VISIBLE_DEVICES=0 fairseq-train $DATA_DIR \  
2     --restore-file $PRETRAINED_MODEL_FILE \  
3     --arch transformer \  
4     --optimizer adam \  
5     --adam-betas '(0.9, 0.98)' \  
6     --clip-norm 1.0 \  
7     --lr-scheduler inverse_sqrt \  
8     --warmup-init-lr 1e-07 \  
9     --warmup-updates 4000 \  
10    --lr 0.001 \  
11    --stop-min-lr 1e-09 \  
12    --dropout 0.3 \  
13    --weight-decay 0.0 \  
14    --criterion label_smoothed_cross_entropy \  
15    --label-smoothing 0.1 \  
16    --max-tokens 5000 \  
17    --max-update 28000 \  
18    --save-dir $MODEL_DIR \  
19    --no-epoch-checkpoints \  
20    --save-interval 1000000000 \  
21    --validate-interval 1000000000 \  
22    --save-interval-updates 100 \  
23    --keep-interval-updates 8 \  
24    --log-format simple \  
25    --log-interval 5 \  
26    --ddp-backend no_c10d \  
27    --update-freq 16 \  
28    --fp16 \  
29    --seed $SEED \  
30    | tee -a ./log/$EXP_NAME.$(date +%Y%m%d').log
```

---



## 付録B TF-IDF を用いた類似文検索

ソースコード B.1: TF-IDF を用いた類似文検索のプログラム

---

```
1 #TF-IDF で類文検索を行う。
2 import numpy as np
3 from sklearn.feature_extraction.text import TfidfTransformer
4 from sklearn.feature_extraction.text import CountVectorizer
5 from sklearn.metrics.pairwise import cosine_similarity
6
7 import pandas as pd
8 import tqdm
9 import torch
10 import pickle
11
12 from IPython.display import display
13
14 def main():
15     #データ読み込み
16     test_100 = []
17     with open("../Data/Kenkyu/test.MeCab.ja") as f:
18         test_100 = f.readlines()
19     src_test = []
20     tgt_test = []
21     with open("../Data/JPara+Kenkyu/JPara+Kenkyu.ja") as f:
22         src_test = f.readlines()
23     with open("../Data/JPara+Kenkyu/JPara+Kenkyu.en") as f:
24         tgt_test = f.readlines()
25
26     #TFIDf
27     #vectorizer の生成。token_pattern=u'(?u)\b\w+\b' で 1文字の語を含む設定
28     #出現回数が 2回以下の単語は排除
29     vectorizer = CountVectorizer(token_pattern=u'(?u)\b\w+\b', min_df
30                                 =2)
31     #transformer の生成。TF-IDF を使用
32     transformer = TfidfTransformer()
33     #ベクトル化
34     src_test_tf = vectorizer.fit_transform(src_test)
35     #TF-IDF
36     src_test_tfidf = transformer.fit_transform(src_test_tf)
```

```

36
37 cs_list = []
38 topK = 1000
39 #1分ずつ処理
40 for text in tqdm.tqdm(test_100):
41     #ベクトル化
42     test_tf = vectorizer.transform([text])
43     #TF-IDF
44     test_tfidf = transformer.transform(test_tf)
45     #コサイン類似度
46     cs = cosine_similarity(test_tfidf, src_test_tfidf)
47
48     cs_tensor = torch.from_numpy(cs.astype(np.float32)).clone()
49     sim, idx = torch.topk(cs_tensor, topK, dim=1)
50     sim_np = sim.to("cpu").detach().numpy().copy()
51     idx_np = idx.to("cpu").detach().numpy().copy()
52
53     cs_list.append(idx_np)
54
55 cs = cs_list
56 #途中保存
57 with open("cs.from_test.bin", "wb") as p:
58     pickle.dump(cs, p)
59
60 df = pd.DataFrame()
61 from_idx_list = []
62 sim_idx_list = []
63 out_en = []
64 out_ja = []
65 #日本語と英語文を取り出す
66 for en, i in enumerate(idx_np):
67     from_idx_list.extend([en] * topK)
68     sim_idx_list.extend(idx_np[en][0])
69     for j in i[0][:topK]:
70         out_ja.append(src_test[j])
71         out_en.append(tgt_test[j])
72
73 df["index"] = from_idx_list
74 df["ja"] = out_ja
75 df["en"] = out_en
76 #重複を削除
77 df_drop = df.drop_duplicates(subset="ja")
78
79 #上位N件を取り出す
80 top(df=df_drop, source=test_100, from_="from_test", topN=1)
81 top(df=df_drop, source=test_100, from_="from_test", topN=10)

```

```

82     top(df=df_drop, source=test_100, from_="from_test", topN=100)
83
84 def top(df, source, from_, topN):
85     """
86     df : 重複を削除したDataFrame
87     source : 検索文
88     frm : 検索文のもと
89     topN : 上位N件
90     """
91     df_out = pd.DataFrame()
92     #上位N件を取り出す
93     for i in range(len(source)):
94         df_out = pd.concat([df_out, df[df["index"] == i][:topN]])
95
96     out_en = df_out["en"]
97     out_ja = df_out["ja"]
98     #保存
99     with open("./" + from_ + ".top" + str(topN) + ".en", "w") as f:
100         f.write(out_en)
101     with open("./" + from_ + ".top" + str(topN) + ".ja", "w") as f:
102         f.write(out_ja)
103
104 if __name__ == "__main__":
105     main()

```

---

# 付 録 C    BaseLine-Proposed の対比較 評価結果

number	input	reference	A	B	評価者A	評価者B	評価者C	評価者D
0	信号が青より赤に変わった。 彼は仕事で京都へ行った。	The signal changed from green to red. He went to Kyoto on business.	The traffic light changed from green to red. He went to Kyoto on business.	The traffic light changed from green to red. He went to Kyoto on business.	o o	X X	X X	X X
1	政府はこの事業の具体的な内容を示した。 信長勢は再び京へ上った。	The government presented some of the features of the project that are concrete. Nobunaga's army went up to Kyoto again.	The government presented the specifics for the project. The Nobunagas went up to Tokyo again.	The government laid out the concrete details of the project. The Nobunaga went up to Kyoto again.	BaseLine top10	BaseLine top10	top10 top10	BaseLine top10
2	私は電車事故で足留めを食った。 彼女は私を通した。	I was stranded as a result of the train accident. She had her own way.	I was kept indoors by a train accident. She passed herself through.	I got stuck in a train accident. She had her own way.	top10 top10	top10 top10	top10 top10	top10 top10
3	ぶらんこが揺れている。 エイズの梅毒など治療法はまだわかっていない。	The swing is swinging. No sure cures for AIDS are known yet.	The swings are swinging. No sure cure for AIDS has yet been found for sure.	The swing is swinging. No sure cure for AIDS is yet known.	BaseLine o	BaseLine top10	BaseLine X	BaseLine top10
4	彼は有罪の宣告を受けた。 その悪はなかなか閉まらない。	He was convicted. The window won't shut.	He was found guilty. The window will not close.	He was declared guilty. That window will not fasten.	top10 o	top10 X	top10 BaseLine X	top10 X
5	その新人女優は体当たりの演技で新人賞を獲得した。 捕虜たちが逃亡を図った。	That started put everything she had into her part and won the prize for new talent. The captives attempted to make their escape.	The new actress won Rookie-of-the-Year honors for her great performance. The prisoners of war tried to escape.	The new actress won the rookie prize for her personal performance. The prisoners made their escape.	o X	o X	X top10	X top10
6	彼女は思いがけない質問にまごついたようだった。 知床では自然破壊が進んでいる。	She seemed to be embarrassed at the unexpected question. The destruction of the natural environment around Shiretoko has become serious.	She seemed to be confused by the unexpected question. Nature is in an advanced stage of destruction in the chimney.	She seemed to have gotten confused with the unexpected question. Nature is destroyed in Shiretoko.	top10 top10	X top10	top10 top10	top10 top10
7	仕事はどうにかうまくか仕上がった。 ガラスは割れやすい。	The work got finished somehow. Glass breaks easily.	The work was finished somehow or other. Glass is liable to break.	I managed to finish the job somehow or other. Glass is liable to break.	top10 o	X X	X X	X X
