



AI翻訳字幕 の市場分析

独立技術評価

 slator

2026年3月5日

目次

- 3 要点
- 5 はじめに
- 7 翻訳の品質
 - 7 総合評価
 - 8 重大度の評価
 - 9 言語別評価
 - 11 言語専門家の選好
- 12 字幕の安定性
 - 13 総合評価
 - 13 言語別評価
- 16 結論
 - 16 総合結果
 - 17 企業への示唆
- 18 付録
 - 18 評価方法

要点

1. DeepL Voiceは、翻訳品質と字幕安定性の両面で業界をリード

DeepL Voiceは、人間による言語評価と自動字幕安定性評価の両方で最高得点を獲得しました。Zoom Meetings版DeepL Voiceは96.4/100、Teams版DeepL Voiceは96.3/100の品質スコアを記録しました。一方、他のプラットフォームは87~89にとどまりました。DeepL Voiceの各製品は、最も安定した字幕を生成し、安定性スコアはそれぞれ88.6と85.8を記録しました。

2. DeepLは、重大な翻訳エラーを大幅に削減

すべての言語ペアにおいて、DeepL Voiceは評価対象の他プラットフォームと比べ、深刻または重度の翻訳エラーの発生率を平均で76%削減しました。

DeepL Voiceでは、言語評価を完全に通過したセグメントの割合は79%であったのに対し、競合ツールでは42%にとどまりました。

すべての言語ペアにおいて、DeepL Voiceはセグメントあたりの平均翻訳エラー数を、Microsoft Teamsと比較して66%、Zoomと比較して64%削減しました。

3. 字幕の安定性はプラットフォーム間で大きく異なる

すべてのプラットフォームで、字幕の書き換え（画面上で翻訳文がちらついたり、何度も書き換えられたりする現象）が確認されました。しかし、DeepL Voice製品は字幕の書き換えが最も少ないことが示されました。

すべての言語ペアにおいて、DeepL Voiceは字幕の変動を、Microsoft Teamsと比較して平均37.6%、Zoomと比較して平均54.7%削減しました。

4. 言語専門家はDeepL Voiceを圧倒的に支持

ブラインド評価の結果、言語専門家の96%が、翻訳字幕のプラットフォームとしてDeepL Voice製品を最も高く評価しました。

5. DeepL Voice製品は、AI翻訳字幕の生成機能で業界をリード

言語的正確性と字幕の安定性の両方における結果を統合するため、Slatorは各プラットフォームを以下の2つの評価軸に基づいて分析しました。

1. 翻訳品質（人間による言語評価に基づく）
2. 字幕の安定性（フレーム単位の自動解析に基づく）

これにより、以下の4つのクワドラントが定義されます。

| クワドラント | 説明 |
|-----------|-------------------------|
| リーダー | 翻訳品質・字幕の安定性ともに高い |
| 高精度システム | 翻訳品質は高いが、安定性はやや低い |
| 高効率システム | 翻訳品質はやや低いが、字幕の安定性は比較的高い |
| 開発段階のシステム | 翻訳品質と安定性の両面で性能が低い |

本評価フレームワークにおいて、DeepL Voice製品は「リーダー」クワドラントに位置づけられています。

AI翻訳字幕プラットフォームのクワドラント

言語品質と字幕表示の安定性に基づく、リアルタイム字幕翻訳システムの比較評価。



はじめに

翻訳機能は、エンタープライズ向けツールにおいてますます導入が進んでおり、AIによる翻訳字幕もグローバルな会議用ソフトウェアに組み込まれるようになってきました。Google Meet、Microsoft Teams、Zoomなどのプラットフォームでは、言語の壁を越えたコミュニケーションを可能にするため、リアルタイムの字幕翻訳機能が利用可能です。

しかし、こうしたシステムの普及が進んでいるにもかかわらず、実際の会議での性能を第三者が評価したベンチマークは限られています。リアルタイム翻訳字幕のユーザー体験は、以下の2つの重要な要素に左右されます。

- 翻訳品質：字幕が音声内容の意味を正確に伝えているかどうか
- 字幕の安定性：音声処理中に頻繁な書き換えやちらつきが発生せず、字幕が画面上に安定して表示されるかどうか

字幕の頻繁な更新や部分的な書き換え、あるいは翻訳の揺れは、たとえ最終的な翻訳が正確であっても、[理解を妨げる可能性があります](#)。したがって、言語的正確性と視覚的安定性の両方を測定することで、リアルタイム字幕システムがエンドユーザーに対してどのように機能しているかを、より包括的に把握することができます。

この課題に対処するため、Slatorは以下の5つのプラットフォームにおいて、AI生成による翻訳字幕の独立評価を実施しました。

1. Google Meet
2. Microsoft Teams
3. Zoom
4. Teams版DeepL Voice
5. Zoom Meetings版DeepL Voice

本研究では、14の言語ペアを対象に、翻訳品質と字幕の安定性の両方を評価しました。英語への翻訳および英語からの翻訳として、それぞれ7言語（スペイン語、フランス語、ドイツ語、イタリア語、ポルトガル語、韓国語、日本語）を対象としています。

本研究では、音声認識による文字起こしやバックエンドの翻訳出力ではなく、ユーザーが画面上で実際に目にする字幕を評価しています。Slatorは、画面録画された会議を取得して分析し、レンダリングされた動画フレームから字幕を直接抽出することで、ユーザーが実際に目にする字幕体験を測定しました。

28名の言語専門家が、各プラットフォームの性能を比較するため、翻訳字幕のブラインド評価を実施しました。評価対象にカスタマイズされたDeepL Voiceシステムが含まれていることは言語専門家に知られておらず、評価は字幕の品質と使いやすさのみに基づいて行われました。

各プラットフォームは、Google Meet、Microsoft Teams、Zoomにおける標準の字幕翻訳設定（初期設定のまま）でテストされました。「Teams版DeepL Voice」および「Zoom Meetings版DeepL Voice」は、翻訳用語集など、エンドユーザーが利用可能な製品本来の機能を用いて評価されました。また、「Teams版DeepL Voice」では、固有名詞や専門用語の文字起こしを強化する発話用語認識機能を用いて評価されました。

音声サンプルは、会話形式でビジネス関連のトピックについて議論する2名の話者によるポッドキャストの録音から収集しました。各録音は、言語ごとに約12分間の連続した音声となるよう長さが編集されました。

この評価手法により、自然な対話パターン、ドメイン固有の用語、および専門的な会議に典型的な実際の話し方の特徴を的確に捉えることができました。本評価方法は、同等の条件下で各プラットフォームにおけるユーザーが実際に目にする字幕体験を測定するよう設計されています。

言語間で言語構造が大きく異なるため、翻訳方向によっては他の方向と比べて字幕の修正回数が多くなる場合があります。例えば、日本語や韓国語のように語順が異なる言語や、動詞が文末に置かれる言語では、追加の文脈が得られるにつれて、翻訳システムが字幕を調整する必要が生じる場合があります。そのため、結果は全体集計だけでなく、言語別にも分析されています。

人間による言語評価に加え、Slatorは、表示された字幕のフレームごとの変更を分析することで、字幕の安定性を定量化する自動測定パイプラインを開発しました。

そのパイプラインは次のとおりです：動画記録 > フレーム抽出（10fps）> 字幕領域の切り出し > OCR（光学文字認識）> テキスト正規化 > フレーム比較 > 変更イベントの検出

これらの人間による分析と自動分析を組み合わせることで、AI翻訳による字幕がリアルタイムの多言語会議においてどのように機能するかを包括的に評価できます。

評価方法とプロセスの詳細については、付録をご覧ください。

翻訳の品質

Slatorは、28名の母語話者である言語専門家を用い、14の言語ペア（英語への翻訳7組、英語からの翻訳7組）にわたるAI翻訳の字幕について、ブラインド評価を実施しました。評価方法の詳細については、付録をご覧ください。

