

# 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

土井 惟成      近藤 真史      山藤 敦史

株式会社 日本取引所グループ

{n-doi, m-kondo, a-santo}@jpx.co.jp

## 1 はじめに

日本の証券市場において、海外投資家の存在感が高まっている。東京証券取引所(以下「東証」)が公表している株式分布状況調査 [1] 及び投資部門別売買状況 [2] によると、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1990年頃よりほぼ一貫して上昇しており、2018年3月末時点の調査では約30%であるほか、海外投資家の取引が株式の売買代金に占める割合についても同様であり、2017年には約58%となっている。

投資家の投資判断における重要な情報源の一つに、上場会社が開示する「コーポレート・ガバナンス\*1に関する報告書」(以下「CG報告書」)がある [3]。東証では、有価証券上場規程において、コーポレートガバナンス・コード\*2(以下「CGコード」)中の各原則に対する遵守状況等をCG報告書に記載して東証に提出することを、上場会社に対して義務付けている\*3。

上場会社が開示するCG報告書は、各社のコーポレート・ガバナンスの取り組み状況等が投資家によって分析可能な形式で記載されている。また、和文のCG報告書と併せて英文のCG報告書も開示できるようになっており、東証では「外国人株主の比率が高い企業においては英文でのCG報告書の開示は望まれるところである」[4]と示しているものの、英文のCG報告書を開示している東証上場会社は、全体の約4.4%(2018年3月末時点)に留まる。

このような状況を踏まえて、海外投資家にとってCG報告書の利用をより容易にする一方法として、機械翻訳の活用が考えられる。ニューラル機械翻訳モデル [5](以下「NMTモデル」)による機械翻訳は、従来の機械翻訳モデルより翻訳品質が高いことで知られているものの、専門用語の多い特定分野の文を対象にすると、汎用的な学習データを用いたNMTモデルでは翻訳品質が低くなる [6]。一方で、CG報告書に特化したNMTモデルを作成するには、学習データとして活用できる言語資源が現時点では乏しい。これら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て用いることで、特定分野を対象にした機械翻訳の品質を高める分野適応(Domain

表1 CG報告書記載項目の区分

|     |   |
|-----|---|
| I   | コーポレート・ガバナンスに関する基本的な考え方及び資本構成、企業属性その他の基本情報    |
| II  | 経営上の意思決定、執行及び監督に係る経営管理組織その他のコーポレート・ガバナンス体制の状況 |
| III | 株主その他の利害関係者に関する施策の実施状況                        |
| IV  | 内部統制システム等に関する事項                               |
| V   | その他   |



図1 CG報告書(和文)の例

Adaptation) という手法がある。分野適応によって作成した機械翻訳モデルのことを「カスタム翻訳モデル(Custom Translation Models)」または「カスタムモデル(Custom Models)」という [7, 8]。

本研究では、CG報告書の機械翻訳における課題を整理の上、既に和文と英文の両方で開示されているCG報告書等から対訳コーパスを構築し、分野適応を施したNMTモデルを作成する。その上で翻訳品質を評価し、CG報告書における機械翻訳の活用可能性を検証した。実験の結果、構築した対訳コーパスを用いてカスタム翻訳モデルを作成することによって、汎用的なNMTモデルと比較してBLEUスコア [9]の上昇が確認できた。

## 2 CG報告書

上場会社が開示するCG報告書は、東証によって様式及び記載要領が定められており、大区分としては表1の5項目が記載されている。図1にCG報告書の例を示す。通常、上場会社が東証を通じて開示する資料のファイル形式はPDFと定められているが、和文のCG報告書についてはPDFファイルと併せてXBRL\*4文書の提出が要請されている。XBRL文書のCG報告書は、各記載項目にXML形式のタグが付与されており、PDFと比較して機械的な処理が容易になっている。

\*4 財務諸表をはじめとする会計用の書類で使用されている、XMLを基にして構築されたマークアップ言語。

\*1 企業統治のこと。単にCGとも略される。CGコード中では、「会社が、株主をはじめ顧客・従業員・地域社会等の立場を踏まえた上で、透明・公正かつ迅速・果敢な意思決定を行うための仕組み」を意味する。

