

ビジネスシーン対話対訳コーパスの構築と対話翻訳の課題

中澤 敏明 李 凌寒 Matīss Rikters
 東京大学

{nakazawa, li0123, matiss}@logos.t.u-tokyo.ac.jp

1 はじめに

機械翻訳の研究には対訳コーパスが必須であり、今や様々な言語対・ドメイン・サイズのデータが自由に使える形で公開されているが、それらの多くは書き言葉を対象としている。例えば Web のクローラデータ¹⁾[1]、ニュース解説²⁾、特許文書 [2]、科学技術論文 [3] などでは全て書き言葉である。話し言葉を対象としたデータもいくつか存在するが、一人が喋り続ける TED のようなモノログ形式のもの [4, 5] であったり、複数名の対話であっても多くのノイズを含んでいたり [6, 7] と、自由に使えるクリーンな対話の対訳コーパスはほとんど存在しないのが現状である。WMT³⁾、IWSLT⁴⁾、WAT⁵⁾などの機械翻訳の評価ワークショップでも書き言葉やモノログ、ノイズを含む対話を翻訳対象としている。唯一、IWSLT で用いられているデータの一つにスペイン語と英語間のクリーンな対話対訳コーパス [8] があるが、このようなコーパスは希少である。

書き言葉やモノログの機械翻訳精度は利用可能な対訳コーパスの増強やニューラルネットワーク技術の発展により非常に向上したが、複数名の対話の機械翻訳についてはまだ精度向上の余地が多く残されている。典型的な例は代名詞補完で、日本語のように代名詞が省略可能な (pro-drop) 言語から英語のように省略できない言語に翻訳する場合に、より自然な訳文とするためには適切な代名詞を補って訳出する必要がある。代名詞の省略は書き言葉よりも話し言葉の方が圧倒的に多いため、対話の翻訳においてはこの問題は重要である。このような問題を解決するために文脈 (ここでは翻訳対象文よりも前の文を指す) を考慮した翻訳モデル [9, 10, 11] が提案されているが、文脈付きの対話対訳コーパスはノイズの

多い OpenSubtitles⁶⁾や JESC⁷⁾などを除くとほとんど存在しない。

そこで我々は新たに日英ビジネスシーン対話対訳コーパス (The Business Scene Dialogue (BSD) corpus) を構築し、その一部を公開した⁸⁾。ライセンスは Creative Commons Attribution-NonCommercial-ShareAlike (CC BY-NC-SA) であるので、商用でなければ自由に利用できる。公開したデータ量やフォーマットの詳細はホームページや本稿の付録を参照していただきたい。

BSD コーパスは他の言語資源を参照して構築されたものではなく、一から構築されたものである。「ビジネス」における「対話」を対訳コーパスのドメインとして選択した理由は二つあり、一つはビジネスシーンがドメインとして広すぎず狭すぎず適当な粒度であると考えた点で、もう一つはノイズの少ない対話の対訳コーパスが機械翻訳研究の新たな方向を開くのに有用と考えた点である。対話のコーパスであるので、機械翻訳だけでなくもちろん対話の研究にも役立つのではないかと考える。

本稿では BSD コーパスの紹介と、これを利用した WAT2020[12] 翻訳タスクの結果の紹介、翻訳結果の誤り分析から得られた対話翻訳の課題を報告する。

2 BSD コーパスの概要

2.1 特徴

BSD コーパスは「シナリオ」を大きな単位として用いている。ビジネスにおける様々なシチュエーションを想定し、そのシチュエーションで行われる一連の対話を一つのシナリオとして収録している。各シナリオには 6 種類のタグからなるシーン情報とシナリオタイトル (より詳細な状況説明) が付与されている。これらの情報を用いることで、例えば対話

1) <https://commoncrawl.org/>

2) <http://www.statmt.org/wmt20/translation-task.html>

3) <http://www.statmt.org/wmt20/>

4) <https://iwslt.org/2021/>

5) <http://lotus.kuee.kyoto-u.ac.jp/WAT/>

6) <https://www.opensubtitles.org/>

7) <https://nlp.stanford.edu/projects/jesc/>

8) <https://github.com/tsuruoka-lab/BSD>

の文脈からだけでは難しいような訳語の選択(日本語の「すみません」を英訳する場合など)が行えたり、将来的な実利用の場面ではカメラにより周囲の状況(シーンや会話のシチュエーション)を認識しつつ適切な翻訳を行うといったマルチモーダル翻訳の実現も考えられる。また各発話には話者の情報も付与されている。