テスト対象となったすべての言語およびプラットフォームにおいて、DeepL Voiceは、総合的な品質と重大な翻訳ミスの削減の両面で、**各言語のネイティブ字幕作成ツールを一貫して上回りました。**

結果は以下の通りです。

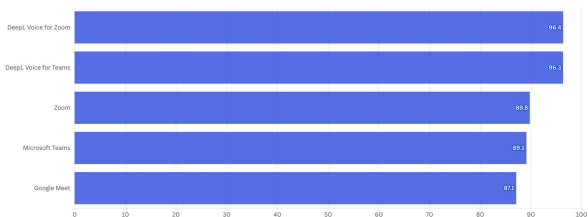
総合評価

Slatorは、評価対象となったすべてのセグメントにおける翻訳エラーの重大度を総合的に反映し、結果を0~100の単一の品質スコアに統合しました。これにより、「5つのプラットフォーム全体におけるAI翻訳字幕の品質はどの程度か？」という問いに答えることが可能となりました。

評価結果は以下の通りです。

Zoom Meetings版DeepL Voiceが、AI翻訳字幕の人間による品質評価で最高評価を獲得

すべてのテスト対象言語ペアおよびプラットフォームにおけるAI翻訳字幕の総合品質スコア（100点満点）。



出典：Slator。総合品質スコアは、各言語ペアごとにネイティブ言語専門家2名が実施した。AI翻訳による字幕の正確性および高解像度に関するブラインド評価に基づいて算出されました。品質スコアは、評価対象となったすべてのセグメントにおけるエラーの平均的な重大度を反映したもので、0~100のスケールに正規化されています。テスト対象となった言語の組み合わせは、英語からスペイン語、フランス語、ドイツ語、イタリア語、ポルトガル語、韓国語、日本語への翻訳、およびその逆方向の翻訳でした。

- Zoom Meetings版DeepL Voiceは、AI翻訳による字幕に関する人間による評価において、100点満点中96.4点という最高スコアを記録しました。
- これに続き、Teams版DeepL Voiceが100点満点中96.3点を記録しました。
- Google Meetは100点満点中87点で、最も低いスコアとなりました。

重大度の評価

Slatorは、上記の品質スコアをさらに詳細に分析し、「AI翻訳の品質は、画面上の字幕に対するエンドユーザーの理解にどの程度影響するのか」という問いに答えました。これにより、各プラットフォームにおける重大な翻訳ミスの実態を把握することができます。

さらにSlatorは深刻または重度の正確性エラーを含むセグメントの割合を追跡しました。言語専門家に対し、翻訳エラーの重大度について、正確性（誤訳、脱落、追加）および流暢さ（文体、文法、綴り）の観点から評価するよう依頼しました。これにより、SlatorはAI翻訳による字幕の全体的な失敗率、すなわち、深刻または重度の正確性エラーを含むセグメントの割合を追跡できるようになりました。これらのエラーは、元の意味の完全な喪失や誤訳を引き起こします。

この失敗率は、字幕がソース音声の意味をどの程度の頻度で著しく歪めたり、失ったりしているかを示すものであり、前述の品質スコアにさらなる示唆を与えます。同様に、合格率は、各プラットフォームがソース音声の意味をどの程度正確に伝えているか、またターゲット言語における流暢さを示しています。

以下に、各カテゴリーの例（Google Meetの翻訳）を示します。

例1（不合格）：

- ソース音声（英語）：“Let me start by saying that I think when you sit in this position as CEO of a company like Merck, there are many stakeholders who have interests.”
- ターゲット字幕（スペイン語）：“Entonces, permítanme comenzar diciendo que creo que cuando uno se sienta en esta posición, como director ejecutivo de una empresa como Killing, hay muchas partes interesadas que tienen intereses.”
- エラーカテゴリー：誤訳（重度）、文体（重度）
- 言語専門家のコメント：直訳では意味上の誤りが生じます。具体的には、「この姿勢で座ると」や「利害関係を持つステークホルダー」といった不自然な表現が見られました。また、「Merck」はスペイン語で「Murder（殺人）」と誤って解釈されています。

例2（条件付き合格）：

- ソース音声（英語）：“If you have to pay that much money, all of the things that I just said are not going to make you pleased with the system.”
- ターゲット字幕（スペイン語）：“Y si tienes que pagar tanto dinero, todas las cosas que acabo de decir no te harán sentir satisfecho con el sistema.”
- エラーカテゴリー：文体（重度）
- 言語専門家のコメント：AIによる翻訳字幕で、二人称の親称（tú）に切り替わっています。これはやや不自然であり、スペイン語では通常、中立的な三人称（例：「Y si hay que pagar tanto dinero...」）が用いられます。

例3 (合格) :

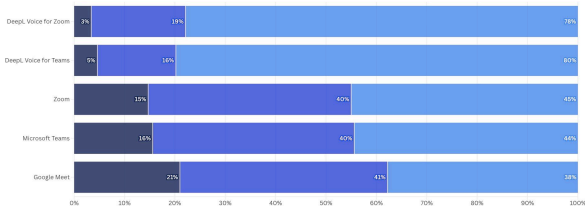
- ソース音声 (英語) : "I think it's my job to ensure that the company functions in a sustainable way to create long term value for all of its stakeholders, including its shareholders."
- ターゲット字幕 (スペイン語) : "Creo que mi trabajo es garantizar que la empresa funcione de manera sostenible para crear valor a largo plazo para todas sus partes interesadas, incluidos los accionistas."
- エラーカテゴリー: 該当なし
- 言語専門家のコメント: 「このセグメントのスペイン語訳は、実際に話されている英語の内容を反映しています。」

この分析の結果は以下の通りです。

DeepL Voiceは深刻なエラーを平均13%削減し、平均合格率は79%

テスト対象のすべての言語組み合わせおよびプラットフォームにおける、AI翻訳による字幕の総合的な合格率、条件付き合格率 (問題あり) 、および不合格率です。

■ 不合格率 ■ 条件付き合格率 (問題あり) ■ 合格率



出典: Slator • 各種割合は、各言語ペアごとにネイティブ言語専門家2名が実施した、AI翻訳による字幕の正確性および流暢さに関するブラインド評価に基づいて算出されました。これらの数値は、評価対象となったすべてのセグメントにおける翻訳エラーの重大度を表しています。テスト対象となった言語の組み合わせは、英語からスペイン語、フランス語、ドイツ語、イタリア語、ポルトガル語、韓国語、日本語への翻訳、およびその逆方向の翻訳でした。

- DeepL Voice製品におけるAIによる翻訳字幕の平均失敗率は4%であり、他のすべてのツールでは平均17%でした。これは、DeepL Voice製品を使用することで、**深刻または重度の正確性エラーが76%削減された**ことを示しています。
- DeepL Voice製品は、AI翻訳された全セグメントにおいて**平均79%の合格率**を達成しており、他のすべてのツールでは平均42%にとどまりました。これは、他の競合ツールと比較して、**完全合格セグメントの割合が88%増加した**ことを示しています。
- Google Meetではセグメントの約60%、ZoomおよびMicrosoft Teamsではその半数に、**重度のエラーまたは重大な翻訳上の問題が含まれていました。**

言語別評価

Slatorは、業界標準 (OOTB) のプラットフォーム (Microsoft Teams、Zoom) とカスタムプラットフォーム (Teams版DeepL Voice、Zoom Meetings版DeepL Voice) の言語ごとの差異をさらに詳しく分析し、次の問いに答えました。

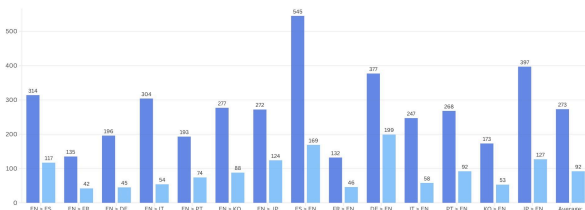
「DeepL Voiceは、業界標準 (OOTB) のプラットフォームと比較して、どの程度優れている、あるいは劣っているのか?」