8	富士山は昔からたくさんのお絵描きに描かれた山だ。 決まり次第お知らせします。	Mount Fuji has been painted by numerous artists since ancient times. I shall let you know as soon as it is decided.	Mount Fuji has been painted in numerous ways since ancient times. I will let you know as soon as fixed.	Mount Fuji has long been painted in various pictures. I'll let you know as soon as it is decided.	BaseLine X o	X X	top10 top10	top10 top10
9	沖の方に船の明かりが見える。 交渉は絶やかに行われた。	You can see ships' lights toward the offing. The negotiations were conducted in a peaceful atmosphere.	I see the lights of the boats offshore. The negotiations were conducted peacefully.	We can see the lights of the ship offshore. The negotiations were carried out calmly.	top10 top10	top10 X	X X	top10 X
10	負債総額は10億円に及ぶ。 彼女は直感的に勘事をとらえる。	The deficit comes to a billion yen. She perceives things intuitively.	The debts amount to a billion yen. She gets at things intuitively.	The total debt amounts to a billion yen. She catches things intuitively.	o X	X top10	X X	X X
11	その後、多くの手筈や一般投資家が参加した。 今日はデパートでたくさん買い物をした。	After that, many speculators and general investors got in on the act. Today, I did a lot of shopping at the department store.	The meeting was followed by a large number of speculators and general investors. I bought a lot of things at a department store today.	Many attended the meeting after that, stockholders and general investors. I did a lot of shopping at a department store today.	top10 o	top10 X	BaseLine X	BaseLine top10
12	昨日のパーティーは、あまり盛り上がりなかった。 今回の法案には無理がある。	Yesterday's party was not very exciting. This bill is hard to pass.	Yesterday's party was not very lively. This bill is impossible.	There was not much excitement at yesterday's party. This bill is unreasonable.	o top10	X top10	X top10	X X

20	どうぞお楽にしてください。	Please make yourself at home .	Make yourself comfortable .	Please make yourself at home .	top10	X	top10	top10
27	この文面は彼の誠実さを証明している。	This letter proves his sincerity .	This sentence exemplifies his sincerity .	This passage exemplifies his sincerity .	o	X	BaseLine	X
28	彼らはその考え方に迷わされた。	They were led astray by that way of thinking .	They were deceived by the idea .	The idea puzzled them .	o	BaseLine	BaseLine	BaseLine
29	彼は自己の権利を主張している。	He claims his own rights .	He asserts his own rights .	He insists upon his rights .	o	X	X	BaseLine
30	この山頂から望む富士はとても美しい。	The view of Mt .	Fuji from the top of this mountain is very beautiful .	Fuji from this mountain peak is very beautiful .	top10	X	X	top10
31	そのけんかは何年も糸を引いた。	The quarrel lasted for years .	The quarrel trailed away for years .	The quarrel ended for many years .	x	X	X	top10
32	これが私にとっては精いっぱいのところだ。	This is the utmost limit of my power .	This is the utmost for me .	This is the most I can do .	o	X	BaseLine	top10
33	彼は額の汗を拭いた。	He wiped the perspiration off his forehead .	He wiped the sweat from his forehead .	He wiped the sweat from his brow .	o	X	X	X
34	田園生活を好んだ。	I gained acquaintance with rural life .	I learned of country life .	I learned how to live in the country .	o	X	BaseLine	BaseLine
35	太陽が堂々と現われた。	The sun appeared majestically .	The sun came out in all its glory .	The sun appeared bright .	top10	X	BaseLine	BaseLine
36	その両国間には交戦状態が存在する。	A state of belligerency exists between the two nations .	There is war between the two countries .	There is a state of belligerency between the two countries .	top10	X	top10	top10
37	その店の前に駐車した。	She parked her car in front of the store .	I parked in front of the store .	We parked in front of the store .	o	X	X	X
38	彼女は栗色の髪をしている。	Her hair is a chestnut color .	She has chestnut-colored hair .	She has chestnut hair .	o	X	X	X
39	昔は大西洋横断に5日かかった。	The Atlantic crossing used to take five days .	It used to take five days to travel to the Atlantic .	In the past it used to take five days to cross the Atlantic Ocean .	BaseLine	BaseLine	BaseLine	BaseLine
40	子供たちの旅行の語は急ぐ間に集めた。	The children's description of the trip centered on the food .	The children's talk about traveling was centered on food .	Talks about the children's travel centered on food .	BaseLine	X	top10	X
41	音波で鼓膜が振動する。	The eardrums vibrate when they are struck by sound waves .	The eardrum vibrates in the sound waves .	The sound waves vibrate the eardrum .	o	X	X	X
42	彼の顔は赤くなくなった。	His face reddened .	A blush mounted to his face .	His face turned red .	o	X	top10	X
43	噴水は四方八方に水をまき散らしていた。	The fountain sprayed water in every direction .	The fountain was scattered in all directions .	The fountain was scattered in all directions .	o	X	X	X
44	彼女の心臓はどきどきしていた。	Her heart thudded .	Her heart was beating fast .	Her heart was thumping against her chest .	top10	X	top10	top10
45	列車は15分間隔で運行している。	Trains are running at 15 - minute intervals .	The trains run every fifteen minutes .	Trains run every fifteen minutes .	o	X	X	X
46	夜みんなメインロジに集まった。	In the evening we gathered in the main lodge .	Everybody gathered at the main lodge at night .	All the night they assembled at the main lodge .	top10	top10	top10	top10
47	その険しい山脈には木が1本も生えていなかった。	No trees grew on the steep mountainside .	The steep hillside lacked a single tree .	No trees grew on the steep hillside .	BaseLine	BaseLine	BaseLine	BaseLine
48	ふたたび長い休止があった。	There was another long pause .	We had a long rest again .	We had a long rest again .	o	X	X	X
49	ある有望な可能性が生じた。	A promising possibility has turned up .	A promising possibility arose .	A promising possibility has arisen .	o	X	X	X
50	もっと早めにより多くの資本を投入すべきだった。	We were not quick enough to inject more capital into the economy .	I should have invested more capital earlier .	We should have invested more capital earlier .	o	X	X	X
51	晩年ほどやがて送った。	He spent his last years on skid row .	We spent the last years in a street .	I spent my last years in town .	BaseLine	X	X	X
52	下水管はこれだけ大量の水を処理しきれない。	The sewers can't cope with so much water .	Sewers cannot process such a large amount of water .	Sewers cannot handle this large amount of water .	o	X	X	BaseLine
53	腹がゴロゴロ鳴った。	His stomach rumbled .	The bowels rumbled .	My stomach rumbled .	top10	X	top10	top10

54	しばしば雨量計が零度を下回る。	The thermometer often goes below zero .	The thermometer is often below zero .	BaseLine X	BaseLine
55	豊富な語彙をもっている。	He has a copious vocabulary .	He possesses a wide vocabulary .	o X	X top10
56	指輪はどこにも見つからなかった。	The ring was nowhere to be found .	The ring was nowhere to be found .	o X	X X
57	その塔は900フィート以上ある。	The tower stands over 900 feet high .	The tower stands over 900 feet .	o X	X X
58	傘が雨返しになった。	The umbrella turned inside out .	My umbrella was turned over .	top10 X	top10 X
59	人は誰も間違いをするものだ。	One is certain to make slips .	Everybody makes a mistake .	o X	top10 top10
60	7マイルは上陸間では歩ききれない距離だ。	Seven miles is a distance too great to walk in an hour .	7 miles are beyond walking in an hour .	top10 X	X X