2.2 構築手順

BSD コーパスは 1) ビジネスシーンの選定、2) シーンに沿った日本語および英語での会話シナリオ作成、3) シナリオの英語および日本語への翻訳の 3 つのステップを経て構築した。なおシナリオ作成には 28 名の日本語シナリオライターおよび 23 名の英語シナリオライターが携わり、日英翻訳は 8 名、英日翻訳は 13 名の翻訳者によって行われた(シナリオ作成と翻訳を兼ねる者も数名いた)。また全体的な質を担保するために、3 名が全体の品質チェックを行った。さらに、シナリオがビジネスシーンとして適切であることを保証するために、広く一般に知られているメディア等でビジネス会話の関連したコンテンツに主として関わった経験があり、日本語、英語ともビジネスシーンにおいて問題なく会話ができ、自らもビジネスに携わった経験を保有しているものを監修者として 1 名置き、全体的な内容の確認を行った。

1) ビジネスシーンの選定

ビジネスシーンは会議、交渉、雑談など様々なビジネスの状況をカバーするように慎重に選定した。また特定の業種にしか該当しないようなシーンは避け、様々な業種で汎用的に用いられるようなシーンを選定した。今回は「対面对話(2名での仕事上の対話)」「電話対応」「雑談」「会議(複数名での対話)」「研修」「プレゼン」の 6 つのシーンを選定した。

2) 単言語でのシナリオ作成

対話のシナリオは選定されたビジネスシーンに沿って、日本語および英語それぞれ半数ずつ、単言語で作成した。半数ずつとした理由は、ビジネスシーンにおいて各言語で特徴的に利用されるような表現をなるべく幅広くカバーするためである。一方の言語でのみシナリオを作成しこれを相手言語に翻訳するだけでは、表現のバリエーションが限られてしまう(いわゆる translationese) 可能性があり、これ

を回避することを狙っている。

3) シナリオの翻訳

作成された単言語のシナリオは人手で相手言語に翻訳した。機械翻訳の利用は禁止とし、正確性だけでなく相手言語における流暢性や会話としての自然さも重視するよう指示した。例えば英語を日本語に翻訳する場合には、逐語的に訳すと英文中の全ての代名詞を日本語文に訳出することになるが、日本語対話においては多くの代名詞(特に一人称と二人称)は省略されることが普通である。逆に主語の省略された日本語を英語に訳す場合には、無理に受動態で翻訳することは避け、適当な主語を補完した自然な能動態として翻訳するよう指示した。

3 WAT2020 での BSD 翻訳タスク

3.1 タスク概要

WAT2020 において BSD コーパスを利用した翻訳タスクを設定した。参加者は BSD コーパスのテストデータの翻訳結果を提出することが求められる。モデルの訓練には BSD コーパスの訓練データだけでなく、他の既存の対訳コーパス(外部データ)を用いることもできる。日英翻訳には 4 チーム、英日翻訳には 3 チームの参加があった。

3.2 人手評価

人手評価基準として特許庁が公開している「特許文献機械翻訳の品質評価手順⁹⁾」の中の「内容の伝達レベルの評価」を採用した。これは機械翻訳結果が原文の実質的な内容をどの程度正確に伝達しているかを、参照訳の内容に照らして 5 段階(評価値 5 が最もよく、1 が最も悪い)の評価基準で主観的に評価するものであり、いわゆる Adequacy(正確性)評価と同等のものである。正確性評価は合計の文数が 400 文になるようにテストデータからランダムにシナリオを選択したものを対象として行い、各文を 2 名の評価者が独立に評価した。また文脈を考慮して適切な訳となっているかという点も考慮に入れている。

図 1 に正確性評価結果を示す。WAT のポリシーに従いチーム名は匿名化して表示しているが、各グラフで同じチーム名のものは実際に同じチームからの投稿であったことを示している。なお Team B

9) https://www.jpo.go.jp/system/laws/sesaku/kikaihonyaku/tokkyohonyaku_hyoutka.html