\*2 実効的なコーポレート・ガバナンスの実現に資する主要な原則を取りまとめた規約であり、有価証券上場規程の別添として位置づけられている。

\*3 有価証券上場規程第419条及び同規程第436条より。

### 3 CG 報告書中の和文の特徴

本研究の予備実験として、CG 報告書中の和文を複数の機械翻訳サービスに入力し、誤訳の傾向と誤訳が発生しやすい和文の特徴を考察した。本予備実験で特筆して見受けられた、CG 報告書の機械翻訳において課題になると考えられる傾向を以下の節に示す。

#### 3.1 専門用語・固有名詞

CG 報告書では、Investor Relations (以下「IR」) や証券の分野における専門用語及び会社名をはじめとする固有名詞が多用されており、汎用的な機械翻訳サービスでこれらを正確に英訳することは難しい傾向にある。また、専門用語によっては、代表的な英訳が一意に定まっていないものがある。これらの例を表 2 に示す。

表 2 専門用語・固有名詞の例

| 原文              | 参照訳 <sup>*5</sup>   | 訳出例  |
|-----------------|---|--|
| 監査役             | Audit & Supervisory Board Member<br>Audit and Supervisory Board Members<br>Statutory auditor<br>Kansayaku | Auditor  |
| 議決権電子行使プラットフォーム | Electronic Voting Platform<br>Platform for the electronic exercise of voting rights                       | Voting right electronic exercise platform        |
| EY 新日本有限責任監査法人  | Ernst & Young ShinNihon LLC   | EY New Japan limited liability audit corporation |
| PwC あらた有限責任監査法人 | PricewaterhouseCoopers Aarata LLC   | PwC Yuru limited liability audit corporation     |

#### 3.2 訳揺れしやすい代名詞

多くの CG 報告書では、自社を指す代名詞として「当社」、直前の人物を指す代名詞として「同氏」といった単語が使われている。これらの英訳には複数の候補が存在しており、英訳時には文脈の把握やスタイルの統一が必要となる。この例を表 3 に示す。

表 3 訳揺れしやすい代名詞の例

| 代名詞 | 英訳例         | 例文  |
|-----|-------------|---|
| 当社  | We          | We fully comply all aspects of the Corporate Governance Code.                   |
|     | The Company | The Company implements each principle of the Corporate Governance Code.         |
|     | (自社名)       | (自社名) implements each of the basic principles of the Corporate Governance Code. |
| 同氏  | He/She      | He has never been the Company's legal advisor.                                  |
|     | (個人名)       | (個人名) has extensive knowledge and experience as a lawyer.                       |

#### 3.3 一文中的並列句

上場会社は自社のコーポレート・ガバナンスに関する取り組み状況を記述するに当たって、項目の細分化や項目の列挙を行うことが多い。そのため、CG 報告書中には、序数を用いた箇条書きや、一文の中に複数の並列句を挿入した文が出現する。特に、改行区切りの箇条書きを含んだ文を、一行ずつ翻訳を行う機械翻訳サービスに入力すると誤訳が発生しやすい。この例を表 4 に示す。また、この文構造は、対訳コーパスを構築する際にも同様に課題となる。

#### 3.4 名詞の非限定列挙

「甲、乙、丙等」における「等」のような、列挙した名詞が一例であること(非限定列挙)を示す語句を含んだ文を正しく英訳するには、前後の文脈やその単語の前提知識を要する場合がある。この例を表 5 に示す。