その結果、すべての言語ペアにおいて、DeepL Voiceは、Microsoft Teamsと比べてセグメントあたりの翻訳エラー数を平均66%、Zoomと比べて64%削減することが明らかになりました。Microsoft TeamsとMicrosoft Teams版DeepL Voiceの比較結果は以下のとおりです。

Teams版DeepL Voiceの利用により、平均字幕エラー率が66%減少

Microsoft TeamsとTeams版DeepL Voiceにおける、言語ペア別のAI翻訳字幕のエラー率の比較です。

■ Microsoft Teams ■ Teams版DeepL Voice



出典: Slator • 数値が低いほど品質が高いことを示しています。翻訳の字数は、言語ペアごとに2名の言語専門家によって確認され、正確性と流暢性に基づいて評価されました。採点において、流暢性の誤りは0.5倍で重み付けされました。

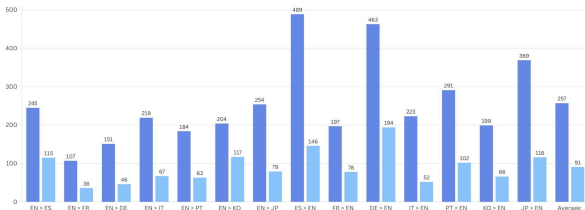
- Teams版DeepL Voiceを使用した場合、セグメントあたりの翻訳エラー数が最も大きく減少したのは、以下の言語ペアでした。
 - 英語からイタリア語 (82%)
 - 英語からドイツ語 (77%)
 - イタリア語から英語 (76%)
- Teams版DeepL Voiceを使用した場合、改善率が50%を下回った言語ペアは1つのみでした (ドイツ語→英語、47%)。

ZoomとZoom Meetings版DeepL Voiceの比較結果は以下のとおりです。

Zoom Meetings版DeepL Voiceの利用により、平均字幕エラー率が65%減少

ZoomとZoom Meetings版DeepL Voiceにおける、言語ペア別のAI翻訳字幕のエラー率の比較です。

■ Zoom ■ Zoom版DeepL Voice



出典: Slator • 数値が低いほど品質が高いことを示しています。翻訳の字数は、言語ペアごとに2名の言語専門家によって確認され、正確性と流暢性に基づいて評価されました。採点において、流暢性の誤りは0.5倍で重み付けされました。

- Zoom Meetings版DeepL Voiceを使用した場合、セグメントあたりの翻訳エラー数が最も大きく減少したのは、以下の言語ペアでした。
 - イタリア語から英語（77%）
 - スペイン語から英語（70%）
 - 英語からドイツ語（70%）
- Zoom Meetings版DeepL Voiceを使用した場合、改善率が50%を下回った言語ペアは1つのみでした（英語→韓国語、43%）。

言語専門家の選好

品質評価の完了後、Slatorは参加した28名の言語専門家全員に対し、ユーザー体験に基づいてAI翻訳字幕の各プラットフォームの優先順位を付けるよう依頼しました。専門家たちは、各プラットフォームに割り当てられたコードネームをもとに盲検評価を行い、最も好ましいプラットフォームを選定しました。

結果は以下の通りです。

- 言語専門家の52%（15名）が、Teams版DeepL Voiceを最も好ましいプラットフォームとして挙げました。
- 言語専門家の44%（12名）が、Zoom Meetings版DeepL Voiceを最も好ましいプラットフォームとして挙げました。
- 全体として、言語専門家の96%（27名）が、Teams版DeepL VoiceまたはZoom Meetings版DeepL Voiceのいずれかを上位2位以内に挙げました。
- Zoom Meetingsは通常3位でしたが、最も好ましいツールとして1票を獲得しました。Microsoft TeamsとGoogle Meetは上位2位には選ばれず、いずれも一貫して4位または5位に位置付けられました。

字幕の安定性

Slatorは、すべての言語およびプラットフォームにおける変更イベント数を自動的に計測しました。以下に、サンプリングした音声ファイルから抽出した変更の例を3つ示します。これらの例は、字幕が最終的に安定する前にどのように表示されていたかを示しています。

例1（文の段階的な生成：フランス語→英語）：

- フレーム1：“Hello Nathalie the topic of the”
- フレーム2：“Hello Nathalie the topic of the day is the”
- フレーム3：“Hello Nathalie the topic of the day is the subject of pharmacy”
- フレーム4（安定状態）：“Hello Nathalie the topic of the day is the subject of pharmacy groups.”

この例では、話者が文を言い終える前に、ライブ字幕に部分的な翻訳が表示されます。追加の単語が処理されるにつれて、字幕は全文が確定するまで順次拡張されていきます。

例2（字幕の書き換え、スペイン語→英語）：

- フレーム1：“When you make a demand what is the minimum economic amount”
- フレーム2：“When you make a claim what is the minimum economic amount”
- フレーム3：“When you make a claim what is the minimum economic amount so that”
- フレーム4（安定状態）：“When you make a claim what is the minimum economic amount so that you can file a case?”

この例では、追加の言語的文脈が得られるにつれて、ライブ字幕は先に表示された語句を修正していきます。

例3（字幕のちらつき、韓国語→英語）：

- フレーム1：“the fourth quarter ended about 10 days”
- フレーム2：“the fourth quarter is now about 10 days”
- フレーム3（安定状態）：“the fourth quarter ended about 10 days ago”

この例では、ライブ字幕は、先に表示された語句をいったん完全に書き換えた後、さらに別の最終的な表示内容へと再度変更されます（A→B→A、A→B→Cなどの挙動）。

例4（高い安定性、韓国語→英語）：

- フレーム1：“Let's talk about Samsung Electronics' performance.”
- フレーム2：“Let's talk about Samsung Electronics' performance.”
- フレーム3：“Let's talk about Samsung Electronics' performance.”

この例では、翻訳の字幕が表示され、その後も変更されることはありませんでした。

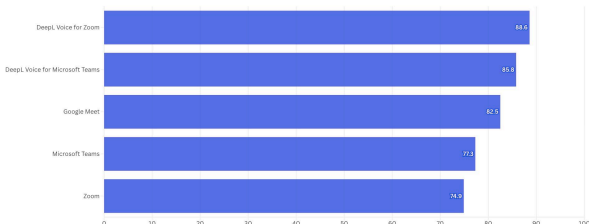
総合評価

Slatorは、すべてのプラットフォームにおける翻訳字幕の全体的な安定性を反映した、0~100の単一の品質スコアに結果を統合しました。これにより、Slatorは次の問いに答えることが可能となりました：「各プラットフォームでは、字幕が最終的に安定するまでに、どの程度の頻度で翻訳がちらついたり更新されたりするのか？」

この点についての評価結果は以下のとおりです。

翻訳字幕の安定性において、Zoom Meetings版DeepL Voiceが最も優れたプラットフォーム

テスト対象となったすべての言語ペアおよびプラットフォームにおける、翻訳字幕の安定性スコアの平均値（100点満点）。



出典：Slator + 各ツールにおける翻訳の字幕のフレームごとの記録を分析し、全言語にわたる重要イベントの総数を算出することで、平均安定性スコアを導出しました。これらのスコアは、0~100の単一の安定性スコアに正規化されました。

- Zoom Meetings版DeepL Voiceは、AI翻訳による字幕の安定性に関する自動評価において、100点満点中88.6点という**最高スコア**を記録しました。
- これに続き、Teams版DeepL Voiceが100点満点中85.8点を記録しました。
- Zoomは100点満点中74.9点で、最も低いスコアとなりました。

言語別評価

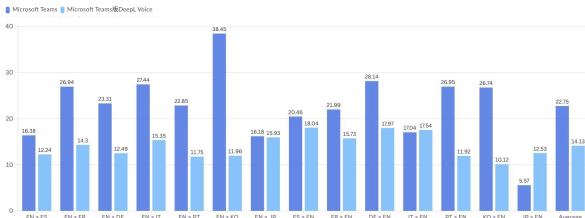
Slatorは、業界標準（OOTB）のプラットフォーム（Microsoft Teams、Zoom）とカスタムプラットフォーム（Teams版DeepL Voice、Zoom Meetings版DeepL Voice）の言語ごとの差異をさらに詳しく分析し、次の問いに答えることが可能となりました。「DeepL Voice製品は、業界標準（OOTB）のプラットフォームと比較して、言語ごとに見た場合、安定性はどの程度優れている、または劣っているのか？」

