

ステークホルダーによる効果的なUTX用語管理

— TBX との使い分けから用語書き換え・ポスト エディットへの応用まで —

Effective UTX Terminology Management by Stakeholders



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、ペレ出版）。

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

1 はじめに

企業や組織では「だれがどのように用語管理すべきか」という問題がある。日本の企業で部門が分かれている場合、それぞれの部門が個別に用語を管理していることが多いのが現状だが、理想的には各部門が協力し、できれば全社規模で文書管理に取り組むことが望ましい。筆者は、2016年のISO/TC37 コペンハーゲン国際会議において、多言語用語集に関する提案を行った。本稿では、この提案とその他の事例に基づいて、UTX 用語集形式に基づく効果的な用語管理、そして用語集のステークホルダー（利害関係者）と用語管理の関係について述べる。

TC37 (technical committee 37) は、ISO の中で「ターミノロジー、その他の言語およびコンテンツ資源」を扱う技術委員会である。TC37 の下には5つのSC (sub committee、分科委員会) があり、筆者は、そのSCのうち、SC3「ターミノロジー、知識、およびコンテンツ管理のシステム」に属している。SC3のワーキンググループ3で、ISO規格の用語集形式であるTBX (ISO 30042)¹が議論されている。

なお本稿でのISOに関係する筆者の言及は、あくまでも筆者個人の見解であり、ISO/TC37の日本委員会やその他の委員会の意見ではないことに留意されたい。

筆者は、AAMT(アジア太平洋機械翻訳協会)からTC37に参加している。ISOのTBXグループとAAMTのUTXグループとの間で意見交換を進めることが、筆者がAAMTから参加している理由の一つである。

2 用語集形式 UTX 1.20

用語集形式 UTX(Universal Terminology eXchange) は、AAMT が策定した、シンプルかつ汎用的な用語集形式である。UTXの詳細は、<http://www.aamt.info/japanese/utx/> を参照されたい。UTXの最近の活用事例として、株式会社デンソーの技術開発センター DP-EDA 改革室 (DENSO Project - Electronic Design Automation) では、用語整理を進めるために、用語集をUTX形式に変換した。これは、SDL Trados用語ベースなどに変換して活用されている。

2016年に、UTXの最新版であるUTX 1.20仕様書がアジア太平洋機械翻訳協会から正式公開された。現行バージョンのUTX 1.11からの変更点はかなりあるが、UTX 1.20で導入された新機能を使わない限り、UTX 1.11から移行するのは簡単である。ヘッダーの一部を書き換えるだけで済むこともある。UTX 1.20仕様書では、全体的に説明や具体例が詳しくなり、ユーザーにとっては使いやすくなった。構造的に大きな変更としては、新しい用語ステータスのしくみである「言語ごと用語ステータス」を追加した。これにより、用語集に含まれる言語ごとに用語ステータスを示すことができる。UTX 1.11では、簡便性を重視して、言語ごとではなく、一列にまとめて用語ステータスを示していた。言語ごと用語ステータスにより、完全に双方向の対訳用語集や、3つ以上の言語を扱う多言語用語集が作成できるようになった。

3 用語管理はステークホルダーが行う

用語管理は、用語集のステークホルダー（利害関係者）が行う必要がある。用語管理とは、ある企業や組織で使用される用語を規定し、各用語に整合性と一貫性がある状態にすることである。その際、文書作成や翻訳でその用語をすぐ適用できるように、専門知識に基づいて、正確性や妥当性を保証し、不足する情報を補い、あいまいな点や不要な情報を取り除く。

用語管理をするうえで、特定用語に関係があり、実際にその語の意味や使う状況を知っている人（すなわちステークホルダー）の意見がまったく反映されない状態は避けるべきである。逆に、ある用語が適切かどうかの判断は、自分でその用語を使わず、該当分野の専門知識がない人（すなわちステークホルダーではない人）にはできない。だが、組織内の用語管理は理想的とはいえないことも多い。よくある問題は、用語管理の権限が必要な人に権限がないことである。また、ステークホルダーであっても、用語集への関わり方には違いがある。

