

Airbnbでの機械翻訳のユーザー体験事例—シンプルな翻訳品質基準の必要性

A user-experience case of machine translation use in Airbnb – the necessity of a simple translation quality standard



秋桜舎代表

山本 ゆうじ

筑波大学を経てシカゴ大学修士号。企業向けに、大規模翻訳・文書管理／作成、日本語作文、英語の講習やコンサルを行う。近著に『IT時代の実務日本語スタイルブック—書きやすく、読みやすい電子文書の作文技法』（2012、ベレ出版）。

✉ <https://cosmoshouse.com/>（連絡用フォームから）

1 はじめに

筆者はこれまで文書での機械翻訳の活用に関わってきた。昨今は特に機械翻訳の活用が様々な場面に広がっていることを実感する。だがそれと同時に、翻訳品質を改善するために、シンプルな翻訳品質基準は、翻訳のエンドユーザー、翻訳者、翻訳会社、翻訳依頼者など翻訳に関わるすべての人にとって重要性を増しつつある。ウェブサイトや海外の町中では、依然として奇妙な日本語を見ることが多い。また英語や他の言語についても似た状況だろう。これらの少なくとも一部は機械翻訳によるものであろう。機械翻訳の品質が上がり、ポストエディットも広まった今必要なのは、機械翻訳と人間翻訳のどちらかにしか使えない翻訳品質基準ではなく、両方に使えるものである。

本稿では、Airbnb を例に、ユーザーの立場から見た機械翻訳の活用とポストエディットについて論考する。

2 Airbnb ゲストとして機械翻訳を使う

Airbnb は、民家を宿泊の場として提供するサービスである。同種のサービスは他にもあり、まとめて「民泊」とも呼ばれる。宿泊者「ゲスト」を泊める側の「ホスト」は素人であり、友人の家に泊まるのにも似ている。ホテルと同じ行き届いたサービスが受けられるわけではない。訪れる現地の人々の暮らしを疑似体験する、画一的なホテルでは得られない経験ができる。場合によってはホテルよりもコスト面で有利なこともあるが、価格はホ

ストの設備や場所などに依存する。

筆者は2018年8月にローマ、ヴェネツィア、パリ、ロンドンの四都取材旅行した。ローマ、ローマ近郊の街のティヴォリ、ヴェネツィア、ロンドンの4箇所までゲストとして Airbnb を使用した。Airbnb を使うのは、今回の旅が初めだった。Airbnb を使うこと自体が旅の目的ではなかったが、オンラインを通じて行う共有経済の一つの形として関心があった。

当初、ホストとのやり取りは挨拶ぐらいで済むのではないかと漠然と考えていた。だが、ゲストとホストのコミュニケーションは単なる挨拶ではなく、必須であることにすぐに気づいた。チェックインの方法は物件により様々であるし、ホストがゲストを直接出迎える場合は、ホストの都合も確認しておく必要がある。24時間フロントが開いているホテルではないから、到着時間はホストと相談しながら決める。到着が遅れる場合や場所が分からない場合はホストに直接尋ねる必要がある。到着する場所や時間を間違えると、異国の地で重い荷物を抱えたまま途方に暮れることになる。Airbnb の担当者がいちいちサポートしてくれるわけではない。

筆者は、英語とフランス語を使ううえでは困ることはない。だが、ティヴォリの宿を予約した時には、ホストが当然のようにイタリア語でメッセージを送ってきたのでやや驚いた。

Airbnb のシステムには機械翻訳（Google 翻訳）が組み込まれており、メッセージはその場で必要な言語に翻訳できる（次頁図）。

様々な国からのゲストが様々な国のホストを訪れるわ



この会話を日本語に翻訳
powered by Google™

けだから、使用する言語の壁を乗り越える必要があるのは当然である。こちらからは英語で返信しても、ホストは機械翻訳して読むことができる。翻訳結果に不安があるならあらかじめこちらで Google 翻訳したイタリア語のメッセージを送ってもよい。このようにして、英語をほとんど話せないホストともスムーズにやり取りができ、宿泊後のお礼まで機械翻訳を使って伝えた。筆者は商業翻訳に 20 年ばかり機械翻訳を活用してきたが、機械翻訳の有用性を新たな形で実感することとなった。

ヴェネツィアのホストは、英語でメッセージを送ってきた。彼はひと通り英語を話せたが、細かいコミュニケーションとなると Google 翻訳での音声翻訳（つまり音声で入力する、通訳に近い形の翻訳）を活用できる画面が数回あった。

ローマとイギリスのホストたちは、英語でのコミュニケーションに問題はなかったため機械翻訳を使う場面はなかった。今回試してみたのはイタリア語と英語、フランス語のみであり、日本語との間の機械翻訳の有効性を確認することはしなかった。宿泊に必要な情報のやり取り程度のコミュニケーションであれば、日本語との機械翻訳でも問題なく翻訳できであろう。

とはいえ、Airbnb で機械翻訳が必ずしも広く使われているわけでもないらしい。後でイタリア在住の別のホストに聞いたところ、5 年間ホストをしてきたが大半はアメリカ人がイタリア語話者で、機械翻訳を使ったことはほとんどなかったと聞いた。Airbnb でどのように機械翻訳が使われているか統計があれば興味深い。