すべての言語ペアにおいて、DeepL VoiceはMicrosoft Teamsと比較して平均37.6%、Zoomと比較して54.7%、字幕の書き換え頻度を低減しました。

Microsoft TeamsとMicrosoft Teams版DeepL Voiceの比較結果は次のとおりです。

Teams版DeepL Voiceの利用により、安定性が平均38%改善

Microsoft TeamsとTeams版DeepL Voiceにおける、各言語ペア間のAI翻訳字幕の書き換え率の比較です。



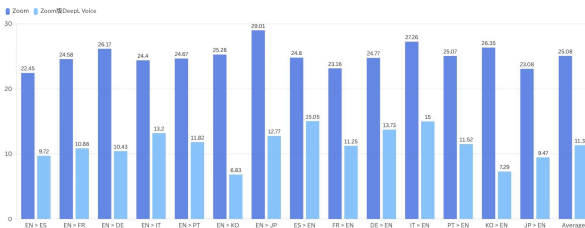
出典: Slator * 字幕の書き換え率とは、表示された字幕が変更されたフレームの割合を指します。これにより、ユーザーが字幕の更新や表示の乱れをどの程度の頻度で経験するかが、直感的に数値化されます。数値はPythonを使用して自動的に算出されました。

- Teams版DeepL Voiceにおける翻訳字幕の安定性において、最も大きな改善が見られた言語ペアは以下の通りです。
 - 英語から韓国語 (+69%)
 - 韓国語から英語 (+62%)
 - ポルトガル語から英語 (+56%)
- Microsoft TeamsのAI翻訳字幕の方がより安定していた言語ペアは2つのみでした。イタリア語から英語 (Microsoft Teamsの方が3%安定) と、日本語から英語 (Microsoft Teamsの方が125%安定) です。

ZoomとZoom Meetings版DeepL Voiceの比較結果は次のとおりです。

Zoom Meetings版DeepL Voiceの利用により、安定性が平均55%改善

ZoomとZoom Meetings版DeepL Voiceにおける、各言語ペア間のAI翻訳字幕の書き換え率の比較です。



出典: Slator * 字幕の書き換え率とは、表示された字幕が変更されたフレームの割合を指します。これにより、ユーザーが字幕の更新や表示の乱れをどの程度の頻度で経験するかが、直感的に数値化されます。数値はPythonを使用して自動的に算出されました。

- **Zoom Meetings版DeepL Voice**における翻訳字幕の安定性において、**最も大きな改善**が見られた言語ペアは以下のとおりです。
 - 英語から韓国語（73%）
 - 韓国語から英語（72%）
 - 英語からドイツ語（60%）
- Zoomと比較して、Zoom Meetings版DeepL Voiceの方が安定性が低かった言語ペアはありませんでした。

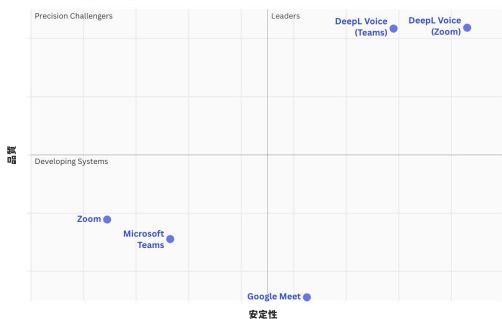
結論

総合結果

以下は、5つのAI翻訳字幕プラットフォームの総合ランキングです。

AI翻訳字幕プラットフォームのクアドラント

言語品質と字幕表示の安定性に基づく、リアルタイム字幕翻訳システムの比較評価。



「Teams版DeepL Voice」と「Zoom Meetings版DeepL Voice」は、品質と安定性の両面で高い評価を獲得し、**AI翻訳字幕ソリューションにおけるリーダー**としての地位を確立しました。リアルタイム字幕システムは、高精度な翻訳を提供しつつ、その表示にかかる遅延を最小限に抑える必要があるため、翻訳精度と字幕の安定性との間でトレードオフに直面する可能性があります。しかし、本研究の結果は、DeepL Voiceがこれらの両方を同時に改善していることを示唆しています。

一方、Microsoft TeamsとZoomはいずれも「開発段階のシステム」と位置付けられており、字幕表示と翻訳品質の両面で平均を下回るパフォーマンスにとどまりました。これらのツールにおいて、安定性の面ではMicrosoft Teamsが、品質の面ではZoomが一方よりわずかに優れていましたが、いずれのツールも既存のサービスを改善するには、各項目においてさらなる改善が必要です。

Google Meetは、市場の他のツールと比較して比較的安定しているものの、Slatorの言語専門家チームによる評価では、品質スコアが最も低い結果となりました。

これにより、Google Meetは総合ランキングにおいて「効率重視型」として位置付けられますが、字幕表示の安定性を維持・向上させつつ、言語ペア全体の品質を改善できれば、「リーダー」となる可能性があります。

企業への示唆

Slatorは本評価の完了を踏まえ、企業に対して以下の示唆を提示しています。

- 1. AI字幕の品質は向上しているものの、プラットフォームによってばらつきが見られる**
 結果は、一般的に利用されている会議プラットフォーム間で、翻訳品質に大きな差があることを示しています。組み込みの字幕翻訳機能に依存している組織では、使用するツールによって翻訳の精度に大きな差が生じる可能性があります。
- 2. 字幕の安定性は使いやすさに直接影響する**
 たとえ翻訳が正確であっても、字幕が頻繁に書き換えられると、理解を妨げる可能性があります。字幕の安定性を測定することは、多言語会議における実際のユーザー体験を把握する上で有用な指標となります。
- 3. カスタマイズ機能を備えた専用プラットフォームは、翻訳品質の向上に寄与する**
 本結果は、専用のAI翻訳ツールの活用と、ドメイン別の用語集や文字起こしの強化を適用することで、多言語にわたるリアルタイムAI字幕の翻訳性能を大幅に向上できることを示唆しています。

付録

評価方法

オーディオサンプルの収集

Slatorは、対象となる各言語のポッドキャスト音声を集めました。ポッドキャストの録音素材が選ばれた理由として、言語を問わず一貫した音質が確保されており、背景ノイズやマイク品質の差異ではなく、翻訳性能を統制された条件下で比較できるためです。

各ポッドキャストでは、2名の話者がビジネス関連の規制対象分野（金融、ライフサイエンス、法務）について対話形式でやり取りしており、背景ノイズは最小限またはほぼ存在しません。このアプローチにより、専門用語の翻訳品質を測定し、各プラットフォームが原文の言語のネイティブスピーカー間の自然な会話をどのように処理するかを評価する実験が構築されました。

各ポッドキャストは編集され、言語ごとに約12分間の独立したソース音声サンプルが作成されました。この音声ファイルは、自動文字起こしツールで処理され、音声の文字起こしデータが生成されました。

技術的な設定

Slatorは、ソース音声を各プラットフォーム（Google Meet、Microsoft Teams、Zoom）にルーティングするための仮想ケーブルを設定しました。これにより、録音中に余分な雑音や意図しない背景音が入ることはありませんでした。Slatorはスクリーンキャプチャツールを使用して、各プラットフォームにおける翻訳字幕を.mp4形式の動画として記録しました。この動画には、ユーザーが実際に目にするのと同じ状態でライブ字幕が表示されています。

Google Meet、Microsoft Teams、Zoomでは、業界標準（OOTB）の翻訳字幕設定が使用されました。Google Meetに限っては、英語からイタリア語への翻訳機能がベータ版であることが確認されました。