UTX 1.20 仕様書ⁱⁱでは、ユーザー ロール、つまりユーザー（ステークホルダー）の用語集への関わり方を規定している。以下は、同仕様書でユーザー ロールを定義している箇所である（図 1 も参照）。

「UTX 用語集には、3つのユーザー ロールがある。これらのロール（役割）は、ユーザーが用語集にどのような操作を行えるかを定義する。

用語集ユーザーは、用語集を使用する個人である。用語集ユーザーは、用語集の分野の知識と、UTX についての最低限の知識を持つ。用語集ユーザーは、用語を追加、削除、変更することはできないが、コメントや他の手段により、用語に関する意見および好みを示すことができる。

用語集提出者とは、用語集への新しい追加項目を提案する個人である。用語集提出者は、用語集の該当分野の知識を十分に持ち、さらに UTX の基本的な知識を持つ必要がある。用語集提出者には、用語集ユーザーのロールが含まれる。

用語集管理者とは、ある用語集に対して責任を持つ個人である。用語集管理者は、用語集の該当分野の専門的知識を持ち、さらに UTX について深い知識を持つ必要がある。用語集管理者は、ある項目が用語集での適切な項目として承認されるか、（必要なら）削除されるべきかを決定する（参照「4.5.2 approved（承認）」）。また用語集管理者は、用語ステータスの割り当てや変更ができる。用語集管理者には、用語集提出者のロールが含まれる。1人または複数の**代理者**が、用語集管理者の役割を代行できる。」

（（参照「4.5.2 approved（承認）」は別項目への参照）

上記のロールは UTX 1.20 で規定されているものだけであり、実際の用語管理ではステークホルダーにはこれ以外のロールもありうる。これらのロールは、ツールやサービスでのアクセス・編集権限として実装できる。

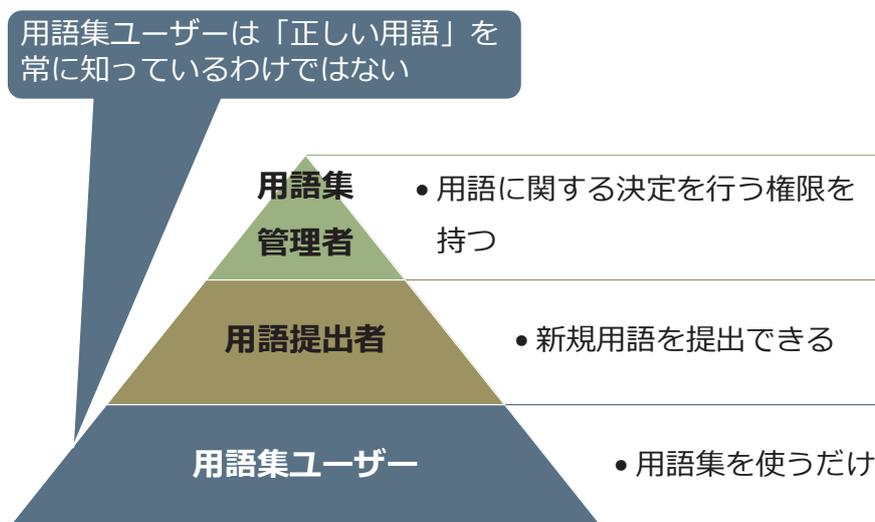


図 1 UTX 1.20 のユーザー ロール

用語集は、単に専門用語の雑多な集合ではない。用語とは、特定の分野、特定のグループで使われる用語の決まり事、合意事項である。用語集はそれらの用語集の集合であるが、それぞれの用語の相互の関係も考える必要がある。さらに、用語管理では、「**用語集をだれがどのよう**に使うのか」を考えることが重要である。専門用語は、その分野に関心のある人にとっては重要でも、そうでない人にとってはまったく関心のない言葉であろう。用語集は、それらの用語を日常的に使う組織でのみその価値が認識される。たとえば、ある用語集が対訳用語集の場合、翻訳者は、言葉を正確に理解して翻訳する義務がある。翻訳者は、ステークホルダーの中では最も直接的に、あいまいな定義や用語のばらつきなど、用語管理の問題の影響を受ける。翻訳者は、「一貫性のない訳語は統一しないと困る」と考えるだろう。それにもかかわらず、翻訳者がステークホルダーとして関われるのは多くの場合、用語提出者としてのみである。