Airbnb は、必要最小限のコミュニケーションだけで、ほとんど言葉を交わさずに宿泊することもできる。だが、知り合いのない旅人が現地人と接触を持つためのよい接点である。ゲストが望み、時間が許せば、ホストとあれこれ話をすることも可能である。今回の旅では時間的余裕がなかったが、音声翻訳を活用して自分の知らない言語での会話を楽しむのもまた興味深い。

また今回タイのバンコクに立ち寄った。タイ語は固有の文字を使うため、写真撮影による文字認識の機械翻訳を試す絶好の機会であった。だが残念ながらタイで使うはずの SIM カードを預け荷物のほうに入れてしまい、取り出すことができなかったので、実際には試すことができなかった

今回の旅で、結論から言えば、音声翻訳が必要な場面はそう多くはなかった。筆者はスマートフォンを常に手元に置いており、海外用の SIM カードで必要ならすぐに通信できる状態にあった。それにもかかわらず音声翻訳は使わなかった。

音声翻訳が実際に使われるには、以下のような条件が揃う必要がある

1. コミュニケーションが明らかに必要であること
2. 音声翻訳を行うための時間的、精神的余裕があること
3. 音声翻訳を行うための機器およびソフト、通信手段（オンライン型の場合）が揃っていること

相手と直接やり取りできるなら、音声翻訳を使わなくても、ボディー ランゲージや不完全な外国語でも用事が足りることも多々ある。

新宿で一度、中国の観光客が、店の閉店時間を尋ねるために音声翻訳を使っている場面に出くわしたことがある。ただこのような活用はまだ比較的まれであるのかもしれない。音声翻訳用の単体で完結する専用機器も注目されているようだが、より一般的に使われるようになるためには、音声翻訳全体の使い勝手のさらなる向上が必要なのかもしれない。

3 日本での民泊

日本人は、Airbnb のようなシステムは敬遠するようである。外国語でのコミュニケーションに不安があるというだけではない。友人の家であればともかく、見知らぬ他人の家に泊まる、あるいは泊めるということに抵抗があるためであろう。事実、統計上は日本での日本人 Airbnb 利用者は 7% である。とはいえ外国からのゲストに限って言えば、2018 年 2 月の Airbnb 利用宿泊客は 580 万人を突破したとのことで、Airbnb はすでに日本でも大いに活用されているようである。

2020 年の東京オリンピック開催時には、さらに多



数の海外からの訪問客が Airbnb や同様の仕組みを活用して日本に滞在するであろう。民泊では、ゲストとホストどちらの立場であっても、機械翻訳を使ったコミュニケーションに対して、食わず嫌いをするかどうかが大きな違いを生む。

機械翻訳、音声翻訳の普及には、啓蒙活動に加えて、実際の活用法をさらに研究・検討しなければ、せっかくのハードウェア、ソフトウェア、サービスが有効に活かせない。そのためには、翻訳品質についても検討が必要になる。

4 シンプルな翻訳品質基準の必要性

機械翻訳の活用が広まる中で、人間翻訳まで含めたうえでの翻訳品質の議論もまた必要である。

文（センテンス）単位でのシンプルな翻訳品質のレベルとして以下の四つを考えた。

優：読者を把握し、文脈をすべて踏まえた、適切な翻訳。

可：不自然な点はあるが実用可能な翻訳。

劣：概訳（後述）。小さな誤訳を含む。

不可：重大な誤訳、訳抜けなど。

案件によっては、優レベルは不要なため、除外することもありうる。

Airbnb でのゲストとホストのやり取りではそれほど高い翻訳品質は必要ではない。実際的な意思疎通ができれば目的は十分に果たせる。このような、精度や品質が重視されない翻訳を「概訳」と呼ぶことにする。会話や会話に近いメッセージのような一過性のコミュニケーションでは概訳でも十分である。

しかし概訳では不十分な場合がある。たとえば高価な商品やサービスの説明で翻訳品質が不十分であれば信用を低下させる。海外の空港や美術館、博物館などの公共設備で日本語の案内がおかしいことはよくある。素人、あるいは非日本語話者が翻訳したとみられる、こなれない日本語である。このような場合は、可レベルではなく、より自然な優レベルの翻訳が求められるはずである。

機械翻訳の結果を改善するポストエディットには、大別して人間翻訳と同等以上の品質を実現するフルエ

ディットと、翻訳品質を犠牲にしてコストと速度を優先した概訳を行うライトエディットの2種がある。言い換えれば、フルエディットは少なくとも可レベルを目指し、ライトエディットでは劣レベルを目指すことになる。

ライトエディットは可能であることを認めながらも、フルエディットの有効性に疑念を抱く人もいる。しかし実際の総合的な難易度からすると、翻訳業務としてはライトエディットのほうが難しい。フルエディットでは、作業とコストのバランスなどを考慮せず全力でポストエディットすればいい。問題点があればすべて直すだけである。