Microsoft TeamsおよびZoomでDeepL Voiceをテストするにあたり、Slatorは用語集を適用し、ポッドキャストの会話が親しみやすい性質であることを踏まえ、DeepLの語調設定でカジュアルなスタイルを選択しました。

さらに、SlatorはDeepL Voiceの「Voiceで使う用語」機能（Microsoft Teamsでのみ利用可能）を有効化しました。この機能は、Voiceで使う用語（略語や固有名詞など）の単一言語リストであり、音声のテキスト化プロセスにおいて、ツールが単一言語の用語を認識する能力を強化するものです。

Slatorは、すべての録音データを自動処理エンジン（「自動品質評価」で説明）で処理し、その録音データを言語専門家と共有しました。

言語専門家の選定

Slatorは、各言語ペアにつき2名の言語専門家を選定し、各専門家が訳文の言語のネイティブスピーカー、またはそれに近いレベルであることを確認しました。選定された言語専門家は、音声ファイルに関連する分野の専門知識に基づいて厳選されており、有資格の専門家です。

選定された言語専門家は、主にメディアローカライゼーション、翻訳、通訳、および品質評価（AIトレーニング/AI評価を含む）の経験を有しています。こうした専門性の組み合わせにより、音声ファイルを聞きながら、同時に画面上の字幕の品質を評価する必要性がさらに高まりました。

人間による品質評価

Slatorは、各ソース音声ファイルごとに品質評価スコアカードを作成しました。以下に例を挙げます。

| 1 | A | B | C | D | E | F | G | H | I | J | K |
|---------------------|------------|----------------|--|-----------------------|----------|----------|-------|---------|----------|--------------------|---|
| Baseline Parameters | | | | Final Caption Quality | | | | | | | |
| Tool Number | Segment ID | Start Time | Original Transcript | Mistranslation | Omission | Addition | Style | Grammar | Spelling | Evaluator Comments | |
| 1 | 1 | [00:00:00.000] | Let me start by saying that I think when you sit in this position as CEO of a company like Merck, there are many stakeholders who have interests. | | | | | | | | |
| 1 | 2 | [00:00:08.120] | I see my job, generally speaking, as to try to meet the needs of multiple stakeholders whose interests are often, if not opposed to one another, it's some dynamic tension with one another. | | | | | | | | |
| 1 | 3 | [00:00:21.720] | That actually gets to the short term versus the long term issue. | | | | | | | | |

スコアカードには以下の内容が含まれます。

- ベースラインパラメータ：
 - ツール番号。これは、各ツール/プラットフォームに対する匿名化された参照であり、言語専門家が評価対象のプラットフォーム名や、各プラットフォームがカスタマイズされているかどうかを把握できないようにするためのものです。言語専門家には、本評価にDeepLのカスタマイズモデルが含まれていることは伝えられていませんでした。
 - セグメントID。これは、ソース音声の文字起こしにおけるセグメントIDです。これにより、Slatorは評価をテキストのセグメント単位に分割し、同一言語ペア内の複数の翻訳者間に加え、言語やプラットフォームをまたいだ分析をより明確に行えるようになりました。
 - 開始時刻。これは原文の文字起こしに基づいており、言語専門家がソース音声ファイル内の特定のセグメントへ必要に応じて簡単に移動できるようになっています。
 - 原文の文字起こし。これは原文の文字起こしです。言語専門家には、この文字起こしに誤りが含まれている可能性があることが伝えられました。Slatorは、言語専門家が翻訳字幕を文字起こしと照らし合わせて評価するのを防ぎ、

代わりに原音声を聞くようにするため、あえて文字起こし内の誤りを修正しませんでした。同様に、同じ理由から、翻訳字幕の文字起こしも提供しませんでした。

- 最終的な字幕の品質：
 - 正確さのカテゴリー：
 - 誤訳。Slatorはこのカテゴリーを「翻訳の字幕の意味が不正確、または誤解を招く状態」と定義しています。分類は以下の通りです。
 - なし：特に問題なし
 - 軽度：意味は概ね正しいが、ニュアンスの欠落がある
 - 重度：意味が一部歪められている
 - 深刻：意味が著しく誤っている
 - 訳抜け。Slatorはこのカテゴリーを「翻訳の字幕から重要な情報が欠落している状態」と定義しています。分類は以下の通りです。
 - なし：特に問題なし
 - 軽度：意味は概ね正しいが、ニュアンスの欠落がある
 - 重度：意味が一部歪められている
 - 深刻：意味が著しく誤っている言語学者には、プラットフォームがそのセグメントを翻訳しなかった場合、または翻訳字幕に原文の言語が表示された場合、そのセグメントを「深刻」として分類するよう指示されました。
 - 余分。Slatorはこのカテゴリーを「音声に含まれていない情報が翻訳の字幕に追加されている状態」と定義しています。分類は以下の通りです。
 - なし：特に問題なし
 - 軽度：意味は概ね正しいが、ニュアンスの欠落がある
 - 重度：意味が一部歪められている
 - 深刻：意味が著しく誤っている
 - 流暢さのカテゴリー：
 - 文体。Slatorはこのカテゴリーを「字幕の文体や語調がライブ字幕として不適切である状態」と定義しています。例：字幕が冗長すぎる、語調が堅すぎる／カジュアルすぎる、書き言葉のように感じられる、など。分類は以下の通りです。
 - なし：特に問題なし
 - 軽度：目立つものの、読みやすさには支障はない
 - 重度：読みづらく、可読性が低下する
 - 文法。Slatorはこのカテゴリーを「文がターゲット言語の文法規則に

違反している状態」と定義しています。（例：一致の誤り、時制や活用の誤り）分類は以下の通りです。

- なし：特に問題なし
- 軽度：目立つものの、読みやすさには支障はない
- 重度：読みづらく、可読性が低下する

- 綴り。Slatorはこのカテゴリーを「読解に影響を及ぼす誤字や表記の誤り」と定義しています。（例：スペルミス、ダイアクリティカルマークの誤り、大文字・小文字の誤り）分類は以下の通りです。

- なし：特に問題なし
- 軽度：目立つものの、読みやすさには支障はない
- 重度：読みづらく、可読性が低下する

- 評価者のコメント。Slatorは、言語専門家に対し、評価の根拠を英語で説明するよう求めました。これにより、Slatorは言語やプラットフォームを横断した照合を行い、指示が適切に遵守されているか、および評価に一貫性があるかを確認することが可能となりました。また、言語専門家は評価を意識的に行うよう促され、評価における一貫性の向上にもつながりました。さらに、さまざまな言語やプラットフォームにわたる翻訳字幕について、定性的なフィードバックも得られました。

言語専門家には、最終的に確定し安定した翻訳字幕のみに基づいて、品質評価スコアカードに記入するよう指示されました。

これに加え、Slatorは言語専門家に対し、ユーザー体験・好みおよび品質の観点から、各プラットフォームを順位付けするよう依頼しました。Slatorはこれらの評価をもとに、どのプラットフォームが言語専門家に好まれているか、またどのプラットフォームのユーザー体験が最も低く評価されているかについて結論を導き出しました。

人間による評価のスコアリング

データのクリーニングおよび重み付け

各言語ペアは2名の独立した言語専門家によって評価されたため、Slatorは結果を集計する前に、評価の一貫性および外れ値を確認しました。評価者間でスコアを平均化し、プラットフォームおよび言語ペアごとのセグメント平均スコアを算出しました。人間による言語評価では、重大度の評価に多少のばらつきが生じることが想定されますが、クロスレビューの結果、評価者間でエラーの分類に高い一貫性があることが確認されました。

Slatorは、品質評価を数値化しました（なし [0点]、軽度 [1点]、重度 [3点]、深刻 [5点]）。

正確性に関連するカテゴリー（すなわち、誤訳、訳抜け、および余分）については、正確性の総合スコアが算出されました。

流暢さに関連するカテゴリー（すなわち、文体、綴り、および文法）についても、流暢さの総合スコアが算出されました。

これら「正確さ」と「流暢さ」の2つのスコアを合算し、「セグメントの総合スコア」を算出しました。同一言語ペアにおける複数の言語専門家のスコアを合算・平均化し、「正確さ」、「流暢さ」、および「セグメントの総合スコア」の平均値を算出しました。