翻訳プロジェクトに関わる立場の中で、翻訳会社は、ステークホルダーではあるが、(用語集管理そのものを依頼されたのでなければ)用語集管理者になることは適切ではない。

文書の最終消費者、(最終的な読み手。たとえば製品マニュアルを読む消費者)もステークホルダーである。最終消費者は、バラバラの用語を読まされて混乱し、不利益を被るにもかかわらず、文書の作成過程にはまったく関わるできない。企業にとって、このような最終消費者は重要な存在のはずだが、問題としては表面化しづらい。用語集管理者は、最終消費者に代わって、彼らにとって分かりにくい用語がないかの意見収集をする必要がある。

翻訳では、発注企業が最も直接的なステークホルダーであり、用語管理の重要性を最も深く理解する必要がある。翻訳発注企業の管理者の認識が、用語管理の成功の鍵となる。前述のように、翻訳者、翻訳会社、最終消費者は、用語管理の問題に影響を受けることはあっても、用語管理には直接関わるできないからである。企業の用語管理では、「**自分たちが使う用語は自分たちで決める**」という**自覚**が重要と思われる。自分がステークホルダーであれば、「だれかがなんとかしてくれる」という他人任せにすべきではない。用語管理には一定のコストがかかるが、コストを上回る費用対効果を上げるこ

とができる。

大規模な用語管理は、一般に想像されるよりも複雑な工程である。適切な用語管理を行うには、該当分野の専門知識に加えて「用語管理」という分野そのものの専門知識を要する。用語は継続的に検討されるため、新語が追加されたり、既存の用語が変更されたりすることもある。用語調整は、harmonizationと呼ばれることもある。利害、意見を調整し、合意形成が得られて、用語が確定する。体系的な方法で取り組まなければ、用語集は混乱してしまう。だが、用語管理の専門家のみがステークホルダーではない。企業の用語担当者は、該当分野の専門知識は持っていても、用語管理の専門知識を持っているとは限らない。むしろ、用語管理専門家はごく一部であり、用語管理の経験がない人、用語管理に間接的にしか関わらない人のほうが多数派であろう。間接的関与は完全に除外することはできないが、非ステークホルダーが用語管理に関わることを減らす一方、ステークホルダーであればなるべく用語管理に関わるようにする体制が必要となる。

用語集管理者のみでなく、用語提出者も、どのような用語をなぜ追加したほうがいいのか説明できる程度の用語管理知識が必要になる。用語ユーザーには、高度な用語管理の知識は要求されないが、用語の重要性について理解しておく必要がある。また、用語ユーザーは目視のチェックだけでなく用語チェックツールを使える技能を習得する必要がある。理想的にはすべての企業文書が配布されるまえに自動的に言語チェックされることが望ましい。ただ文書は、文字入力を終えた直後にメールなどさまざまな方法で配布可能であるから、自動であってもチェックを徹底することは困難でもある。スペルチェッカーも用語管理の応用の一つだが、軽視されていることが多い。誤検出を減らして用語チェックを効果的に行うには、用語データに不適切な用語が含まれないよう管理する必要がある。

次節では、用語管理の具体的な手法としてクラウドを使用する方法について触れる。

4 クラウドと用語承認過程

用語管理とは合意形成の過程であるともいえる。ステークホルダーたちが合意形成しながら用語データを共

5 ISO 用語集規格 TBX の状況

TBX は、ISO で定められた用語集規格 (ISO 30042) であり、XML 形式に基づく。日本では必ずしも TBX の知名度は高くない。この節では、TBX の状況について述べる。