<sup>\*5</sup> 本稿における参照訳とは、英文の CG 報告書に見られた表現の例をいう。以下同様。

表 4 一文中的並列句の例

| 原文  | 参照訳  | 訳出例   |
|---|--|---|
| 強化すべき点としては、当社事業がグローバルに拡大する中において、ガバナンス機能の更なる充実に向けた取組みが重要との認識に立ち、 | Recognizing that efforts for further enhancement of the governance function is vital to Santen in the midst of the global expansion of its business, the Company has decided to exert further efforts with respect to items that need to be strengthened including those listed below: | As a point of strengthening, we recognize that efforts aimed at further enhancement of the security function are important as our business expands to global. |
| ・監督機能をより強化するため、取締役会で意思決定した重要な事項に対する継続的なモニタリングをより充実させること         | With the aim of strengthening the monitoring function, further enhancing the continuous monitoring of material matters that are decided at meetings of the Board of Directors; and   | ・ Enhance continuous monitoring of important matters made by the Board of Directors in order to enhance the supervision function.                             |
| ・リスクマネジメント視点での議論を更に強化すること                                       | Further strengthening discussions from the viewpoint of risk management.   | ・ To further strengthen the discussion from the perspective of risk management.   |
| などについて、取り組むこととしました。   |  | I decided to work on what.  |

表 5 名詞の非限定列挙の例

| 原文              | 参照訳   | 訳出例   |
|-----------------|---|---|
| 最高経営責任者等の後継者の計画 | Succession planning for positions such as president | Planning for successors of chief executive officers |

#### 3.5 見出し符号(見出し番号)

CG 報告書では多様な見出し符号が用いられており、「(1) (2) (3)」や「a) b) c)」のような、英文でも利用されるものの他に、丸数字、漢数字、片仮名といった、英文に即する形への変換が必要なものがある。特に、片仮名の「(イ)」については、「(ア)(イ)(ウ)」と「(イ)(ロ)(ハ)」の二通りが存在しており、前後の文脈を踏まえた英訳が必要となる。

#### 3.6 造語・詩的な表現

CG 報告書には経営理念等を記載する項目がある。当該項目の文では、日本語特有のニュアンスを踏まえた造語や詩的な表現が用いられる傾向にある。これらの文を機械翻訳サービスに入力すると、それらのニュアンスが失われ、正しく意味が伝わらない恐れがある。この例を表 6 に示す。

表 6 造語・詩的な表現の例

| 原文  | 参照訳例   | 訳出例   |
|---|--|---|
| 当社は、経営理念(Mission)を「日本を世界一豊かに。その未来へ心を尽くす一瞬一瞬の「いちご」とし、定款に定めております。 | Ichigo's mission is to preserve and improve real estate and contribute to a sustainable society. | We have defined our management philosophy (Mission) as "the world's richest in the world." "Ichigo in a once-in-a-lifetime meeting to the future", stipulated in the Articles of Incorporation. |
| 挑戦することを真剣に楽しみ、独創的な「道(どう)」を極めます。                                 | Embrace challenges and seek to master the Doh ("Way" or "Path") of creativity.                   | I am seriously enjoying challenging and continuing to keep my ingenious "way".  |

### 4 CG 報告書対訳コーパス

本研究に際して、CG コード及び和文と英文の両方で開示されている CG 報告書を基に、翻訳会社に対応付け(アライメント)を依頼することで対訳コーパスを構築した。この対応付けとは、ある和文に対して意味の対応する英文を割り当てることで対訳を作成する作業であり、今回は改行や句点を文の区切りとした。なお、和文と英文が一对一で対応するとは限らないため、和文または英文が二文以上となる対訳も存在した。

CG コードは和文と英文がそれぞれ公開されており、今回は初版の CG コード(2015 年 6 月公開、A4 用紙 40 ページ)を対象とした。また CG 報告書については、2018 年 3 月末時点で和文と英文の両方を開示している 160 社の、同時点で最新の CG 報告書を対象とした。

構築した対訳コーパスについては、対訳の意味の整合等を目的として、以下をはじめとする修正作業を施し

た。この過程において、3.3 節に挙げた例をはじめとす  
る、和文と英文で意味の整合が困難な対訳は削除した。

- 文字コードの正規化<sup>\*6</sup>
- 整合しない固有名詞の置換<sup>\*7</sup>
- 整合しない記号の削除<sup>\*8</sup>
- 訳注の削除
- URL の削除
- 人名のみ・数値のみの対訳の削除
- 重複する対訳の削除

本稿ではこれらの作業済みの対訳コーパスを「CG 報告書対訳コーパス」と呼ぶ。CG 報告書対訳コーパスに含まれる対訳数は 43,158 であった。この和文の文字列長と英文の単語数の分布を図 2 に示す。