「セグメントの総合スコア」において、「流暢さ」スコアには0.5倍の重み付けが適用されました。これにより、スペル、文法、または文体に関する1つの重大なエラーのみでセグメントが「不合格」となることを防ぎ、言語専門家の間でより品質への影響が大きいと見なされる誤訳や訳抜けといった正確さの問題に重点を置いて、セグメント全体のスコアが算出されました。

データの分類

また、「セグメントの総合スコア」において、各セグメントが「合格」、「条件付き合格」、または「不合格」のいずれに該当するかも示されました。これにより、Slatorはプラットフォームおよび言語をまたいだ不合格率を算出できるようになりました。

前述の通り、正確性を重視した重み付けが行われているため、以下の条件のいずれかに該当する場合、そのセグメントは自動的に「不合格」と分類されます。

- 正確さに関する項目（誤訳、訳抜け、または余分）で、セグメントが「深刻」と評価された場合
- 正確さに関する項目で、セグメントに「重度」のエラーが2つ以上含まれる場合（例：重度の訳抜けと重度の余分）
- 誤訳が「重度」と評価された場合（または、項目1に従い「深刻」と評価された場合）

以下の条件を満たす場合、そのセグメントは「条件付き合格」と分類されます。

- 流暢さに関する項目（スペル、文法、文体）で、セグメントが「重度」と評価された場合、または
- 正確さに関する項目で、セグメントが「重度」と評価された場合（ただし、「不合格」には該当しない場合）、または
- すべてのカテゴリーにおいて、軽度のエラーが3件を超える場合

軽度のエラーが1件または2件のみの場合、そのセグメントは合格条件を満たします。

データのスコアリング

各プラットフォームと言語ペアごとに「セグメントの総合スコア」が算出され、これにより直接の比較が可能となりました。

加重された「セグメントの総合スコア」は以下の通りです（数値が低いほどスコアが高くなります）。

| 言語ペア | Google Meet | Microsoft Teams | Zoom | Teams版DeepL Voice | Zoom版DeepL Voice |
|-----------|-------------|-----------------|------|-------------------|------------------|
| 英語→スペイン語 | 288 | 314 | 245 | 117 | 115 |
| 英語→フランス語 | 164 | 135 | 107 | 42 | 36 |
| 英語→ドイツ語 | 179 | 196 | 151 | 45 | 46 |
| 英語→イタリア語 | 252 | 304 | 219 | 54 | 67 |
| 英語→ポルトガル語 | 210 | 193 | 184 | 74 | 63 |
| 英語→韓国語 | 237 | 277 | 204 | 88 | 117 |
| 英語→日本語 | 291 | 272 | 254 | 124 | 79 |
| スペイン語→英語 | 531 | 545 | 489 | 169 | 146 |
| フランス語→英語 | 274 | 132 | 197 | 46 | 78 |
| ドイツ語→英語 | 408 | 377 | 463 | 199 | 194 |
| イタリア語→英語 | 311 | 247 | 223 | 58 | 52 |
| ポルトガル語→英語 | 390 | 268 | 291 | 92 | 102 |
| 韓国語→英語 | 329 | 173 | 199 | 53 | 66 |
| 日本語→英語 | 671 | 397 | 369 | 127 | 116 |

不合格率および合格率は、各カテゴリ（不合格／条件付き合格／合格）におけるセグメント数を集計して、次のように算出されました。

| ツール | 不合格率 | 条件付き合格率（問題あり） | 合格率 |
|-------------------|------|---------------|------|
| Google Meet | 307 | 603 | 555 |
| Microsoft Teams | 229 | 586 | 650 |
| Zoom | 216 | 590 | 659 |
| Teams版DeepL Voice | 67 | 228 | 1170 |
| Zoom版DeepL Voice | 50 | 274 | 1141 |

これらの数値を、全言語にわたる総セグメント数（1465）で割ることで、全体に対する割合を算出しました。

| ツール | 不合格率 | 条件付き合格率（問題あり） | 評価合格率 |
|-------------------|-------|---------------|-------|
| Google Meet | 21.0% | 41.2% | 37.9% |
| Microsoft Teams | 15.6% | 40.0% | 44.4% |
| Zoom | 14.7% | 40.3% | 45.0% |
| Teams版DeepL Voice | 4.6% | 15.6% | 79.9% |
| Zoom版DeepL Voice | 3.4% | 18.7% | 77.9% |

データ処理

ツール間および言語ペア間の比較を明確にするため、「セグメントの総合スコア」を0～100の品質スコアに正規化しました。各ツールについて、まずセグメントごとの平均ペナルティを算出しました。これは、評価対象となったすべてのセグメント（すべての言語ペアおよび双方向）における「セグメントの総合スコア」を、評価対象となったセグメントの総数で割った値として定義されます。

エラーポイントは、あらかじめ定義された採点枠組み（正確さおよび流暢さのカテゴリーに、重大度に基づく重み付けを適用したもの）に従って割り当てられました。つまり、セグメントごとの理論上の最大ペナルティは24点となります（正確さカテゴリーから15点、流暢さカテゴリーから9点）。その後、正規化スコアは次のように算出されました：
 $100 \times (1 - \text{セグメントごとの平均ペナルティ} \div 24)$

この計算式では、スコアが100の場合、エラーが一切観測されないシステムを示します。スコアがゼロの場合、すべてのセグメントで理論上可能な最大ペナルティが課された極端なケースを示します。

データは以下の通りです。

| ツール | 加重総合スコア | 総セグメント数 | 平均ペナルティ値 | 正規化スコア (0~100) |
|-------------------------------|---------|---------|----------|-------------------|
| Google Meet | 4534 | 1456 | 3.09 | 87.10537543 |
| Microsoft Teams | 3828 | 1456 | 2.61 | 89.11333902 |
| Zoom | 3592 | 1456 | 2.45 | 89.78313424 |
| Teams版 DeepL Voice | 1286 | 1456 | 0.88 | 96.34172355 |
| Zoom Meetings版 DeepL Voice | 1277 | 1456 | 0.87 | 96.36732082 |

観測されたツールのスコア（87～96）は、この理論上の最大エラー上限に対する相対的な性能を反映しており、ツール間、言語間、および将来の評価における比較可能性を保証しています。

自動化された品質評価

Google Meet、Microsoft Teams、Zoom、DeepL Voice for Microsoft Teams、およびZoom Meetings版DeepL Voiceにおいて、ライブ字幕が画面上で時間の経過とともにどの程度安定または不安定に表示されるかを定量化する自動測定パイプラインを構築しました。また、言語固有のOCRを活用して、複数言語にわたる分析も実施しています。

簡単に言うと、このシステムは、録画されたすべての会議動画（言語およびプラットフォームごとに1本）を取り込み、各フレームから字幕領域を抽出し、その領域にOCRを適用して、

その時点で閲覧者に表示されていたテキストを復元します。その後、字幕の安定性指標を算出し、要約レポートを作成します。

その目的は、基盤となるASRの文字起こしではなく、ユーザーが実際に目にする字幕体験を測定することです。字幕がちらついたり、書き換えられたり、表示が不安定に変動したりする場合でも、Slatorはレンダリングされた動画フレームからその挙動を直接取得します。

具体的には、Slatorは録画ファイルごとに1秒あたり約10フレームを抽出し、それらを画像ファイルとして保存しました。これにより、言語やプラットフォームにわたり、画面に表示された内容をフレーム単位で分析することが可能となり、一瞬のうちに発生する急速な書き換えも分析に反映することが可能です。

抽出された各フレームについて、Slatorのシステムは、字幕オーバーレイ領域のみを含むように画面をトリミングし、無関係な画面コンテンツを除外しました。

Slatorは、各字幕言語を認識できるOCRエンジンであるTesseractを活用し、トリミングされた各字幕画像に対してOCR処理を行い、表示されている字幕テキストを抽出し、フレームごとに1件のレコードとして保存しました。