単純化すると、ISO 参加国の中で TBX を使う国とそうでない国、また、各国内で TBX を使う企業とそうでない企業が二極化している。

TBX を使う国は、一般的に欧米が多い。多言語翻訳、ひいては多言語用語集の必要性が高く、用語管理の経験が豊富な国である。一方、日本は TBX を使わない国である。最大の理由は、体系的翻訳の歴史が浅く、実務翻訳が企業や大学から軽視されていることである。また、翻訳の需要が英日・日英翻訳に集中しており、多言語用語集は必要性が少ない。多言語翻訳では日本語を軸とすることは難しく、英語を軸にすることが多いということもある。**多言語を常に念頭に置く欧米と、外国語はひとまず英語のことだけを考えておけばいい日本では、用語に対する意識に大きな違いがある。**

企業というレベルで考えると、TBX を使う企業は、大規模で、多言語翻訳の需要があり、なおかつ体系的翻訳に取り組む意識が高い企業である。これに対して小規模の企業では、TBX について学習し、活用する余裕がない。また大企業でも、企業の各部門が縦割りで、全社を統括する文書管理部門がないこともある。この場合、文書資産・翻訳資産を前者で集約的に管理するという発想がなく、各部門が個別に管理している。そして、各部門に文書管理の専門家がいるわけではないため、文書の品質を維持できない。

TBX 形式は、構造を深く理解しなくとも、ある程度は用語ツールで利用できる。ただ、TBX をゼロから作成するなど、高度な利用をするには、TBX 形式の複雑な構造について学ぶ必要がある。また用語管理全体の理解とともに、XML についても知る必要がある。一方、単純な用語管理のみで必要十分な場合、TBX は複雑すぎる。TBX は、欧米での需要とは一致しているが、日本では大企業以外は TBX を活かすことがしにくい。次節では、ISO コペンハーゲン国際会議での、筆者の UTX と TBX を組み合わせた提案について紹介する。

同編集する方法には、ファイルベース、クラウドベース、サーバーベースの3種類がある。ファイルベースの方法では、ファイル送信サービスやメールで用語データファイル（用語ベース、TBX、UTX など）をやり取りし編集する。これに対して、クラウドベースやサーバーベースの用語管理では、用語への変更は即時に行われる。これは一見便利ようだが、慎重な検討と承認の過程を要する用語管理では注意が必要である。前述のロールに基づいた編集権限を設定する必要があるし、暫定的な用語の登録を許可するか、もしそうならそれらをどのように扱うかも決定する必要がある。

クラウドベースの用語管理には専門ツールまたはサービスもあるが、単純または小規模の用語集であれば Google ドキュメント (Google シート) や OneDrive (Excel Online) などの無料のクラウド表計算サービスも使用できる。UTX 1.20 のページには Google シートの実例を掲載してある。この例では表計算シートを HTML に埋め込み、編集不可としているが、表計算シートそのものを共有することもできる。

理屈からすると適切な用語であっても、実際にはだれにも使われていないことがある。その場合は、広く使われている用語のほうを採用する、という判断がされることもある。クラウド用語管理では、このような用語ユーザーの意見を比較的収集しやすい。

ただし、用語ユーザーの意見とは関係なく、用語集管理者の観点が重要なこともある。既存の用語が分かりにくい、誤解を生むなどの問題があれば、その用語がどれほど定着していても、新しい用語に合わせるべきだ、という判断がされることもある。用語を収集する元データとなるコーパス内の文はすべて適切であるわけではない。Google でのヒット数を参考にして、文書の用語、または翻訳の訳語を決定することはよくある。しかし、ヒット数が多くても誤用、不適切な使用がされていることもしばしばある。統計は非常に有用な手法ではあるものの、統計結果をそのまま受け入れて用語・訳語決定することはできない。統計はあくまで用語・訳語決定抽出と、その決定の参考情報として使われるべきであり、最終的には用語専門家（用語集管理者）の判断が必要になる。