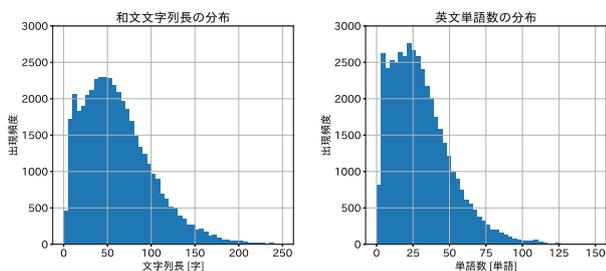


図 2 和文文字列長の分布 (左図) と英文単語数の分布 (右図)

## 5 実験

### 5.1 実験設定

本研究では、Google Cloud Platform<sup>TM</sup>が提供する AutoML Translation[10]<sup>\*9</sup>を利用して実験を行った。AutoML Translation では、入力した対訳コーパスを用いて Google NMT[11] (以下「GNMT」) に分野適応を施し、その対訳コーパスに適したカスタム翻訳モデルを作成するサービスを提供している<sup>\*10</sup>。

実験では、今回構築した CG 報告書対訳コーパス (以下「全量版コーパス」) と、当該コーパスから「和文の文字列長が閾値 (100 字) 以上の対訳」及び「英文の単語数が閾値 (50 単語) 以上の対訳」を削除した対訳コーパス (以下「長文除去版コーパス」) を利用した。この閾値は、文字列長及び単語数に比して十分な対訳数が確保できていない値として、図 2 における変曲点と思われる値を踏まえつつ決定した。これらの対訳数等を表 7 に示す。各対訳コーパスについて 3 回実験を実施

<sup>\*6</sup> 和文の CG 報告書は Unicode のテキストデータ (XBRL) が提供されているものの、英文は PDF 形式のみの公開となっており、和文と英文で文字コードが整合しない場合がある。また、和文に CJK 互換漢字 (U+F900~U+FAFF) が含まれている場合があるため、これらを CJK 統合漢字の代表字に正規化を行った。(例: U+F9BE (料) → U+6599 (料))

<sup>\*7</sup> 和文では代名詞の使用や主語等の省略を行っているものの、英文では固有名詞を記載している対訳について、その固有名詞を代名詞に置換した。例として、和文の主語は「当社」だが、英文の主語は自社名となっているような対訳について、英文の主語を「The Company」に置換する等が挙げられる。

<sup>\*8</sup> 和文と英文の片方だけに、括弧や米印等の記号が見られる対訳について、それらの記号を削除するといった修正。

<sup>\*9</sup> 2019 年 1 月 15 日現在、β 版。

<sup>\*10</sup> AutoML Translation で行う分野適応の具体的なアルゴリズム等は公開されていない。

し、それぞれについて BLEU による翻訳の評価を行い、その平均値を算出した。

表 7 実験データの内訳 (対訳数)

|       | 学習用    | 開発用   | 評価用   | 合計     |
|-------|--------|-------|-------|--------|
| 全量版   | 34,527 | 4,316 | 4,315 | 43,158 |
| 長文除去版 | 27,641 | 3,455 | 3,455 | 34,551 |

### 5.2 実験結果と考察

表 8 各実験設定における BLEU スコア (いずれも平均値)

|       | カスタム版  | ベースライン | 上昇量    |
|-------|--------|--------|--------|
| 全量版   | 25.572 | 19.631 | +5.941 |
| 長文除去版 | 26.022 | 19.781 | +6.241 |

通常の GNMT をベースラインとし、GNMT に分野適応を施したカスタム翻訳モデルと BLEU スコアを比較した。この結果を表 8 に示す。

まず、カスタム翻訳モデルではどちらの対訳コーパスを用いた場合でも 6 ポイント程度の BLEU スコアの上昇が見られた。また、長文除去版コーパスによるカスタム翻訳モデルの方が BLEU スコアの上昇量が大きい。