61	ボタンは糸いれでしっかりと布地につけられた。	The button was anchored to the cloth with heavy thread .	The button was firmly draped on the cloth with thick thread .	BaseLine X	top10 top10
62	幸福で彼女の目は輝いた。	Happiness kindled her eyes .	Happiness kindled her eyes .	o X	BaseLine BaseLine
63	誰だって一人ではできない。	No one man can do it .	Nobody can do it by himself .	o top10	X top10
64	我々は特別番組を放送した。	We put on special programs .	We broadcast a special program .	o X	X X
65	彼はそこで戦闘中に殺された。	He was killed in action there .	He was killed there during the battle .	o	BaseLine BaseLine X
66	当社製品はよく売れる。	Our products are marketable .	Our product sells well .	o X	X X
67	我々の大学は長い輝かしい歴史を誇っている。	Our university can boast a long and glorious history .	Our university has a long , brilliant history .	top10 X	top10 X
68	彼は町の有力者を抱き込んだ。	He managed to secure the help of a town boss .	Our university boasts a long , glorious history .	X	X top10 X
69	彼の胸は興奮で躍動していた。	His heart was throbbing with excitement .	His heart was racing with excitement .	o X	X X
70	その事件はうやむやに葬られた。	The matter has been hushed up .	The matter was hushed up .	BaseLine	BaseLine BaseLine
71	革靴が上着を濡らした。	The wheels rubbed my coat .	The wheels robbed my coat .	BaseLine X	X X
72	どうぞ着席して下さい。	Please have a seat .	Please be seated .	BaseLine	BaseLine BaseLine X
73	英語をやり直さねばならない。	I must brush up my English .	I have to redo my English .	top10 X	top10 top10
74	お酒の量がわかりません。	I'm afraid your point escapes me .	I don't understand the gist of what you said .	o X	top10 X
75	彼はバイクを木にぶつけた。	He hit his bike against a tree .	He ran his motorbike into a tree .	o	BaseLine X
76	私は酒で頭がぼんやりしてしまいました。	The spirits muddled my brain .	My brain has gotten muddled with alcohol .	BaseLine	BaseLine BaseLine X
77	雨が試合が濡れた。	The game was rained out .	The game was washed out by the rain .	top10	top10 top10
78	そのポンプは蒸気で動く。	The pump is run by steam .	The pump is driven by steam .	o X	X X
79	彼女の声は震えた。	Her voice wavered .	Her voice vibrated .	o X	top10 X
80	いつかは当局が動き出すだろう。	Someday the authorities will go into action .	The authorities will take action some day .	o	top10 top10
81	時計がからちと時を刻んでいる。	The clock is ticking away the time .	The clock is ticking away the minutes .	o X	top10 BaseLine
82	お先に失礼します。	Excuse me , but I must be going now .	Please excuse me the first time .	o	top10 X

83	彼は冗談が過ぎた。	His joke went too far .	He carried a joke too far .	He passed the joke .	X	BaseLine	X	top10
84	値段は大き過ぎて。 彼は手紙を破り書きました。	The prices offer according to size . He scribbled his letters .	Prices vary according to the size . He scribbled a letter .	The prices vary according to the size .	O	X	X	X
85	彼が得た報酬は破々たるものだった。	He only got a small reward .	His remuneration he received was substantive .	He scribbled a letter .	O	X	X	X
86	その部屋はむんむんしていた。	It was stuffy in the room .	The room was stuffy .	He paid me a trifle .	top10	top10	top10	X
87	転職を考えています。	I'm thinking of changing my jobs .	I'm considering changing jobs .	The room was trembling .	top10	top10	top10	top10
88	基本的に、米国は日本に二つの行動を求めている。	Essentially, the United States is asking Japan to do two things .	Basically, the United States urges Japan to take two actions .	I'm thinking of changing jobs .	O	X	top10	X
89	ソマリアを他山の石とした旧ゴP K O関係者が武力行使に傾斜だったので、米国は不満だった。	The United States was dissatisfied with the United Nations since those involved in the peacekeeping operation in the former Yugoslavia were cautious about using armed forces because of their experience in Somalia .	The United States was dissatisfied because an official involved in the peacekeeping operations of Anglo-Soviet Yugoslavia, who was said to have been deliberate on the use of weapons .	Thus, the United States has asked Japan for two actions .	top10	top10	top10	top10
90	田中首相はオーストラリア大使館の元職員による公金流用疑惑について、杉浦正徳副大臣に再調査を指示した。	Foreign Minister Makiko Tanaka on Thursday instructed Senior Vice Foreign Minister Masataka Sugiyama to conduct a new probe into the alleged misappropriation of public funds by a former embassy official in Australia .	Foreign Minister Masaken Sugiyama was asked by Foreign Minister Masaken Sugiyama on suspicion of using public money for officials at the Japanese Embassy in Australia .	The United States was dissatisfied with the fact that former Yugoslavia involved in peacekeeping operations had been cautious about using force .	X	X	X	top10
91	韓国にもまた、一層の努力を求めたい。	We exhort South Korea to make stronger efforts as well .	We urge South Korea to make further efforts .	Tanaka reportedly instructed former members of the Australian Embassy to conduct a re-examination of suspicions using public money .	X	X	X	top10
92	12号線延伸は一九九二年に着工、来年十二月に開業を予定している。	Construction on the subway line began in 1992, and the line is scheduled to be put into operation in December next year .	Construction will begin on Track 12 in 1992 and will begin next December .	We urge South Korea to make greater efforts to this end .	top10	top10	top10	top10
93	従業員は約千人で、投資額は五千万米ドルに及ぶ。	The average monthly wage for a Chinese worker is about 1,000 yuan, or about 115 dollars .	With about 1,000 employees, the company plans to invest 50 million US dollars .	Construction will begin on Route 12, which is scheduled to begin in December next year .	X	BaseLine	top10	top10
94	北鋼会は一九九九年に荷の設立認可を受け、大阪厚生院と大和川病院を継ぎ。	They are Osaka Ensei Hospital in Higashi Sumiyoshi Ward, Osaka, and Yamatagawa Hospital in Kashihara, Osaka prefecture .	The Kitanishiki Kaikan was approved by the prefectural government in 1994 and operates the Osaka Ensei Hospital and the Daiwaigawa Hospital .	The company employs about \$50 million and employs about 1,000 people .	X	BaseLine	top10	X
95	この会社には人材が揃っている。	This company has a good staff of talents .	This company has a staff of people .	The Kitanishiki Kaikan was established in 1964 by the prefectural government and runs the Osaka Ensei Hospital and the Owakawa Hospital .	O	X	X	BaseLine
96	高次の回折が得られる。	High degree refraction is obtained .	A higher order round is obtained .	This firm has a good staff of workers .	BaseLine	BaseLine	X	BaseLine
97	彼の論文が私の弟見の手掛りとなりました。	I found a clue for my discovery in his paper .	His paper was a start for my discovery .	Higher order fractures are obtainable .	X	X	top10	BaseLine
98	太陽が雲層から突然顔を出した。	The sun flashed from behind the cloud .	The sun broke through the clouds .	His paper was beginning to take shape in my discovery .	top10	X	top10	top10
99				The sun broke through the clouds .	O	X	X	X



# 付 録 D    google-BaseLine の対比較評 価結果

number	input	reference	A		B		評価者A		評価者B		評価者C		評価者D	
			The traffic light changed from blue to red .	The traffic light changed from green to red .	The signal changed from green to red .	The traffic light changed from green to red .	He went to Kyoto on business .	He went to Kyoto on business .	fine	fine	fine	fine	fine	fine