次節では、ファイルベース用語管理で使われる TBX の状況について述べる。

6 ISO コペンハーゲン国際会議での多言語用語集に関する提案と反応

筆者は、ISO コペンハーゲン国際会議で現地時間2016年6月28日9時からのSC3 WG3会議で、十数分程度、多言語用語集に関する提案について発表と質疑を行った。繰り返しとなるが、本稿でのISOに関係する筆者の言及は、あくまでも筆者個人の見解であり、ISO/TC37の日本委員会やその他の委員会の意見ではないことに留意されたい。

筆者は、当初、コペンハーゲンに直接行って会議に参加することを想定していたが、残念ながら予算の関係上叶わなかったため、この会議を含め、ほぼすべてのSC3の会議にオンラインで参加した。筆者の提案の骨子は、以下のとおりである。(ISO/TC37は用語の技術委員会であるが)現在のISO/TC37の各規格では、英語2234語とフランス語967語の訳語は決まっているが、それ以外の言語では用語集として用語が規定されていない。各ISO規格には"Terms and definitions"の節があり、そこで規定された用語があるので、各国語でそれらに訳を付けたうえで、多言語用語集としてまとめるのはどうか、というものであった(図2)。また、非専門家を含むステークホルダーの意見を反映する現実的な手法として、シンプルなUTX用語集が最適である。UTXはTBXに変換(または統合)できる。さらに、ISO多言語用語集はCreative Commonsなどで広く公開するのがISOの普及にも貢献するのではないかと提案した。

この、多言語用語集に関する提案は、臨時の議題であるにもかかわらず、参加者の関心は高く、活発な質問とコメントがあった。参加者からはたとえば以下のような質問があった。「多言語用語集作成を各国委員会に強制するのは難しいのではないのか」。筆者の回答は、強制

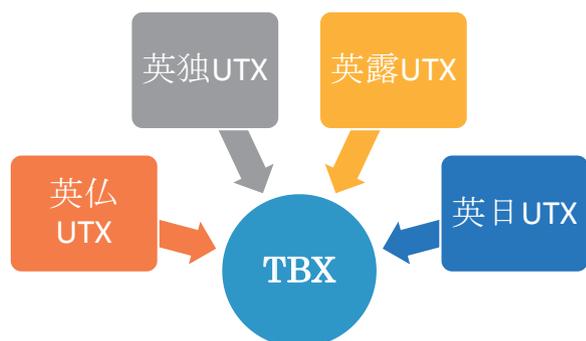


図2 多言語用語集

は必要ない、というものである。ISO規格の各国語翻訳と同じように、用語策定も、特定規格に関心のあるステークホルダーのみが自発的に行うべきである。日本のTC37で日本語に翻訳されたのはISO 17100など一部のISO規格のみであり、それはその規格のステークホルダーがいたからである。ただ、TC37以外のISO委員会でも用語の調整は大きな課題となっている。たとえばコペンハーゲン会議では、ウクライナのTC20委員会から航空・宇宙業界での用語調整についての説明があった。定義された用語は複数規格にまたがっていることも多く、また重複も多い。安全管理が人命に関わる航空・宇宙業界では、確実な理解のために用語調整は特に重要である。

もう一つの質問は、「TBXやその他のファイル形式があるのになぜUTXが必要なのか」ということである。前述のように、TBXは、欧米などターミノロジスト(用語専門家)のいる翻訳先進国では活用されている。一方、UTXは、専門知識が不要なため、用語管理の経験がない委員や、これまでTBXを敬遠してきた委員でも取り組める。

提案にも含めたとおり、私見ではあるが、TC37全体の用語集を直接多言語化するよりも、規格に記載済みの用語を取り出し、それらの対訳を付けるという手順が現実的と思われる。言い換えれば**トップダウンではなくボトムアップでの作業が望ましい**。ある用語が(その同義語・異表記を含め)各言語で同じ概念を示すことは一意の「概念ID」によって識別される(UTX 1.20仕様書を参照)。