評価用に用いた和文の一部を、長文除去版コーパスから作成したカスタム翻訳モデル及び通常の GNMT に入力した結果を表 9 に示す。例 1、例 2 の入力文には、それぞれ専門用語と固有名詞が含まれているものの、カスタム翻訳モデルの訳出結果では正しく翻訳できていることが確認できる。また、例 3 の訳出結果を見ると、ベースラインでは「最重要課題」を「the most important tasks」と英訳しているところ、カスタム翻訳モデルでは「the most important issues」となっており、より文脈に適した英訳となっている。

一方で、本実験で作成したカスタム翻訳モデルでは翻訳品質の向上が見られなかった例を次に挙げる。例 4 には、3.2 節に例示した「同氏」を含んでおり、参照訳では「she」となっているところ、ベースライン及びカスタム翻訳モデルの訳出結果では「he」となっている。例 5 には、3.4 節に例示した「等」を含んでおり、全文体ではカスタム翻訳モデルの訳出結果はベースラインより改善しているものの、「等」で列挙している語句の範囲がどちらも参照訳と異なっている。例 6 には、3.5 節に例示した「(イ)」を含んでおり、参照訳では「(ii)」となっているところ、ベースライン及びカスタム翻訳モデルの訳出結果では「(a)」となっている。例 7 には、3.6 節に示した造語・詩的な表現を含んでおり、ベースライン及びカスタム翻訳モデルの訳出結果では、入力文のニュアンスが失われている。これらの正しい英訳には前後の文脈や各単語の前提知識が必要となるため、一文ごとに機械翻訳を行う NMT モデルにおける分野適応では解決が難しいと考えられる。

## 6 おわりに

本研究では、既存の CG 報告書を用いて対訳コーパスを構築し、それを用いたカスタム翻訳モデルを評価することで、CG 報告書における機械翻訳の活用可能性を検証した。実験により、専門用語や固有名詞をはじめ

表9 機械翻訳モデルの訳出結果の例

|     |                               |  |
|-----|-------------------------------|--|
| 例 1 | 入力文<br>参照訳<br>ベースライン<br>カスタム版 | 株式会社 I C J の提供する議決権電子行使プラットフォームに参加しております。<br>The Company is participating in the platform for the electronic exercise of voting rights operated by ICJ, Inc.<br>We participate in the platform for exercising voting rights electronics provided by ICJ Co., Ltd.<br>The Company participates in the electronic voting platform provided by ICJ, Inc.  |
| 例 2 | 入力文<br>参照訳<br>ベースライン<br>カスタム版 | 2016年度において、当社の会計監査人は新日本有限責任監査法人であります。<br>The Accounting Auditor of the Company for the year ended March 31, 2017 is Ernst & Young ShinNihon LLC.<br>In fiscal 2016, our accounting auditor is New Japan limited liability audit corporation.<br>In fiscal 2016, the Company's Accounting Auditor was Ernst & Young ShinNihon LLC.  |
| 例 3 | 入力文<br>参照訳<br>ベースライン<br>カスタム版 | 当社グループは、企業価値の最大化を目指し、コーポレート・ガバナンスの徹底を最重要課題の一つと位置付け様々な施策を講じています。<br>The Group is aiming to maximize corporate value, and has been implementing various measures as rigorous corporate governance is our highest priority.<br>The Group aims to maximize corporate value and positions thorough corporate governance as one of the most important tasks and takes various measures.<br>The Group considers thoroughness of corporate governance as one of the most important issues with the aim of maximizing corporate value, and takes various measures.  |
| 例 4 | 入力文<br>参照訳<br>ベースライン<br>カスタム版 | さらに同氏は取締役会議長として、当社グループの経営の基本方針等について、取締役会としての決議に向け議案審議を主導いたしました。<br>Furthermore, as the Chairman of the Board of Directors, she led the Board of Directors to make decisions on proposals, including a proposal for basic management policy of the Group.<br>In addition, as Chairman of the Board of Directors, he led the deliberation on the agenda for resolutions as the Board of Directors regarding the Group's basic management policies and others.<br>In addition, as Chairman of the Board of Directors, he led the deliberation of the Board of Directors on the basic policy on the management of the Group. |
| 例 5 | 入力文<br>参照訳<br>ベースライン<br>カスタム版 | 外部識者による講演会の開催、社内 WEB サイトでの情報発信、座談会実施等による啓発活動<br>Holding awareness campaigns through round-table discussions, publishing of information on the Company's internal website, and hosting lectures by visiting experts.<br>Held lectures by outside experts, disseminate information on internal website and raise awareness through implementation of round-table discussion etc.<br>Held lectures by outside experts, disseminate information on the internal website, and conduct awareness-raising activities such as holding round-table talks.   |
| 例 6 | 入力文<br>参照訳<br>ベースライン<br>カスタム版 | 最近1年間において、(イ)または当社の業務執行者に該当していた者<br>A person who falls into (ii) or was an executor of business for the Company in the last year.<br>In the past year, (a) or those who corresponded to the Company's business executor.<br>A person who has fallen under either of the above (a) or an executive of the Company's business for the past year.   |
| 例 7 | 入力文<br>参照訳<br>ベースライン<br>カスタム版 | 当社グループの経営理念は、「いつも、人から。」です。<br>The Group's management philosophy is "Putting People First."<br>The management philosophy of the Group is "always from people."<br>The Group's management philosophy is "always from a person."  |