0	信号が青より赤に変わった。	The signal changed from green to red .	The traffic light changed from blue to red .	The traffic light changed from green to red .	He went to Kyoto on business .	He went to Kyoto on business .	fine	fine	O	X	X	fine	fine	same
1	彼は仕事で京都へ行った。	He went to Kyoto on business .	The government presented some of the features of the project that are concrete .	The government presented the specifics for the project .	The government presented some of the features of the project that are concrete .	The government presented the specifics for the project .	google	google	google	google	X	google	X	fine
2	政府はこの事業の具体的な内容を示した。	The government presented some of the features of the project that are concrete .	Nobunaga's army went up to Kyoto again .	The Nobunagas went up to Tokyo again .	Nobunaga's army went up to Kyoto again .	The Nobunagas went up to Tokyo again .	O	X	O	X	google	google	google	google
3	信長勢は再び京へ上った。	Nobunaga's army went up to Kyoto again .	I was stranded as a result of the train accident .	I was stranded in a train accident.	I was stranded as a result of the train accident .	I was stranded in a train accident.	google	google	google	google	X	google	X	google
4	私は電事故で定置機を食った。	I was stranded as a result of the train accident .	She had her own way .	She had her own way .	She had her own way .	She had her own way .	fine	fine	fine	fine	fine	fine	fine	fine
5	彼女は我を通した。	She had her own way .	The swing is swinging .	The swing is swinging .	The swing is swinging .	The swing is swinging .	O	X	O	X	X	X	X	same
6	ぶらんこが揺れている。	The swing is swinging .	No sure cures for AIDS are known yet .	There is still no known definitive cure for AIDS.	No sure cures for AIDS are known yet .	There is still no known definitive cure for AIDS.	google	google	google	google	google	google	google	fine
7	エイズの確実な治療法はまだわかっていない。	No sure cures for AIDS are known yet .	He was convicted .	He was found guilty .	He was convicted .	He was found guilty .	O	X	O	X	X	X	X	same
8	彼は有罪の宣告を受けた。	He was convicted .	The window won't shut .	That window will not fasten .	The window won't shut .	That window will not fasten .	fine	fine	fine	fine	X	X	X	fine
9	その窓はなかなか閉まらない。	The window won't shut .	That starlet put everything she had into her part and won the prize for new talent .	The rookie actress won the rookie of the year award for her performance.	That starlet put everything she had into her part and won the prize for new talent .	The rookie actress won the rookie of the year award for her performance .	google	google	google	google	X	google	X	google
10	その新人女優は伝当たりの演技で新人賞を獲得した。	That starlet put everything she had into her part and won the prize for new talent .	The captives attempted to make their escape .	The prisoners of war tried to escape .	The captives attempted to make their escape .	The prisoners of war tried to escape .	fine	fine	fine	fine	X	google	X	google
11	捕虜たちが逃亡を図った。	The captives attempted to make their escape .	She seemed to be embarrassed at the unexpected question .	She seemed puzzled by the unexpected question.	She seemed to be embarrassed at the unexpected question .	She seemed puzzled by the unexpected question.	google	google	google	google	X	google	X	google
12	彼女は思いがけない質問にまごついたようだった。	She seemed to be embarrassed at the unexpected question .	The destruction of the natural environment around Shiretoko has become serious .	Nature destruction is progressing in Shiretoko.	The destruction of the natural environment around Shiretoko has become serious .	Nature is in an advanced stage of destruction in the chimney .	google	google	google	google	google	google	google	google
13	知床では自然破壊が進んでいる。	The destruction of the natural environment around Shiretoko has become serious .	The work got finished somehow .	I managed to finish the job somehow or other .	The work got finished somehow .	Somehow or another the work was done.	fine	fine	fine	fine	X	X	X	both
14	仕事はどうかにか仕上がった。	The work got finished somehow .	Glass breaks easily .	Glass breaks easily .	Glass breaks easily .	Glass is liable to break .	google	google	google	google	X	X	X	both
15	ガラスは割れやすい。	Glass breaks easily .	Mount Fuji has been painted by numerous artists since ancient times .	Mount Fuji has long been painted in various pictures .	Mount Fuji has been painted by numerous artists since ancient times .	Fuji has long been a mountain that has been depicted in many painters.	google	google	google	google	X	X	X	both
16	富士山は昔からたくさんさんの絵描きに描かれた山だ。	Mount Fuji has been painted by numerous artists since ancient times .	I shall let you know as soon as it is decided .	We will let you know as soon as it is decided.	I shall let you know as soon as it is decided .	I will let you know as soon as fixed .	fine	fine	fine	fine	fine	fine	fine	both
17	決まり次第お知らせします。	I shall let you know as soon as it is decided .	You can see ships' lights toward the offing .	You can see the light of the ship offshore.	You can see ships' lights toward the offing .	I see the lights of the boats offshore .	google	google	google	google	X	X	X	google
18	沖の方に船の明かりが見える。	You can see ships' lights toward the offing .	The negotiations were conducted in a peaceful atmosphere .	The negotiations were carried out calmly .	The negotiations were conducted in a peaceful atmosphere .	Negotiations were peaceful.	fine	fine	fine	fine	X	X	X	fine
19	交渉は穏やかに行われた。	The negotiations were conducted in a peaceful atmosphere .	The deficit comes to a billion yen .	The total amount of debt reaches 1 billion yen.	The deficit comes to a billion yen .	The debts amount to a billion yen .	fine	fine	fine	fine	X	X	X	google
20	負債総額は10億円に及ぶ。	The deficit comes to a billion yen .	She perceives things intuitively .	She gets at things intuitively .	She perceives things intuitively .	She perceives things intuitively.	google	google	google	google	X	X	X	google
21	彼女は直感的に物事をとらえる。	She perceives things intuitively .	After that , many speculators and general investors got in on the act .	Many attended the meeting after that , stockholders and general investors .	After that , many speculators and general investors got in on the act .	After that, many traders and general investors participated.	google	google	google	google	google	google	google	google
22	その後、多くの仕手筋や一般投資家に参加した。	After that , many speculators and general investors got in on the act .	Today , I did a lot of shopping at the department store .	I did a lot of shopping at the department store today .	Today , I did a lot of shopping at the department store .	I bought a lot of things at a department store today .	O	X	O	X	X	X	X	both
23	今日はデパートでたくさん買い物をした。	Today , I did a lot of shopping at the department store .	Yesterday's party was not very exciting .	There was not much excitement at yesterday's party .	Yesterday's party was not very exciting .	Yesterday's party wasn't very lively.	O	X	O	X	X	X	X	google
24	昨日のパーティーは、あまり盛り上がりなかった。	Yesterday's party was not very exciting .												