この他、「日本を含め用語管理が浸透していないアジアでは、UTXは用語管理の入り口として有用であろう」というご意見をいただいた。

ISO/TC37で多言語用語集がどのように実現されていくかは今後の議論に注目したい。またこのような用語管理がTC37以外でも広まることに期待したい。次節では、軽量のUTXと強力なTBXの使い分けについて述べる。

7 軽量のUTXと強力なTBXの使い分け

TBXとUTXは、いわゆる適材適所として相互補完的に使い分けることが望ましい。TBXの仕様は、多言

語やマルチメディアなどを含む複雑な用語集を想定している。だが、シンプルな対訳用語集のみが必要十分である場合には複雑すぎ、編集も困難である。一方、UTXは、翻訳や文書作成の現場作業に必要な情報のみにフォーカスしている。

また、UTXは、日本で策定されたことから、日本語に関するいくつかの事情が考慮されている。UTXは表形式であることから用語の一覧表示が容易である。さらに、標準化された用語ステータスに基づいてExcelなどで色付けの書式設定をすることで、**用語を「見える化」することができる**（図4を参照）。どのような用語の概念で多様な表記があるか、禁止語を伴う用語がどれかなどが視覚的にはっきり分かるため、合意形成がしやすい。

日本語を含む用語集で一覧表示が特に必要とされるのは、漢字、ひらがな、カタカナなどの表記が多様であり、用語統一以前に、用語を一覧して「用語の表記統一」が必要となるからである。英語にもアメリカ英語やイギリス英語などのつづりの違いはあるが、center/centre、-ize/-iseといったわずかな差異でしかも規則的なものが多い。日本語の用語の表記統一はこれと比較すると、統一すべきバリエーションが多岐にわたる。また日本語を含む用語集では複合語が多い。単一の語のみでなく、複合語の構成要素を一覧で比較すると、前述の表記ゆれが見つかることが多い。たとえば「先物取引」と「先物取り引き」のような不統一である。

ISOのTBX検討グループ(SC3 WG3)でも、複雑化しすぎたTBXの問題を認識しており、TBXを簡略化した「方言」の検討が進められている。TBX-Basicは、TBXを簡略化した方言で、TBX-MinはTBX-Basicをさらに簡略化した方言である。これらの方言に対して本来のTBXはTBX-Defaultと呼ばれる。これらの試みはすばらしいが、今後、用語の専門家でない人にとってTBX自体の構造が理解しやすくなる可能性はほぼない。簡略化された方言であっても、TBXを作成するには、用語集の階層構造(TMF(Terminological Markup Framework)ISO 16642)とその意義を理解することが必要となるからである。これは、UTXのようなシンプルな用語集形式は今後とも存在意義があるということでもある。

UTXは、翻訳の作業に必要な最小限の情報から出発し、必要に応じて拡張できる。この方向性は、用語集をゼロから作る時、専門の用語集管理者を常時用意できないとき、用語集の規模が比較的小規模、または構造が単純なときに最適である。一方、TBXは、人的資源の豊富な環境での大規模翻訳で、最大限に効果を発揮できる。用語集をゼロから作る場合は、当初はUTXで作成して、用語集が大きくなった場合、より複雑な構造が必要になった場合にTBXに移行することが最適と考えられる。

日本でよくみられる対訳用語集(日英・英日など二言

アメリカ英語／日本語／中国語（簡体字）

#UTX 1.20; lang: en-US/ja/zh-CH; copyright: AAMT (2016); license: CC BY 4.0					
#term:en-US	tgt:ja	term status:en-US	term status:ja	concept ID	term:zh-CH
#用語 (アメリカ英語)	用語 (日本語)	用語ステータス (アメリカ英語)	用語ステータス (日本語)	概念ID	用語 (中国語簡体字)
Asia-Pacific Association for Machine Translation	アジア太平洋機械翻訳協会	approved	approved		亚洲太平洋机器翻译协会
contributor	用語提出者	approved	provisional		用語提交者
entry	項目	approved	approved	1	条目
entry	エントリー	approved	forbidden	1	
unidirectional	一方向	approved	approved	2	单向
monodirectional	一方向	non-standard	approved	2	