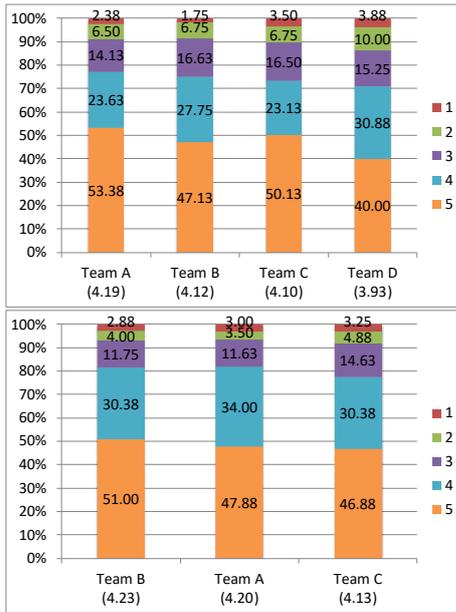


図1 日英(上)および英日(下)BSD 翻訳タスクの正確性評価結果(グラフ内の数字は各評価値の割合を、チーム名の下の括弧内の数字は評価の平均を表す)

は本稿の著者らの結果である。両方向とも評価の平均値は最も高いもので4.2程度であり非常に高精度で訳されていると思われるが、BSD コーパスの性質上、相槌など非常に短い文なども含まれているため、4.2という値がどの程度満足のいくものなのかはより深く分析する必要があると思われる。

3.3 各チームの手法と知見

評価結果はどのチームも近い値であったが、用いた翻訳モデルや外部データは異なっている。Team Aは翻訳対象文と直前の3文を文脈として同時に使ったmBART[13]を用いている。他のチームは文脈情報は用いず、単文での翻訳だった。どのチームも大規模な外部データを併用して翻訳モデルの訓練を行い、BSDなどの対話コーパスでfine-tuningするという戦略を取った。

興味深い知見として、BSD コーパスのみを用いたfine-tuningは効果がなく、むしろ翻訳精度が悪化すると多くのチームから報告があった。しかしながら、BSDに加えて外部の対話データ(OpenSubtitlesやJESC)も併用したfine-tuningは効果的で、翻訳精度が向上するとも報告された。このことから機械翻訳ではfine-tuningにおいてもある程度の量の対訳コーパスが必要であると結論づけられる。

またTeam Dからの報告では、sub-wordの語彙サイズを小さく(報告では6,000を用いていた)する方

表1 翻訳誤りの傾向調査結果

分類	誤訳	訳抜け	過剰訳	文脈上不適切	代名詞誤り	合計
日英	45	7	1	66	32	151
英日	10	7	0	115	0	132

が翻訳精度が向上することだった。はっきりとした理由は不明だが、会話文においては使われる語彙がニュース記事や科学技術論文などと比べてそれほど大きくないためである可能性がある。この点については今後深く分析する必要がある。

4 対話翻訳における課題

現状の対話翻訳においてどのような課題があるのかを分析するために、WAT2020のBSD翻訳タスクでのTeam B(本稿の著者ら)の翻訳結果について、翻訳誤り傾向を調査した。誤りのタイプを「誤訳(文脈に関係なく誤り)」「訳抜け」「過剰訳」「文脈上不適切」「代名詞誤り」の5種類に分類し、2名の評価者で分類が一致したものを計数したものを表1に示す。なお1つの文で複数の分類が付与される場合もあるため、数字は延べ数である。ここでは特に「文脈上不適切」「代名詞誤り」の2つについて例を示しつつ分析を行う。

4.1 文脈上不適切

文脈上不適切な訳になっている例を図2および図3に示す。図2は文脈情報を使うことで適切な訳が出力できるであろう例で、実際に一つ前の文を連結して用いる手法[9, 14]を用いて翻訳を行うと一番右の列のように翻訳が改善することがわかる。一方、図3に示す例は単に前の文の情報を用いだけでは解決が難しく、ぼんやりと会話全体の流れをとらえ、さらにその流れに沿って適切な表現を選択する必要がある。これにはmBARTのような事前学習モデルを利用することで解決が期待できるため、まずはTeam Aの翻訳結果の分析を行う予定である。