しかし、十分なコストがかけられないライトエディットでは、想定より手間がかかってしまうことも多い。これが今、多くの翻訳会社、翻訳発注企業、そしてポストエディターが悩んでいる点であろう。

誤訳の有無は、翻訳品質で重要な問題である。しかし機械翻訳に誤訳があるかどうかを簡単に見分ける方法はなく、人間がチェックをするしかない。そもそも機械的に誤訳を見つける方法があるなら、その方法はすでに機械翻訳に含まれているはずである。ポストエディットは、現時点での機械翻訳の制限を人間が補填する手法なのだから、問題のしわ寄せが人間に来て、機械翻訳の尻拭いをすることになる。見合うコストが支払われていない労働を翻訳者に強いると、ポストエディットの案件はやり手がなくなってしまう。

この問題に解決策がないわけではない。後述のように用語データを活用すれば、先の見えないポストエディット作業を半自動化できる。

5 今後の機械翻訳

今後、機械翻訳はどのようになっていくのか。定型的な文章では、機械が確実に人間を置き換えるであろう。

原理的にはビッグデータを活用することにより、一定の法則さえ見出すことさえできれば、ほとんどどのような翻訳でも機械翻訳できるはずではある。場合によっては人間よりも文脈をうまく捉えたかのように見える結果を機械翻訳が出すこともある。だがこれはまだ現時点では例外的と言える。日本語に訳す場合や日本語から訳す場合では、前後の文脈が微妙すぎて機械翻訳では対応できないケースも多々ある。

下手くそな日本語原文でもニューラル機械翻訳なら対応できるようになったことは事実である。だが、人間が作文の能力を鍛える必要性は依然としてある。その意味で産業日本語研究会〈<https://www.tech-jpn.jp/>〉のように日本語作文の品質を向上する取り組みは今後とも重要である。

またニューラル機械翻訳では、用語の一貫性がないことがポストエディットの手間を増やしている。ヤラクゼンのように用語を翻訳結果に反映させることができるサービスは有効である。さらに用語データは機械翻訳・人間翻訳を問わず、用語チェックにも使用できる。

用語データを使うには、事前に用意する必要がある。中・大規模なプロジェクト、または繰り返される案件では、用語データを用意するコストは十分に回収できる。用語データは、AAMT が策定した UTX 用語集形式〈<http://aamt.info/japanese/utx/>〉の仕様に沿っ

て作成することで、簡単に作成でき、またデータ構造を標準化して再利用できる。AAMT は、許諾を得て、特許庁の用語データを UTX 用語集形式に変換して無償で公開した〈<http://aamt.info/japanese/utx/jpo-utx.htm>〉。

本データに含まれる用語の例を示す【表 1】。

このような用語データを活用することでポストエディットをより効率的に行うことができる。

6 まとめ

今回、Airbnb で機械翻訳・音声翻訳をいちユーザーとして使ったのは非常に興味深い体験であった。機械翻訳を活用できなければ、今後、日本のインバウンドは大きな機会損失が見込まれる。精度は向上したものの、ユーザーの抵抗をなくし、即時機械翻訳・音声翻訳を使いやすくするには、多くの実践と啓蒙活動が必要と実感した。

またシンプルな翻訳品質基準と用語データを活用することで、ポストエディットに関する重大な問題を解決する糸口になるであろう。

これらの事柄は今に始まったことではなく、筆者がここ数年訴えてきていることではある。残念ながら翻訳および通訳業界・機械翻訳業界ではこのような事柄はまだ十分に理解・認識されておらず、もどかしさを感じる。業界外部からの新しい風が必要なのかもしれない。

参考文献

- Airstair. “2月のAirbnb利用宿泊客は580万人を突破 物件数は約62,000件に1物件あたりの受け入れでは大阪市が最多”. <https://airstair.jp/airbnb-japan-data/>, (参照 2018-8-30).
- MINPAKU.Biz. “Airbnb (エアビーアンドビー)”. <https://min-paku.biz/list/airbnb>, (参照 2018-8-30).
- AAMT. “UTX 1.20仕様書”. AAMT. <http://www.aamt.info/japanese/utx/utx1.20-specification-j.pdf>, (参照 2018-8-30).

表 1 特許庁 UTX 用語データの例

分野 (正確には意味素性)	項目数	例	
植物 (通称、品名、学名など)	5484	白いぼキュウリ	white spine cucumber
		メラレウカ・アルテルフォリア	Melaleuca Alternifolia
		いらくさ科植物	urticaceous plant
動物 (通称、学名など)	3009	ヤブカ	striped mosquito
		モンシロチョウ	Pieris rapae
		ユーグレナ	Euglena
人 (人名、職位など)	1284	昌聰	Yoshiaki
		調香士	perfumer
		登壇者	presenter
企業・組織名	4454	日本醸造協会	Brewing Society of Japan
		猟友会	hunters' association
		インド技術研究所	Indian Institute of Technology
その他	44875	化学、医学、機械、工学、その他専門用語	
分類情報なし	64726	オキシジフタル酸二無水物	oxydiphthalic dianhydride
		ガス不透過プレート	gas-impermeable plate
		パッシェン曲線	Paschen curve