25	今回の法案には無理がある。	This bill is hard to pass .	This bill is impossible .	This bill is unreasonable.	google	X	google
26	どうぞお楽にしてください。	Please make yourself at home .	Please enjoy yourself.	Make yourself comfortable .	fine	fine	fine
27	この文面は彼の誠意さを証明している。	This letter proves his sincerity .	This sentence proves his sincerity.	This sentence exemplifies his sincerity .	google	google	both
28	彼らはその考え方に迷わされた。	They were led astray by that way of thinking .	They were misled by the idea.	The idea puzzled them .	google	X	google
29	彼は自己の権利を主張している。	He claims his own rights .	he claims his rights.	He insists upon his rights .	O	X	both
30	この山頂から望む富士はとても美しい。	The view of Mt .	The view of Mt.Fuji from this peak is very beautiful.	Fuji from this mountain peak is very beautiful .	google	fine	google
31	そのけんかは何年も糸を引いた。	The quarrel lasted for years .	The quarrel dragged on for years.	The quarrel ended for many years .	google	google	google
32	これが私にとっては精いっぱいのところだ。	This is the utmost limit of my power .	This is the most I can do .	This is the best part for me.	fine	fine	fine
33	彼は額の手を拭いた。	He wiped the perspiration off his forehead .	He wiped the sweat from his forehead.	He wiped the sweat from his brow .	google	X	both
34	田園生活を知った。	I gained acquaintance with rural life .	I learned about rural life.	I learned of country life .	google	google	both
35	太陽が壁々と現われた。	The sun appeared majestically .	The sun appeared majestically.	The sun appeared bright .	google	google	google
36	その両国間には交戦状態が存在する。	A state of belligerency exists between the two nations .	A state of war exists between the two countries.	There is war between the two countries .	google	X	google
37	その店の前に駐車した。	She parked her car in front of the store .	I parked in front of the store.	I parked in front of the store .	O	X	same
38	彼女は栗色の髪をしている。	Her hair is a chestnut color .	She has chestnut hair.	She has chestnut-colored hair .	O	X	fine
39	昔は大西洋渡航に5日かかった。	The Atlantic crossing used to take five days .	It used to take five days to cross the Atlantic.	In the past it used to take five days to cross the Atlantic Ocean .	O	X	fine
40	子供たちの旅行の語は食べ物に集中した。	The children's description of the trip centered on the food .	The children's travel stories centered around food.	The children's talk about traveling was centered on food .	O	X	fine
41	音波で鼓膜が揺動する。	The eardrums vibrate when they are struck by sound waves .	The eardrum vibrates in the sound waves .	Sound waves vibrate the eardrum.	O	X	google
42	彼の顔は赤くなった。	His face reddened .	his face turned red.	A blush mounted to his face .	google	X	google
43	噴水は四方八方に水をまき散らしていた。	The fountain sprayed water in every direction .	The fountain was spraying water in all directions.	The fountain was scattered in all directions .	google	X	google
44	彼女の心臓はときどきどきしていた。	Her heart thudded .	Her heart was pounding.	Her heart was thumping against her chest .	O	google	google
45	列車は1.5分間隔で運行している。	Trains are running at 15 - minute intervals .	Trains run every 15 minutes.	The trains run every fifteen minutes .	O	X	same
46	夜みんなメインロウジに集まった。	In the evening we gathered in the main lodge .	All the night they assembled at the main lodge .	We all gathered at the main lodge for the night.	google	google	google
47	その険しい山脈には木が1本も生えていなかった。	No trees grew on the steep mountainside .	Not a single tree grew on the steep hillside.	No trees grew on the steep hillside .	fine	X	both
48	ふたたび長い休止があった。	There was another long pause .	We had a long rest again .	There was another long pause.	google	fine	google
49	ある有望な可能性が生じた。	A promising possibility has turned up .	A promising possibility has arisen .	A promising possibility arose.	google	X	both
50	もっと早くにより多くの資本を投入すべきだった。	We were not quick enough to inject more capital into the economy .	More capital should have been put in earlier.	I should have invested more capital earlier .	fine	X	both
51	晩年はどや街で送った。	He spent his last years on skid row .	I spent my later years in Doyamachi.	We spent the last years in a street .	fine	google	google

52	下水管はこれだけ大量の水を処理しきれない。	The sewers can't cope with so much water .	Sewer pipes cannot handle this much water.	Sewers cannot handle this large amount of water .	O	X	X	google
53	腹がゴロゴロ鳴った。	His stomach rumbled .	The bowels rumbled .	My stomach was rumbling.	google	google	google	both
54	しばしば寒暖計が零度を下回る。	The thermometer often goes below zero .	Thermometers often drop below zero.	The thermometer is often below zero .	fine	X	X	google
55	豊富な語彙をもっている。	He has a copious vocabulary .	I have a rich vocabulary.	He possesses a wide vocabulary .	O	X	X	google
56	指輪はどこにも見当たらなかった。	The ring was nowhere to be found .	The ring was nowhere to be seen.	The ring was nowhere to be found .	fine	X	X	fine
57	その塔は900フィート以上ある。	The tower stands over 900 feet high .	The tower stands over 900 feet .	The tower is over 900 feet.	fine	google	X	google
58	傘が裏返しになった。	The umbrella turned inside out .	My umbrella was turned over .	The umbrella turned inside out.	google	google	X	google
59	人は誰しも間違いを犯すものだ。	One is certain to make slips .	Everyone makes mistakes.	Everyone is liable to error .	google	google	X	google
60	7マイルは1時間では歩ききれない距離だ。	Seven miles is a distance too great to walk in an hour .	Seven miles is too far to walk in one hour.	7 miles are beyond walking in an hour .	google	google	X	google
61	ボタンは太い糸でしっかりと布地につけられた。	The button was anchored to the cloth with heavy thread .	The button was firmly attached to the fabric with a thick thread.	The buttons were heavily threaded on the cloth .	google	google	google	google
62	幸福で彼女の目は輝いた。	Happiness kindled her eyes .	Happiness kindled her eyes .	Happiness lit up her eyes.	O	X	X	fine
63	誰だって1人ではできない。	No one man can do it .	Nobody can do it by himself .	No one can do it alone.	google	google	google	google
64	我々は特別番組を放送した。	We put on special programs .	We broadcast a special program.	We broadcast a special program .	O	X	X	same
65	彼はそこで戦場中に殺された。	He was killed in action there .	He was killed in battle there.	He was killed in the battle there .	O	X	X	same
66	当社製品はよく売れる。	Our products are marketable .	Our products sell well.	Our product sells well .	fine	X	X	same
67	我々の大学は長い輝かしい歴史を誇っている。	Our university can boast a long and glorious history .	Our university boasts a long and glorious history.	Our university has a long , brilliant history .	fine	google	X	google
68	彼は町の勢力家を抱き込んだ。	He managed to secure the help of a town boss .	He hugged the influential people of the town .	He embraced the town's powerhouses.	X	X	X	X
69	彼の胸は興奮で躍動していた。	His heart was throbbing with excitement .	His chest was beating with excitement.	His heart was racing with excitement .	google	X	X	fine
70	その事件はうやむやに葬られた。	The matter has been hushed up .	The incident was buried in obscurity.	The matter was hushed up .	fine	google	fine	google
71	車輪が上着を掃った。	The wheel rubbed my coat .	The wheels grazed my jacket.	The wheels robbed my coat .	google	google	google	google
72	どうぞ着席して下さい。	Please have a seat .	Please be seated.	Please be seated .	O	X	X	same
73	英語をやり直さねばならない。	I must brush up my English .	I'll have to improve my English .	I have to redo my English.	google	google	X	google
74	お話の要点がわかりません。	I'm afraid your point escapes me .	I don't understand the point of your story.	I don't understand the point you said .	fine	X	X	google
75	彼はバイクを木にぶつけた。	He hit his bike against a tree .	He crashed his bike into a tree.	He ran his motorbike into a tree .	google	X	X	fine
76	私は酒で頭がぼんやりしてしまった。	The spirits muddled my brain .	I've become dizzy with alcohol.	My brain has gotten muddled with alcohol .	fine	X	X	fine