図3 UTX用語ステータスによる「見える化」

語の場合) や、小規模かつ単純な用語集での用語管理では、用語ベースではなく、UTX (あるいは UTX に基づく Excel) で管理するほうが簡単である。完成度の高い用語集をごく一部修正するだけなら TBX や用語ベースでもよいだろう。ただ、日本で一般的な傾向の一つは、方針が不明確なまま作られた用語集を整理する必要があるという状況である。多数の不適切な項目が混入しており、分野分けもされていない。このように、用語集の内容の大きな見直しが予想される場合、TBX や用語ベースでは編集がしづらい。また、英日・日英翻訳のみの場合も、UTX で必要十分といえる。たとえば、言語が日英のみで、原語、訳語、品詞、定義など 4 列しか情報がないのに、TBX のような XML を使用して管理する必要はない。次節では、機械翻訳での UTX の活用について述べる。

8 用語書き換えデータとしての UTX

UTX は、「機械支援による言い換え処理」にも利用できる。「産業日本語」は、「産業・技術情報を人に理解しやすく、かつ、コンピュータ (機械) にも処理しやすく表現するための日本語」と定義されている。また「機械支援による言い換え処理は今後の課題」とも記されている。ⁱⁱⁱ

UTX は用語を決定する枠組みとして、さらにはその用語を活用する枠組みとして利用できる。たとえば、あいまい、難しいなどの理由から「望ましくない用語がど

れか」ということの合意形成ができる。さらに、UTX 変換ツールで禁止語リストや置換リストの抽出が可能である (図 4 を参照)。図の中で、approved が承認語、forbidden が禁止語である。禁止語リストは禁止語と指定された語のみのリスト、置換リストは禁止語を承認語で置換するためのリストである。これらの禁止語リストや置換リストは、さまざまな用語チェック・置換ツールなどで活用できる。ただ、UTX は名詞の書き換えには効果的だが、動詞や形容詞などの活用などを含む置換は、UTX 以外の方法も必要となる。

今後、産業日本語でも、用語書き換えのしくみとして UTX とその関連ツールを活用いただければと考えている。

9 機械翻訳でのポスト エディットと UTX

アジア太平洋機械翻訳協会で策定された UTX は、当初は機械翻訳、特にルールベース方式のパッケージ翻訳ソフトで使用されることが想定されていた。その後、機械翻訳ではルールベース方式に代わり、統計方式が目まぐるしく注目を集めている。ただ、統計機械翻訳でも用語データが重要であることに代わりはない。

しばしば見過ごされているが、翻訳の原文文書の品質を上げるためにも用語集は必要である。UTX 用語データおよびチェック ツールを活用して用語統一を図ることで、原文文書の品質を上げることができる。統計機械翻訳のベースとなるコーパスは、本来であれば表記基準

UTX用語集で原語が同じで訳語が異なる場合

term:en	term:ja	term status:ja	yomi:ja
bedsore	褥瘡	forbidden	じょくそう
bedsore	床ずれ	approved	とこずれ
aggravate	増悪させる	forbidden	ぞうあくさせる
aggravate	悪化させる	approved	あつかさせる

↓ 置換リストとして抽出できる

置換前	置換後
褥瘡	床ずれ
増悪させる	悪化させる

置換して
分かりやすい用語に統一できる

図 4 禁止語と承認語

に基づき、用語の一貫性がある品質を持つことが望ましい。欧米言語では、体系的翻訳の取り組みがあるため、表記基準の準拠と用語の一貫性はある程度実現されている。だが、日本語に関しては、体系的翻訳の発想が根拠いておらず、また論理的作文が徹底していないこと、さらに表記が多様であることも災いして、欧米言語と比較するとコーパスの品質は低い。