めとする英訳の改善によって、BLEU スコアの上昇を確認した。本研究で構築した対訳コーパスは比較的少量であることから、この拡充と品質の向上によって、翻訳品質のさらなる向上が期待される。ただし、前後の文脈等を踏まえた翻訳が必要な文については、前処理や後処理等が必要であると考えられるため、これについても今後調査を行いたい。

以上を踏まえ、今後の課題としては次が挙げられる。

- 機械翻訳の課題
  - 他の機械翻訳サービス等による検証
  - 各記載項目に XML 形式のタグが付与されていることから、各項目ごとに最適な機械翻訳モデルの検討
- 外部の課題
  - 証券・IR 分野における対訳コーパスの品質向上及び拡充
  - 証券・IR 分野における用語集及びスタイルガイド等の検討
  - 特定分野の対訳コーパスの構築に即した作業依頼内容の検討
  - NMT モデルに適した和文の書き方等の体系的な調査

## 謝辞

本研究において、グーグル合同会社の小田悠介氏に有益なご助言を戴いた。ここに記して謝意を表する。

## 参考文献

- [1] 株式会社東京証券取引所, 株式会社名古屋証券取引所, 証券会員制法人福岡証券取引所, 証券会員制法人札幌証券取引所. 2017年度株式分布状況調査結果の概要.
- [2] 株式会社 東京証券取引所情報サービス部. 投資部門別売買状況. <https://www.jpix.co.jp/markets/statistics-equities/investor-type/00-02.html>, 2017. (参照 2019-01-15).
- [3] 投資家フォーラム. 投資家フォーラム-第1・2回会合-報告書. <https://investorforum.jp/>. (参照 2019-01-15).
- [4] 株式会社東京証券取引所. 東証上場会社 コーポレート・ガバナンス白書, 2017.
- [5] 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.
- [6] Philipp Koehn and Rebecca Knowles. Six challenges for neural machine translation. In *Proceedings of the First Workshop on Neural Machine Translation*, pp. 28-39. Association for Computational Linguistics, 2017.
- [7] 株式会社みらい翻訳. 職場英語力を TOEIC 900 点相当に引き上げる日英双方向機械翻訳サービスをリリース. <https://mirai.translate.com/uploads/2017/12/befdf2e9eca64235a2042cd9f50a3db.pdf>, 12 2017. (参照 2019-01-15).
- [8] Translation API - Dynamic Translation | Translation API. <https://cloud.google.com/translate/>. (参照 2019-01-15).
- [9] Kishore Papineni, Salim Roukos, Todd Ward, and Wei jing Zhu. Bleu: a method for automatic evaluation of machine translation. pp. 311-318, 2002.
- [10] AutoML Translation | Google Cloud. <https://cloud.google.com/translate/automl/docs/>. (参照 2019-01-15).
- [11] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, Vol. abs/1609.08144, , 2016.