77	雨で試合が流れた。	The game was rained out .	The game was washed away by the rain.	The game was washed out by the rain .	fine	X	X	fine
78	そのポンプは蒸気で動く。	The pump is run by steam .	The pump is driven by steam .	The pump runs on steam.	google	X	fine	fine
79	彼女の声は震えた。	Her voice wavered .	her voice trembled.	There was a tremble in her voice .	fine	fine	X	google

80	いつかは当局が動き出すだろう。	Some day the authorities will go into action .	At some point the authorities will act .	google	X	google	fine
81	時計がちかちかと時を刻んでいる。	The clock is ticking away the time .	The clock is ticking .	fine	X	google	same
82	お先に失礼します。	Excuse me , but I must be going now .	Pardon me , I have to go now .	google	google	google	X
83	彼は冗談が過ぎた。	His joke went too far .	He carried a joke too far .	google	google	google	fine
84	値段は大きすぎて着う。	The prices differ according to size .	The prices vary according to the size .	fine	X	X	both
85	彼は手紙を殴り書きました。	He scribbled his letters .	He scribbled a letter .	O	X	X	same
86	彼が得た報酬は微々たるものだった。	He only got a small reward .	His reward was meager .	google	google	X	google
87	その部屋はむんむんしていた。	It was stuffy in the room .	The room was trembling .	google	google	google	google
88	転職を考えています。	I'm thinking of changing my jobs .	I'm considering changing jobs .	google	X	X	both
89	基本的に、米国は日本に二つの行動を求めている。	Essentially , the United States is asking Japan to do two things .	Essentially , the United States is asking Japan to take two actions .	O	google	google	google
90	ソマリアを他山の石とした旧ユーゴPKO関係者が武力行使に慎重だったため、米国は不満だった。	The United States was dissatisfied with the United Nations since those involved in the peacekeeping operation in the former Yugoslavia were cautious about using armed forces because of their experience in Somalia .	The United States was dissatisfied with the fact that former Yugoslav PKO officials, who considered Somalia to be a rock, were cautious about using force .	google	google	google	X
91	田中首相は在オーストラリア大使館の元館員による公金流用疑惑について、杉浦正輝副大臣に再調査を指示した。	Foreign Minister Makiko Tanaka on Thursday instructed Senior Vice Foreign Minister Masatake Sugura to conduct a new probe into the alleged misappropriation of public funds by a former embassy official in Australia .	Foreign Minister Masatake Sugura was asked by Foreign Minister Masatake Sugura on suspicion of using public money for officials at the Japanese Embassy in Australia .	google	google	google	X
92	韓国にもまた、一層の努力を求めたい。	We exhort South Korea to make stronger efforts as well .	I would also like to ask South Korea to make further efforts .	O	google	X	google
93	1 2号線環状部は一九九二年に着工、来年十二月に開業を予定している。	Construction on the subway line began in 1992 , and the line is scheduled to be put into operation in December next year .	Construction will begin on Track 12 in 1992 and will begin next December .	google	google	google	X
94	従業員は約千人で、投資額は五千万米ドルに及ぶ。	The average monthly wage for a Chinese worker is about 1,000 yuan , or about 115 dollars .	The company employs about \$50 million and employs about 1,000 people .	google	google	google	google
95	北朝会は一九六九年に府の設立認可を受け、大阪円生病院と大和川病院を運営。	They are Osaka Ensei Hospital in Higashi Sumiyoshi Ward , Osaka , and Yamatogawa Hospital in Kashihara , Osaka prefecture .	The Kitanshikkai was established in 1964 by the prefectural government and runs the Osaka Ensei Hospital and the Owakawa Hospital .	google	X	google	google
96	この会社には人材が揃っている。	This company has a good staff of talents .	This firm has a good staff of workers .	fine	fine	fine	X
97	高次の回折が得られる。	High degree refraction is obtained .	Higher orders of diffraction are obtained .	google	google	google	google
98	彼の論文が私の発見の手掛りとなりました。	I found a clue for my discovery in his paper .	His paper was beginning to take shape in my discovery .	google	google	google	google
99	太陽が雲間から突然顔を出した。	The sun flashed from behind the cloud .	The sun broke through the clouds .	google	fine	google	google

# 付 録 E    google-Proposed の対比較評 価結果

number	input	reference	A		B		評価者A		評価者B		評価者C		評価者D	
							top10	top10	top10	top10	top10	top10	top10	top10
0	信号が青より赤に変わった。	The signal changed from green to red .	The traffic light changed from green to red .	The traffic light changed from blue to red.										
1	彼は仕事で京都へ行った。	He went to Kyoto on business .	He went to Kyoto on business .	He went to Kyoto on business.										
2	政府はこの事業の具体的な内容を示した。	The government presented some of the features of the project that are concrete .	The government showed the specific contents of this project.	The government laid out the concrete details of the project .										
3	信長勢は再び京へ上った。	Nobunaga's army went up to Kyoto again .	Nobunaga's army went up to Kyoto again.	The Nobunaga went up to Kyoto again .										
4	私は電事故で定置機を食った。	I was stranded as a result of the train accident .	I was stranded in a train accident.	I got stuck in a train accident .										
5	彼女は我を通した。	She had her own way .	She passed herself through .	she got me through										
6	ぶらんこが揺れている。	The swing is swinging .	The swings are swinging .	The swing is swinging.										
7	エイズの確実な治療法はまだわかっていない。	No sure cures for AIDS are known yet .	There is still no known definitive cure for AIDS.	No sure cure for AIDS is yet known .										
8	彼は有罪の宣告を受けた。	He was convicted .	he was found guilty.	He was declared guilty .										
9	その窓はなかなか閉まらない。	The window won't shut .	The window will not close .	The window is hard to close.										
10	その新人女優は伝説的な演技で新人賞を獲得した。	That starlet put everything she had into her part and won the prize for new talent .	The rookie actress won the rookie of the year award for her performance.	The new actress won the rookie prize for her personal performance .										
11	捕虜たちが逃亡を図った。	The captives attempted to make their escape .	The prisoners tried to escape.	The prisoners made their escape .										
12	彼女は思いがけない質問にまごついたようだった。	She seemed to be embarrassed at the unexpected question .	She seemed to be confused by the unexpected question .	She seemed puzzled by the unexpected question.										
13	知床では自然破壊が進んでいる。	The destruction of the natural environment around Shiretoko has become serious .	Nature destruction is progressing in Shiretoko.	Nature is destroyed in Shiretoko .										
14	仕事はどのようにか仕上がった。	The work got finished somehow .	The work was finished somehow or other .	Somehow or another the work was done.										
15	ガラスは割れやすい。	Glass breaks easily .	Glass is liable to break .	Glass breaks easily.										
16	富士山は昔からたくさん人の絵描きに描かれた山だ。	Mount Fuji has been painted by numerous artists since ancient times .	Mount Fuji has been painted in numerous ways since ancient times .	Fuji has long been a mountain that has been depicted in many painters.										
17	決まり次第お知らせします。	I shall let you know as soon as it is decided .	We will let you know as soon as it is decided.	I'll let you know as soon as it is decided .										
18	沖の方に船の明かりが見える。	You can see ships' lights toward the offing .	You can see the light of the ship offshore.	We can see the lights of the ship offshore .										
19	交渉は穏やかに行われた。	The negotiations were conducted in a peaceful atmosphere .	The negotiations were conducted peacefully .	Negotiations were peaceful.										
20	負債総額は10億円に及ぶ。	The deficit comes to a billion yen .	The total amount of debt reaches 1 billion yen.	The total debt amounts to a billion yen .										
21	彼女は直感的に物事をとらえる。	She perceives things intuitively .	She perceives things intuitively.	She catches things intuitively .										