4.2 代名詞誤り

代名詞の誤り出誤りは日英翻訳における典型的な翻訳誤りであるが、その解決に必要な情報源には我々が観測した範囲では三つのタイプが存在する。一つ目は文脈文に出現する名詞等の情報を用いたり、目的言語側の過去の出力を利用[15]するなどすれば解決可能なものである。二つ目は図4に示すように、文脈文を用いずとも当該文の動詞に後続する

話者	入力	正解	単文出力	+文脈出力
受付係	お名前頂戴してもよろしいでしょうか？	Could I have your name please?	May I have your name?	May I have your name?
高岡	はい、高岡渉と申します。	Sure, it's Wataru Takaoka.	Hello, my name is Wataru Takaoka.	Yes, my name is Wataru Takaoka.
受付係	ありがとうございます。	Thank you.	Thank you very much.	Thank you.
受付係	神田は会議室で待っております。	Kanda is expecting your visit in the meeting room.	Kanda is waiting in the conference room.	Kanda is waiting in the meeting room.
受付係	ご案内いたします。	Please follow me.	Let me show you.	I'll take you to the meeting.
高岡	ありがとうございます。	Thank you.	Thank you very much.	Thank you.
受付係	こちらでございます。	Here it is.	This is it.	Here you are.

図2 日英翻訳における文脈上不適切な訳の例(文脈文を用いることである程度解決が可能)

話者	入力	正解	出力
Yutaka	To be honest with you, I am a little bit nervous.	正直なところ、少し緊張しているんだ。	正直なところ、ちょっと緊張しています。
Dennis	It's understandable.	それは当然だよ。	それは理解できます。
Yutaka	It is going to be a big adjustment.	馴れるまで大変だろうな。	大きな調整になります。
Yutaka	The new office has twice as many employees as this one.	新しいオフィスにはこの倍の従業員がいるんだよ。	新オフィスの従業員数は、このオフィスの倍です。
Yutaka	It seems like a competitive environment.	競争が激しそうだし。	競争環境みたいですね。
Dennis	Don't feel discouraged.	弱気にならないで。	落胆しないでください。

図3 英日翻訳における文脈上不適切な訳の例(全体的な会話の流れを掴んだり、一般常識等を用いる必要がある)

話者	入力	正解	出力
	多国籍チームだけの話じゃない、営業グループ全体について言ってるんだよ。	I'm not just talking about Multinational team, I'm talking about the entire Sales Group.	It's not just about Multinational team, it's talking about the whole sales group.
	それが何か意味あることだと思ってるのさ。	She thinks it's an indication of something.	I think that's something meaningful.
	その方向で進めて、佐藤部長に相談してみようかな。	Let's go with that and maybe I'll discuss this with Sato-san.	Maybe we can proceed with that and discuss it with Mr. Sato.

図4 原文の機能語列の情報を利用して省略された代名詞が推測可能な例

話者	入力	正解	出力
左枝さん	体調が1番だから、これは他のものに任せていいよ。	Your health is the most important so others can do these.	You're in top shape so you can leave this to others.
木内さん	ありがとうございます。	Thank you very much.	Thank you.
左枝さん	産休の申請もしないといけないな。	You will need to apply for maternity leave too.	I need to apply for maternity leave.
左枝さん	6月出産なら4月から産休がいいのかな？	If you're due in June, then you will take maternity leave from April?	Should I take maternity leave from April if I give birth in June?
木内さん	はい、それぐらいからお休みいただくと助かります。	Yes, it would be great if I can take my leave from around that time.	Yes, that would be great if you can take some time off.

図5 正しい代名詞を補完するために話者情報が必要となる例(3文目や4文目は木内さんの発言の可能性もある)

機能語列から代名詞が推測可能なものであり、例えば Kudo ら [16] の手法などにより解決可能なものである。三つ目は図5に示すように、話者情報が必要となるものである。例えば3文目は木内さんの発言としても解釈可能であり、その場合は出力文の方が正しい訳となる。

5 まとめ

本稿では BSD コーパスおよび WAT2020 での BSD 翻訳タスクの結果の紹介を行い、翻訳結果の分析から対話翻訳の課題について議論した。公開した BSD コーパスがこれらの課題の解決の一助となれば幸いである。BSD を利用した翻訳タスクは WAT2021¹⁰⁾ においても設定しているので、興味のある方はぜひ参加していただきたい。WAT2021 では新たな開発