生のOCR出力は、スペースの違い、Unicodeの差異、および不可視文字の影響により異なる場合があります。フレームを比較する前に、システムは厳密にテキストを正規化し、比較結果がOCRノイズではなく、実際に目に見える変化を反映します。

正規化には以下の処理が含まれます。

- Unicodeの正規化
- タイムスタンプの削除
- 話者名の削除
- UIテキストの削除（例：「翻訳字幕が有効です」）
- 連続する空白の統合
- 先頭・末尾の空白の削除
- ゼロ幅文字の削除
- 言語グループ別の処理（ラテン文字とCJK）

これにより、意味のある差異を保持しつつ、可能な限り「変更」が実際にユーザーに見える書き換えが反映されます。

システムはフレームごとのOCR結果を読み取り、時間の経過に沿ってテキストを比較することで、字幕が更新されるタイミング、書き換えの頻度、およびちらつきや不安定な変動の有無を検出します。

このパイプラインは、動画ごとの主要な指標（アクティブなフレーム数や変更イベント数など）を要約したCSVファイルを出力します（詳細は後述の「自動評価のスコアリング」を参照）。これらの結果は、プラットフォーム間、言語間、およびテスト条件（DeepLの使用の有無）で直接比較できるよう設計されています。

自動評価のスコアリング

字幕の安定性を客観的に測定するため、Slatorは、フレーム間で字幕がどの程度の頻度で変更されるかを測定しました。単語の追加、セグメントの再分割、あるいは一時的な変動やちらつきなど、目に見えるすべての変更は、変更イベントとして計測されました。具体的には、以下の変更イベントが検出されます。

- 文字の追加または削除
- 単語の補完
- 軽微な置換
- 句読点の調整
- 書式の変更
- 表示の往復的な変動
- テキストの不安定な変動
- 視認性を下げるちらつき

目に見える変化がなかったフレームの割合を、0~100の「安定性スコア」に換算しました。このスコアにおいて、100は完全に安定した字幕（フレーム間の変化がない状態）を示し、スコアが低いほど字幕の更新頻度が高いことを示します。

併せて、表示された字幕が変更されたフレームの割合として定義される字幕の書き換え率も算出しました。「安定性スコア」は視覚的に安定したフレームの割合を示す一方で、字幕の書き換えは、ユーザーが字幕の更新や表示の乱れをどの程度の頻度で経験するかを直接数値化します。

「安定性スコア」と「字幕の書き換え率」を組み合わせることで、プラットフォーム間で字幕のパフォーマンスを比較するための、透明性が高く、挙動に基づく枠組みを提供します。これらは、正確性、翻訳品質、または文法や意味の正確性を測定するものではありません（これらは人間による言語評価を通じて別途評価されました）。

安定性に関する生データ

生データは以下の通りです。

安定性ランキング（ツール比較）

このデータにより、Slatorは、同じプラットフォームでDeepL Voiceを使用した場合と使用しない場合を比較し、その増減率を算出することができました。

| ツール | 安定性スコアの平均値 (0~100) | 字幕の書き換え (ちらつき) |
|-----------------------------|--------------------|----------------|
| Zoom版DeepL Voice | 88.6 | 11.35% |
| Microsoft Teams版DeepL Voice | 85.8 | 14.20% |
| Google Meet | 82.5 | 17.50% |
| Microsoft Teams | 77.3 | 22.75% |
| Zoom | 74.9 | 25.08% |

改善率は、DeepL Voice製品と対応するネイティブプラットフォームにおける字幕の書き換え率の減少に基づいて算出されました。

安定性ランキング (言語比較)

言語別の安定性スコアの平均値は以下の通りです。

| 言語ペア | Zoom版DeepL Voice | Microsoft Teams版DeepL Voice | Google Meet | Microsoft Teams | Zoom |
|-----------|------------------|-----------------------------|-------------|-----------------|------|
| 英語→スペイン語 | 90.3 | 87.8 | 87.5 | 83.6 | 77.6 |
| 英語→フランス語 | 89.1 | 85.7 | 70.5 | 73.1 | 75.4 |
| 英語→ドイツ語 | 86.8 | 87.5 | 77.5 | 76.7 | 73.8 |
| 英語→イタリア語 | 88.2 | 84.7 | 85.4 | 72.6 | 75.6 |
| 英語→ポルトガル語 | 93.2 | 88.2 | 91.4 | 77.2 | 75.3 |
| 英語→韓国語 | 93.2 | 88.0 | 80.2 | 61.5 | 74.7 |
| 英語→日本語 | 87.2 | 84.1 | 67.2 | 83.8 | 71.0 |
| スペイン語→英語 | 84.9 | 82.0 | 88.7 | 79.5 | 75.2 |
| フランス語→英語 | 88.8 | 84.3 | 89.4 | 78.0 | 76.8 |
| ドイツ語→英語 | 86.3 | 82.0 | 75.5 | 71.9 | 75.2 |
| イタリア語→英語 | 85.0 | 82.5 | 88.4 | 83.0 | 72.7 |
| ポルトガル語→英語 | 88.5 | 88.1 | 78.6 | 73.1 | 74.9 |
| 韓国語→英語 | 92.7 | 89.9 | 87.9 | 73.3 | 73.7 |
| 日本語→英語 | 90.5 | 87.5 | 86.7 | 94.4 | 76.9 |
| 平均 | 88.6 | 85.9 | 82.5 | 77.3 | 74.9 |

字幕の書き換え率 (つまり、表示される字幕が変更されたフレームの割合) について、DeepL Voice製品と既製品との比較結果は以下の通りです。

| 言語ペア | Zoom版DeepL Voice | Microsoft Teams版DeepL Voice | Google Meet | Microsoft Teams | Zoom | Teams版DeepL VoiceとTeamsの差分 | Zoom Meetings版DeepL VoiceとZoomの差分 |
|----------|------------------|-----------------------------|-------------|-----------------|--------|----------------------------|-----------------------------------|
| 英語→スペイン語 | 9.72% | 12.24% | 12.53% | 16.38% | 22.45% | 25% | 57% |
| 英語→フランス語 | 10.86 | 14.30% | 29.45% | 26.94% | 24.58% | 47% | 56% |
| 英語→ドイツ語 | 10.43% | 12.49% | 22.54% | 23.31% | 26.17% | 46% | 60% |
| 英語→イタリア語 | 13.20% | 15.35% | 14.63% | 27.44% | 24.40% | 44% | 46% |

| | | | | | | | |
|-----------|--------|--------|--------|--------|--------|-------|-----|
| 英語→ポルトガル語 | 11.82% | 11.75% | 8.57% | 22.85% | 24.67% | 49% | 52% |
| 英語→韓国語 | 6.83% | 11.96% | 19.76% | 38.45% | 25.28% | 69% | 73% |
| 英語→日本語 | 12.77% | 15.93% | 32.78% | 16.18% | 29.01% | 2% | 56% |
| スペイン語→英語 | 15.05% | 18.04% | 11.32% | 20.46% | 24.80% | 12% | 39% |
| フランス語→英語 | 11.25% | 15.73% | 10.57% | 21.99% | 23.16% | 28% | 51% |
| ドイツ語→英語 | 13.73% | 17.97% | 24.53% | 28.14% | 24.77% | 36% | 45% |
| イタリア語→英語 | 15.00% | 17.54% | 11.58% | 17.04% | 27.26% | -3% | 45% |
| ポルトガル語→英語 | 11.52% | 11.92% | 21.39% | 26.95% | 25.07% | 56% | 54% |
| 韓国語→英語 | 7.29% | 10.12% | 12.11% | 26.74% | 26.35% | 62% | 72% |
| 日本語→英語 | 9.47% | 12.53% | 13.30% | 5.57% | 23.08% | -125% | 59% |
| 平均 | 11.35% | 14.13% | 17.50% | 22.75% | 25.08% | 38% | 55% |