22	その後、多くの仕手筋や一般投資家が参加した。	After that , many speculators and general investors got in on the act .	The meeting was followed by a large number of speculators and general investors.	After that, many traders and general investors participated.										
23	今日はデパートでたくさん買い物をした。	Today , I did a lot of shopping at the department store .	I did a lot of shopping at the department store today.	I did a lot of shopping at a department store today .										
24	昨日のパーティーは、あまり盛り上がりなかった。	Yesterday's party was not very exciting .	Yesterday's party was not very lively .	Yesterday's party wasn't very lively.										

25	今回の法案には無理がある。	This bill is hard to pass .	This bill is unreasonable .	o	X	X
26	どうぞお楽にしてください。	Please make yourself at home .	Please enjoy yourself .	top10	top10	top10
27	この文面は彼の誠意さを証明している。	This letter proves his sincerity .	This sentence proves his sincerity .	o	X	google X
28	彼らはその考え方に迷わされた。	They were led astray by that way of thinking .	They were deceived by the idea .	google	google	google
29	彼は自己の権利を主張している。	He claims his own rights .	He asserts his own rights .	google	X	top10 google
30	この山頂から望む富士はとても美しい。	The view of Mt .	Fuji from the top of this mountain is very beautiful .	google	X	google X
31	そのけんかは何年も糸を引いた。	The quarrel lasted for years .	The quarrel trailed away for years .	X	google	google google
32	これが私にとっては精いっぱいのところだ。	This is the utmost limit of my power .	This is the utmost for me .	o	top10	top10
33	彼は額の手を拭いた。	He wiped the perspiration off his forehead .	He wiped the sweat from his forehead .	o	X	X
34	田園生活を知った。	I gained acquaintance with rural life .	I learned about rural life .	google	X	google
35	太陽が堂々と現われた。	The sun appeared majestically .	The sun came out in all its glory .	o	google	google google
36	その両国間には交戦状態が存在する。	A state of belligerency exists between the two nations .	A state of war exists between the two countries .	top10	X	top10 google
37	その店の前に駐車した。	She parked her car in front of the store .	I parked in front of the store .	o	X	X
38	彼女は栗色の髪をしている。	Her hair is a chestnut color .	She has chestnut hair .	o	X	X
39	昔は大西洋渡航に5日かかった。	The Atlantic crossing used to take five days .	It used to take five days to travel to the Atlantic .	google	google	google google
40	子供たちの旅行の話は食べ物に集中した。	The childrens description of the trip centered on the food .	The childrens travel stories centered around food .	google	X	top10
41	音波で鼓膜が振動する。	The eardrums vibrate when they are struck by sound waves .	Sound waves vibrate the eardrum .	o	X	X
42	彼の顔は赤くなった。	His face reddened .	His face turned red .	o	X	X
43	噴水は四方八方に水をまき散らしていた。	The fountain sprayed water in every direction .	The fountain was spraying water in all directions .	o	google	google top10
44	彼女の心臓はときどき止っていた。	Her heart thudded .	Her heart was beating fast .	google	X	google X
45	列車は15分間隔で運行している。	Trains are running at 15 - minute intervals .	Trains run every 15 minutes .	o	X	X
46	夜みんなメイソングに集まった。	In the evening we gathered in the main lodge .	Everybody gathered at the main lodge at night .	google	X	X top10
47	その険しい山腹には木が1本も生えていなかった。	No trees grew on the steep mountainside .	The steep hillside lacked a single tree .	google	google	google google
48	ふたたび長い休みがあった。	There was another long pause .	There was another long pause .	o	X	google X
49	ある有望な可能性が生じた。	A promising possibility has turned up .	A promising possibility arose .	o	X	X
50	もっと早めにより多くの資本を投入すべきだった。	We were not quick enough to inject more capital into the economy .	More capital should have been put in earlier .	o	X	top10 X
51	晩年はどや町で渡った。	He spent his last years on skid row .	I spent my later years in Doyamachi .	google	X	google X



52	下水管はこれだけ大量の水を処理しきれない。	The sewers can't cope with so much water .	Sewers cannot process such a large amount of water .	Sewer pipes cannot handle this much water .	o	X	google
53	腹がゴロゴロ鳴った。	His stomach rumbled .	My stomach was rumbling .	My stomach rumbled .	top10	X	X
54	しばしば寒暖計が零度を下回る。	The thermometer often goes below zero .	The mercury often falls below zero .	Thermometers often drop below zero .	google	X	google
55	豊富な語彙をもっている。	He has a copious vocabulary .	I have a rich vocabulary .	He has a large vocabulary .	o	X	X
56	指輪はどこにも見当たらなかった。	The ring was nowhere to be found .	The ring was nowhere to be seen .	The ring was nowhere to be found .	top10	X	top10
57	その塔は900フィート以上ある。	The tower stands over 900 feet high .	The tower stands over 900 feet .	The tower is over 900 feet .	google	X	top10
58	傘が裏返しになった。	The umbrella turned inside out .	The umbrella turned inside out .	The umbrella was turned inside out .	google	X	google
59	人は誰しも間違いをするものだ。	One is certain to make slips .	Everybody makes a mistake .	Everyone makes mistakes .	o	X	top10
60	7マイルは1時間では歩ききれない距離だ。	Seven miles is a distance too great to walk in an hour .	Seven miles is too far to walk in one hour .	Seven miles are beyond walking in an hour .	o	X	google
61	ボタンは太い糸でしっかりと布地につけられた。	The button was anchored to the cloth with heavy thread .	The button was firmly draped on the cloth with thick thread .	The button was firmly attached to the fabric with a thick thread .	X	google	google
62	幸福で彼女の目は輝いた。	Happiness kindled her eyes .	Happiness lit up her eyes .	Happiness shone in her eyes .	o	X	google
63	誰だって1人ではできない。	No one man can do it .	No one can do it alone .	Nobody can do it alone .	o	X	X
64	我々は特別番組を放送した。	We put on special programs .	We broadcast a special program .	We broadcast a special program .	o	X	X
65	彼はそこで戦闘中に殺された。	He was killed in action there .	He was killed there during the battle .	He was killed in battle there .	o	top10	X
66	当社製品はよく売れる。	Our products are marketable .	Our product sells well .	Our products sell well .	o	X	X
67	我々の大学は長い輝かしい歴史を誇っている。	Our university can boast a long and glorious history .	Our university boasts a long , glorious history .	Our university boasts a long and glorious history .	o	google	X
68	彼は町の勢力家を抱き込んだ。	He managed to secure the help of a town boss .	He embraced the town's powerhouses .	He embraced the powerhouses in the town .	X	X	X
69	彼の胸は興奮で躍動していた。	His heart was throbbing with excitement .	His chest was beating with excitement .	His heart was racing with excitement .	o	google	google
70	その事件はうやむやに葬られた。	The matter has been hushed up .	The incident was buried in obscurity .	The case was buried peacefully .	google	google	google
71	車輪が上着を掃った。	The wheel rubbed my coat .	The wheels grazed my jacket .	The wheels robbed me of my coat .	o	google	google
72	どうぞ着席して下さい。	Please have a seat .	Please be seated .	Please sit down .	google	google	google
73	英語をやり直さねばならない。	I must brush up my English .	I have to redo my English .	I have to redo my English .	o	X	X
74	お話の要点がわかりません。	I'm afraid your point escapes me .	I don't understand the gist of what you said .	I don't understand the point of your story .	o	X	X
75	彼はバイクを木にぶつけた。	He hit his bike against a tree .	He crashed his bike into a tree .	He smashed his motorcycle into a tree .	o	google	google
76	私は酒で頭がぼんやりしてしまった。	The spirits muddled my brain .	I've become dizzy with alcohol .	My head went blank from drinking .	google	google	google