データとテストデータを追加で公開予定である。

また本稿では触れていないが、AMI Meeting Corpus[17] を和訳することで対訳コーパスとしたデータも公開している¹¹⁾ ので、合わせて利用していただければと思う。

謝辞

本研究成果は独立行政法人情報通信研究機構(NICT)の委託研究「多言語音声翻訳高度化のためのディープラーニング技術の研究開発」により得られたものです。

またビジネスシーン対話対訳コーパスの構築にあたっては、株式会社バオバブ様に設計の段階からご協力いただきました。ここに御礼申し上げます。

10) <https://lotus.kuee.kyoto-u.ac.jp/WAT/WAT2021/>

11) <https://github.com/tsuruoka-lab/AMI-Meeting-Parallel-Corpus>

参考文献

- [1] Makoto Morishita, Jun Suzuki, and Masaaki Nagata. JParaCrawl: A large scale web-based English-Japanese parallel corpus. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pp. 3603–3609, Marseille, France, May 2020. European Language Resources Association.
- [2] Isao Goto, Bin Lu, Ka Po Chow, Eiichiro Sumita, and Benjamin Tsou. Overview of the patent machine translation task at the ntcir-9 workshop. In *Proc. of NTCIR-9 Workshop Meeting*, pp. 559–578, 12 2011.
- [3] Toshiaki Nakazawa, Manabu Yaguchi, Kiyotaka Uchimoto, Masao Utiyama, Eiichiro Sumita, Sadao Kurohashi, and Hitoshi Isahara. Aspec: Asian scientific paper excerpt corpus. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2016)*, pp. 2204–2208, Portorož, Slovenia, may 2016. European Language Resources Association (ELRA).
- [4] Mauro Cettolo, Christian Girardi, and Marcello Federico. Wit³: Web inventory of transcribed and translated talks. In *Proceedings of the 16th Conference of the European Association for Machine Translation (EAMT)*, pp. 261–268, Trento, Italy, May 2012.
- [5] Mattia Antonino Di Gangi, Roldano Cattoni, Luisa Bentivogli, Matteo Negri, and Marco Turchi. MuST-C: a Multilingual Speech Translation Corpus. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, Minneapolis, MN, USA, June 2019.
- [6] Jörg Tiedemann. Finding alternative translations in a large corpus of movie subtitle. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pp. 3518–3522, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA).
- [7] R. Pryzant, Y. Chung, D. Jurafsky, and D. Britz. JESC: Japanese-English Subtitle Corpus. *Language Resources and Evaluation Conference (LREC)*, 2018.
- [8] Elizabeth Salesky, Susanne Burger, Jan Niehues, and Alex Waibel. Towards fluent translations from disfluent speech. In *Proceedings of the IEEE Workshop on Spoken Language Technology (SLT)*, Athens, Greece, December 2018.
- [9] Jörg Tiedemann and Yves Scherrer. Neural machine translation with extended context. In *Proceedings of the Third Workshop on Discourse in Machine Translation*, pp. 82–92, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.
- [10] Elena Voita, Pavel Serdyukov, Rico Sennrich, and Ivan Titov. Context-aware neural machine translation learns anaphora resolution. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1264–1274, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [11] Elena Voita, Rico Sennrich, and Ivan Titov. When a good translation is wrong in context: Context-aware machine translation improves on deixis, ellipsis, and lexical cohesion. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 1198–1212, Florence, Italy, July 2019. Association for Computational Linguistics.
- [12] Toshiaki Nakazawa, Hideki Nakayama, Chenchen Ding, Raj Dabre, Shohei Higashiyama, Hideya Mino, Isao Goto, Win Pa Pa, Anoop Kunchukuttan, Shantipriya Parida, Ondřej Bojar, and Sadao Kurohashi. Overview of the 7th workshop on Asian translation. In *Proceedings of the 7th Workshop on Asian Translation*, pp. 1–44, Suzhou, China, December 2020. Association for Computational Linguistics.
- [13] Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, Vol. 8, pp. 726–742, 2020.
- [14] Rico Sennrich and Barry Haddow. Linguistic input features improve neural machine translation. In *Proceedings of the First Conference on Machine Translation: Volume 1, Research Papers*, pp. 83–91, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [15] Hideya Mino, Hitoshi Ito, Isao Goto, Ichiro Yamada, and Takenobu Tokunaga. Effective use of target-side context for neural machine translation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 4483–4494, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics.
- [16] Taku Kudo, Sora Ichikawa, and Hideto Kazawa. Language independent null subject prediction for statistical machine translation. *Journal of Natural Language Processing*, 2015.
- [17] Iain McCowan, Jean Carletta, Wessel Kraaij, Simone Ashby, S Bourban, M Flynn, M Guillemot, Thomas Hain, J Kadlec, Vasilis Karaiskos, et al. The ami meeting corpus. In *Proceedings of the 5th International Conference on Methods and Techniques in Behavioral Research*, Vol. 88, p. 100, 2005.