最後の2つの列では、正の値は、業界標準（OOTB）のプラットフォームと比較してDeepL Voice製品を使用した場合の改善率を示しています。負の値は、DeepL Voice製品と比較して、業界標準（OOTB）のプラットフォームの方が安定性が高いことを示しています。

完全データセット

以下は、すべてのツールと言語に関する完全なデータセットです。

| ツール | 言語ペア | アクティブフレーム数 | フレームレート (fps) | 総変更イベント数 |
|---------------------------|-----------|------------|---------------|----------|
| Teams版DeepL Voice | ドイツ語→英語 | 8987 | 11 | 1615 |
| Teams版DeepL Voice | 英語→ドイツ語 | 7212 | 11 | 901 |
| Teams版DeepL Voice | 英語→スペイン語 | 7167 | 11 | 877 |
| Teams版DeepL Voice | 英語→フランス語 | 7099 | 10 | 1015 |
| Teams版DeepL Voice | 英語→イタリア語 | 7005 | 10 | 1075 |
| Teams版DeepL Voice | 英語→日本語 | 6957 | 10 | 1108 |
| Teams版DeepL Voice | 英語→韓国語 | 7021 | 10 | 840 |
| Teams版DeepL Voice | 英語→ポルトガル語 | 7249 | 11 | 852 |
| Teams版DeepL Voice | スペイン語→英語 | 7057 | 10 | 1273 |
| Teams版DeepL Voice | フランス語→英語 | 8123 | 11 | 1278 |
| Teams版DeepL Voice | イタリア語→英語 | 7214 | 10 | 1265 |
| Teams版DeepL Voice | 日本語→英語 | 5921 | 10 | 742 |
| Teams版DeepL Voice | 韓国語→英語 | 5285 | 10 | 535 |
| Teams版DeepL Voice | ポルトガル語→英語 | 6360 | 10 | 758 |
| Zoom Meetings版DeepL Voice | ドイツ語→英語 | 8766 | 11 | 1204 |
| Zoom Meetings版DeepL Voice | 英語→ドイツ語 | 7112 | 10 | 742 |
| Zoom Meetings版DeepL Voice | 英語→スペイン語 | 7246 | 11 | 704 |
| Zoom Meetings版DeepL Voice | 英語→フランス語 | 7192 | 11 | 781 |
| Zoom Meetings版DeepL Voice | 英語→イタリア語 | 7157 | 11 | 945 |
| Zoom Meetings版DeepL Voice | 英語→日本語 | 7241 | 11 | 925 |
| Zoom Meetings版DeepL Voice | 英語→韓国語 | 7272 | 11 | 497 |
| Zoom Meetings版DeepL Voice | 英語→ポルトガル語 | 7239 | 11 | 856 |

| | | | | |
|---------------------------|-----------|------|----|------|
| Zoom Meetings版DeepL Voice | スペイン語→英語 | 7268 | 11 | 1094 |
| Zoom Meetings版DeepL Voice | フランス語→英語 | 7923 | 10 | 891 |
| Zoom Meetings版DeepL Voice | イタリア語→英語 | 7299 | 10 | 1095 |
| Zoom Meetings版DeepL Voice | 日本語→英語 | 6117 | 11 | 579 |
| Zoom Meetings版DeepL Voice | 韓国語→英語 | 5380 | 11 | 392 |
| Zoom Meetings版DeepL Voice | ポルトガル語→英語 | 6258 | 10 | 721 |
| Google Meet | ドイツ語→英語 | 8512 | 10 | 2088 |
| Google Meet | 英語→ドイツ語 | 6185 | 9 | 1394 |
| Google Meet | 英語→スペイン語 | 6712 | 10 | 841 |
| Google Meet | 英語→フランス語 | 6183 | 9 | 1821 |
| Google Meet | 英語→イタリア語 | 7141 | 11 | 1045 |
| Google Meet | 英語→日本語 | 6977 | 10 | 2287 |
| Google Meet | 英語→韓国語 | 7139 | 11 | 1411 |
| Google Meet | 英語→ポルトガル語 | 7097 | 10 | 608 |
| Google Meet | スペイン語→英語 | 7237 | 11 | 819 |
| Google Meet | フランス語→英語 | 7783 | 10 | 823 |
| Google Meet | イタリア語→英語 | 6917 | 9 | 801 |
| Google Meet | 日本語→英語 | 6022 | 11 | 801 |
| Google Meet | 韓国語→英語 | 5045 | 10 | 611 |
| Google Meet | ポルトガル語→英語 | 6260 | 10 | 1339 |
| Microsoft Teams | ドイツ語→英語 | 8443 | 10 | 2376 |
| Microsoft Teams | 英語→ドイツ語 | 6957 | 10 | 1622 |
| Microsoft Teams | 英語→スペイン語 | 6793 | 10 | 1113 |
| Microsoft Teams | 英語→フランス語 | 7035 | 10 | 1895 |
| Microsoft Teams | 英語→イタリア語 | 6796 | 10 | 1865 |
| Microsoft Teams | 英語→日本語 | 6731 | 10 | 1089 |
| Microsoft Teams | 英語→韓国語 | 7053 | 10 | 2712 |
| Microsoft Teams | 英語→ポルトガル語 | 7170 | 11 | 1638 |
| Microsoft Teams | スペイン語→英語 | 7141 | 10 | 1461 |
| Microsoft Teams | フランス語→英語 | 7748 | 10 | 1704 |
| Microsoft Teams | イタリア語→英語 | 7117 | 10 | 1213 |
| Microsoft Teams | 日本語→英語 | 5923 | 10 | 330 |
| Microsoft Teams | 韓国語→英語 | 5071 | 10 | 1356 |
| Microsoft Teams | ポルトガル語→英語 | 6160 | 10 | 1660 |
| Zoom | ドイツ語→英語 | 7528 | 9 | 1865 |
| Zoom | 英語→ドイツ語 | 6780 | 10 | 1774 |
| Zoom | 英語→スペイン語 | 7176 | 11 | 1611 |
| Zoom | 英語→フランス語 | 7023 | 10 | 1726 |
| Zoom | 英語→イタリア語 | 7099 | 10 | 1732 |
| Zoom | 英語→日本語 | 7066 | 10 | 2050 |
| Zoom | 英語→韓国語 | 6787 | 10 | 1716 |
| Zoom | 英語→ポルトガル語 | 6805 | 10 | 1679 |
| Zoom | スペイン語→英語 | 6257 | 9 | 1552 |
| Zoom | フランス語→英語 | 6903 | 9 | 1599 |
| Zoom | イタリア語→英語 | 6328 | 8 | 1725 |

| | | | | |
|------|-----------|------|---|------|
| Zoom | 日本語→英語 | 5207 | 9 | 1202 |
| Zoom | 韓国語→英語 | 4619 | 9 | 1217 |
| Zoom | ポルトガル語→英語 | 5504 | 9 | 1380 |

最終的な結論

本報告書で示された最終的な結論は、DeepLおよびその他の第三者から独立したものです。Slatorは、本分析において特定のプラットフォームを人為的または意図的に優遇することはなく、完全に独立かつ中立的な分析を提供するよう努めました。Slatorは本調査手法を独自に設計し、分析および調査結果に関する編集上の決定権を全面的に保持しました。

Slator について

Slatorは、翻訳、ローカライゼーション、通訳、言語AIに関する調査および市場分析の主要な情報源です。Slatorのアドバイザー部門は、M&Aサービスや独立した分析を求めのお客様にとって、信頼できるパートナーです。

Slatorはチューリッヒ（本社）とロンドンにオフィスを構え、アジア、ヨーロッパ、米国にアナリストを配置しています。

チームメンバー

FLORIAN FAES

Slator
マネージングディレクター
E : florian@slator.com

ALEX EDWARDS

Slator
コンサルティング部門
責任者
E : alex@slator.com

ROCIO TXABARRIAGA

Slator
シニアリサーチアナリスト
E : rocio@slator.com