77	雨で試合が流れた。	The game was rained out .	The game was washed away by the rain .	The game was rained out .	top10	top10	top10
78	そのポンプは蒸気で動く。	The pump is run by steam .	The pump runs on steam .	The pump is driven by steam .	top10	X	google
79	彼女の声は震えた。	Her voice wavered .	Her voice vibrated .	her voice trembled .	o	X	google

80	いつかは当局が動き出すだろう。	Someday the authorities will go into action .	The authorities will take action some day .	At some point the authorities will act .	top10	X	X	X
81	時計がちかちかと時を刻んでいる。	The clock is ticking away the time .	The clock is ticking away the minutes .	The clock is ticking .	top10	X	top10	google
82	お先に失礼します。	Excuse me , but I must be going now .	Please excuse me the first time .	Pardon me , I have to go now .	0	X	top10	google
83	彼は冗談が過ぎた。	His joke went too far .	He joked too much .	He passed the joke .	google	google	google	google
84	値段は大きすぎて違う。	The prices differ according to size .	Prices vary according to the size .	Prices vary by size .	0	google	X	X
85	彼は手紙を殴り書きました。	He scribbled his letters .	He scribbled the letter .	He scribbled a letter .	0	X	X	X
86	彼が得た報酬は微々たるものだった。	He only got a small reward .	His remuneration he received was substantive .	His reward was meager .	google	google	google	google
87	その部屋はむんむんしていた。	It was stuffy in the room .	The room was stuffy .	The room was stuffy .	0	X	X	X
88	転職を考えています。	I'm thinking of changing my jobs .	I am thinking of changing jobs .	I'm thinking of changing jobs .	0	X	X	X
89	基本的に、米国は日本に二つの行動を求めている。	Essentially , the United States is asking Japan to do two things .	Basically , the United States urges Japan to take two actions .	Essentially , the United States is asking Japan to take two actions .	0	X	X	X
90	ソマリアを他山の石とした旧ユーゴPKO関係者が武力行使に慎重だったため、米国は不満だった。	The United States was dissatisfied with the United Nations since those involved in the peacekeeping operation in the former Yugoslavia were cautious about using armed forces because of their experience in Somalia .	The United States was dissatisfied with the fact that former Yugoslav PKO officials, who considered Somalia to be a rock , were cautious about using force .	The United States was dissatisfied with the fact that former Yugoslav PKO officials, who considered Somalia had been cautious about using force .	X	top10	X	google
91	田中首相は在オーストラリア大使館の元館員による公金流用疑惑について、杉浦正輝副大臣に再調査を指示した。	Foreign Minister Makiko Tanaka on Thursday instructed Senior Vice Foreign Minister Masatake Sugitara to conduct a new probe into the alleged misappropriation of public funds by a former embassy official in Australia .	Foreign Minister Tanaka instructed State Minister Masatake Sugitara to re-investigate allegations of misappropriation of public funds by a former staff member of the Australian embassy .	Tanaka reportedly instructed former members of the Australian Embassy to conduct a re-examination of suspicions using public money .	X	google	X	google
92	韓国にもまた、一層の努力を求めたい。	We exhort South Korea to make stronger efforts as well .	We urge South Korea to make further efforts .	I would also like to ask South Korea to make further efforts .	0	google	X	top10
93	12号線環状部は一九九二年に着工、来年十二月に開業を予定している。	Construction on the subway line began in 1992 , and the line is scheduled to be put into operation in December next year .	Construction of the Route 12 loop began in 1992 and is scheduled to open in December next year .	Construction will begin on Route 12 , which is scheduled to begin in December next year .	google	google	X	google
94	従業員は約千人で、投資額は五千万米ドルに及ぶ。	The average monthly wage for a Chinese worker is about 1,000 yuan , or about 115 dollars .	With about 1,000 employees , the company plans to invest 50 million US dollars .	The company employs about 1,000 people and has invested 50 million US dollars .	google	google	X	google
95	北朝会は一九六九年に府の設立認可を受け、大阪円生病院と大和川病院を運営。	They are Osaka Ensei Hospital in Higashi Sumiyoshi Ward , Osaka , and Yamatogawa Hospital in Kashihara , Osaka prefecture .	The Kitanshiki Kaikan was approved by the prefectural government in 1964 and operates the Osaka Ensei Hospital and the Daivagawa Hospital .	In 1969 , Kitanshikkai was approved by the prefectural government to operate Osaka Ensei Hospital and Yamatogawa Hospital .	top10	google	google	top10
96	この会社には人材が揃っている。	This company has a good staff of talents .	This company has a staff of people .	This company is fully staffed .	google	google	google	google
97	高次の回折が得られる。	High degree refraction is obtained .	A higher order round is obtained .	Higher orders of diffraction are obtained .	google	X	google	X
98	彼の論文が私の発見の手掛りとなりました。	I found a clue for my discovery in his paper .	His paper was a start for my discovery .	His paper led to my discovery .	top10	google	google	X
99	太陽が雲間から突然顔を出した。	The sun flashed from behind the cloud .	The sun suddenly appeared from behind the clouds .	The sun broke through the clouds .	google	google	google	top10

## 謝辞

最後に、一年間に渡り、本研究のご指導をいただきました鳥取大学工学部知能情報工学科自然言語処理研究室の村上仁一准教授、村田真樹教授に深く感謝すると共に、厚く御礼申し上げます。そして、日常の議論を通じて多くの知識や示唆を頂いた同研究室の皆様に深謝いたします。

人手評価に協力してくれた、以下の学生の協力を得ました。感謝致します。(丸山 京祐, 深谷 諒, 松本 武尊, 柳原 弘哉)

## 参考文献

- [1] Chenhui Chu and Rui Wang. A survey of domain adaptation for neural machine translation. In *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 1304–1319, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics.
- [2] 森下睦, 帖佐克己, 鈴木潤, 永田昌明. Jparacrawl v3.0: 大規模日英対訳コーパス. 言語処理学会 第28回年次大会 発表論文集 (2022年3月), 2022.
- [3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, Vol. 30, , 2017.
- [4] Marta Bañón, Pinzhen Chen, Barry Haddow, Kenneth Heafield, Hieu Hoang, Miquel Esplà-Gomis, Mikel L Forcada, Amir Kamran, Faheem Kirefu, Philipp Koehn, et al. Paracrawl: Web-scale acquisition of parallel corpora. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 4555–4567, 2020.
- [5] Philipp Koehn and Rebecca Knowles. Six challenges for neural machine translation. *arXiv preprint arXiv:1706.03872*, 2017.
- [6] Alberto Poncelas, Gideon Maillette de Buy Wenniger, and Andy Way. Transductive data-selection algorithms for fine-tuning neural machine translation. *arXiv preprint arXiv:1908.09532*, 2019.
- [7] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine

- learning in Python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830, 2011.
- [8] Minh-Thang Luong and Christopher D Manning. Stanford neural machine translation systems for spoken language domains. In *Proceedings of the 12th International Workshop on Spoken Language Translation: Evaluation Campaign*, 2015.
- [9] 村上仁一. 日英対訳データベースの作成のための 1 考察. 言語処理学会第 17 回年次大会発表論文集, D4-5, pp. 979–82, 2011.
- [10] Taku Kudo and John Richardson. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. *arXiv preprint arXiv:1808.06226*, 2018.
- [11] Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. fairseq: A fast, extensible toolkit for sequence modeling. *arXiv preprint arXiv:1904.01038*, 2019.
- [12] Matt Post. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pp. 186–191, Belgium, Brussels, October 2018. Association for Computational Linguistics.
- [13] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311–318, 2002.