表 2 公開したビジネスシーン対話対訳コーパスの統計量(シナリオ数/文数)

シーン	訓練データ		開発データ		テストデータ	
	日英	英日	日英	英日	日英	英日
対面対話	122 / 3,525	103 / 2,986	11 / 319	12 / 314	12 / 381	11 / 345
電話応対	68 / 1,944	75 / 2,175	6 / 176	7 / 185	6 / 163	7 / 212
雑談	61 / 1,915	72 / 1,883	7 / 223	8 / 248	7 / 211	8 / 212
会議	56 / 1,964	58 / 1,787	7 / 240	7 / 219	7 / 228	7 / 229
研修	12 / 562	19 / 463	1 / 40	1 / 23	1 / 38	1 / 30
プレゼン	6 / 607	18 / 189	1 / 31	1 / 33	1 / 31	1 / 40
合計	325 / 10,000	345 / 10,000	34 / 997	35 / 1,054	34 / 1,052	35 / 1,068

起点言語：日本語 シーン：電話応対 シナリオタイトル：伝言への折り返し電話 打ち合わせ日程調整

日本語		英語	
話者	内容	話者	内容
土井さん	はい、K社システム開発部です。	Doi-san	Hi this is the systems development department of Company K.
高市さん	H社の高市と申します。	Takaichi-san	My name is Takaichi from Company H.
高市さん	いつもお世話になっております。	Takaichi-san	Thank you as always.
土井さん	こちらこそ、お世話になっております。	Doi-san	Thank you as always as well.
高市さん	稲田さんはいらっしゃいますか？	Takaichi-san	Is Inada-san there?
高市さん	1時間ほど前に、お電話いただいたそうなんですけど。	Takaichi-san	It seems I got a call from him around 1 hour ago.
...

図 6 BSD コーパスの対訳例

A 公開データ数とシーンごとの内訳

表 2 に公開したデータのシナリオ数および文数、またシーンごとの内訳を示す。訓練データは合計 2 万対訳文で、開発データとテストデータはそれぞれ約 2,000 文ずつである。なお WAT での利用を想定して、毎年新たな開発データとテストデータを同量程度ずつ追加公開する予定である。

B BSD 対訳例

図 6 に BSD コーパスの対訳例を示す。なおデータは json 形式で配布されている。

C BSD コーパス構築の詳細仕様

BSD 構築における詳細な仕様のうち、主だったものを以下に示す。

- 口語でそのまま発音しない略語 (NY など) は使わず、会話で使われる表記を用いる (New York など)
- 時刻は数字ではなく文字表記で訳出し、「:」は使用しない
- 翻訳時にあまりにも冗長であったり不自然な場合は訳文を 2 文にしてもよい
- 1 シナリオあたり対話は 3 ターン (6 文) 以上、上限は 20 ターン (40 文) まで
- 「プレゼン」と「研修」はモノログ (登場人物が 1 名のみ) でもよい
- 日本語では標準語を英語ではアメリカ英語を極力用いること (方言やスラングなど極端にくだけた表現は避ける)
- 1 シナリオ内の登場人物は 4 人以下とする
- 個別の実在する会社名の使用は避け、A 社、B 社などを用いる (ただしサービス名としての Facebook や twitter、インスタなどの使用はよしとする)
- 対面ではない発話の状況を示すような情報、また話者の役職、肩書等は、本文とは別に「付属情報」に記載する