また、原文の品質向上に加えて、ポストエディットでの対訳用語チェックでも UTX 用語集を活用できる。欧米言語と比較して英日・日英などでは、前述のようにコーパスの品質が低いことに加え、言語構造の違いが大きいことから、機械翻訳は精度が低い。特に用語レベルでの翻訳正確性は、統計機械翻訳では低くならざるを得ない。「文としては正確性が高い・自然である」場合でも、用語が間違えているのは致命的である。特にポストエディットでは、全面的な書き換えを必要とする事態が多い。UTX 用語集で対訳用語チェックを行えば、少なくとも用語レベルでの正確性を保証できる。UTX での用語チェックを含む各種ツールと、用語チェックの具体的な方法はこちらのページ^{iv}で示している。

10 TC シンポジウム

最後に、TC シンポジウムについて述べておきたい。

2016年8月24日・25日の両日、筆者は、テクニカルコミュニケーター協会が主催する TC (technical communication) シンポジウムにアジア太平洋機械翻訳協会から参加した。このシンポジウムは、産業日本語と方向性が重なる点が多く、興味深かった。マニュアル作成を行う企業が多数参加しており、「分かりやすく書く」ことを日々実践されている現場の方たちの話を聞いたのは貴重であった。

24日午後には、「テクニカルコミュニケーション分野における専門用語の統一活動—Terminology of Technical Communication—」を聴講した。ドイツを中心とするテクニカル コミュニケーション組織 tekomp の Michael Fritz 氏による講演である。Fritz 氏の用語集作成に関わる問題意識は、筆者が日頃から考えていることと共通する点が多く興味深かった。また、日英機械翻訳に関する別の発表では、日英の場合はルールベース方式のほうが統計方式よりも精度が高かったとい

う事例が紹介された。どちらの方式が良いかは実際にはさまざまな要因が関係してくるとはいえ、これは一般論としては筆者の経験と一致する結果である。近年では統計機械翻訳が注目を集めているが、用語データを活用でき、外部への情報流出の危険がないルールベース機械翻訳ももっと見直してもよいのではないだろうか。

TC シンポジウムのその他の各種セッションでは、今後、用語管理への関心がもっと高まって欲しいと感じた。「用語集の重要性は分かるが今はまだ取り組めていない」という声も聞かれた。また「用語統一してしまうとマニュアルが機械的になる」という意見も伺った。この懸念はもっともであるが、用語統一されていないと、訳語にもばらつきが出るなど翻訳に大きな支障があるのもまた事実である。

11 まとめ

本稿では、用語管理とステークホルダーの関係、クラウドと用語承認過程、ISO 用語集規格 TBX の状況、ISO コペンハーゲン国際会議での多言語用語集に関する提案と反応、軽量の UTX と強力な TBX の使い分け、用語書き換えデータとしての UTX などといった項目について述べてきた。

今後、UTX チームが取り組む方向性としては、各種変換ツールでの UTX 1.20 対応の促進、機械翻訳での活用の実証研究、UTX と TBX その他の形式との間での変換をより確実にする補足仕様書の策定などがある。

参考文献

- i Introduction to TermBase eXchange (TBX). (n.d.). Retrieved August 30, 2016, from <http://www.tbxinfo.net/tbx-dialects/>
- ii AAMT. “UTX 1.20 仕様書”. AAMT. <http://www.aamt.info/japanese/utx/utx1.20-specification-j.pdf>, (参照 2016-8-30).
- iii 特許版・産業日本語. “特許版・産業日本語について”. 特許版・産業日本語. <http://japio-tjp.org/gaiyo.html>, (参照 2017-8-29).
- iv 秋桜舎. “UTX 関連のツール”. 秋桜舎. <http://cosmoshouse.com/utx/tools.htm>, (参照 2016-9-1).