

UTXと、用語管理、翻訳品質、および標準化の関連性

－ ISO/TC37 松江国際会議に参加して－

UTX and its relation with terminology management, translation quality, and standardization
 – Report on the ISO/TC37 2015 annual working meetings and plenary –

秋桜舎 代表 **山本 ゆうじ**

PROFILE

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』。

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

1 はじめに

本稿では、用語集形式 UTX と、用語管理、翻訳品質、および標準化の関連性について、特に ISO/TC37 松江国際会議に筆者が参加して得た知見に基づいて考察する。

2015年6月20日から26日にかけて、ISO/TC37 松江国際会議が、島根県松江市のくにびきメッセで開催された(写真1)¹。アジア太平洋機械翻訳協会(AAMT) WG3 (通称 UTX チーム) グループリーダー 山本ゆうじ(筆者)が、24日から26日にかけて本会

議に参加した。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。TC37では、SC5「翻訳、通訳および関連技術」のメンバーが最も多いが、翻訳・通訳そのものはその一部でしかない。とはいえ、翻訳・通訳は、SC3を含め、TC37の他のSCにも関係している。

用語集形式 UTX(Universal Terminology eXchange) <<http://www.aamt.info/japanese/utx/>> は、AAMT(アジア太平洋機械翻訳協会)が策定した、シンプルかつ汎用的で、オープンな用語集形式である。

本稿の各節では、用語集と産業日本語の関係、用語管理による翻訳品質の改善といった背景を先に紹介する。

¹ 会議サイト <<http://lang.cs.tut.ac.jp/isotc37/>> なお ISO/TC37 国際会議が日本で開催されるのは今回が初めてである。また、ISO/TC37 日本国内委員は、登録上、JISC(日本工業標準調査会)の名前で会議に参加する。



写真1 会場・くにびきメッセ

その後、ISO/TC37 松江国際会議で筆者が参加したTBX グループ（後述）での会議、そこで行った UTX のプレゼンテーション、UTX と TBX の関係、会議参加者についての情報と考察、会議総会とそこで行われた決議について述べる。

2 用語集と産業日本語の関係

ISO/TC37 松江国際会議の報告の前に、用語集、用語管理と産業日本語の関係について簡単に述べる。特許版・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²。

本稿での**用語集**とは、文書や翻訳の品質を用語レベルで向上することを目的として、専門用語、そしてその対訳を定め、集積し整理したものを指す。

また本稿での**用語管理**とは、用語集を単に専門用語の雑多な集合とするのではなく、用語が文書作成や翻訳で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、あいまいな点や不要な情報を取り除き、各用語に整合性と一貫性がある状態にすることである。特に、特許のように、専門用語が極めて多く、新規の概念を含む分野では、用語管理により、文書作成・翻訳・検索での効率向上が望める。

用語集、より正確には用語データと用語ツールを活用することで、文書の翻訳や作成をする際に、正確な用語の適用はもちろんだが、分かりやすい表現の適用と管理もできる。特に、不必要な新語・造語・難解な語、読者の混乱を招く異表記を、「用語」という観点からチェックできる。

このことから、用語管理による理解度の向上は、産業日本語が目指す方向と合致すると思われる。具体的には、制限言語ほど厳しく制限を行わなくても、「分かりにくい表現や語句」を用語データとして集めることで、自動的なチェックが可能になる。また、そのような分かりにくい表現の代わりに使用すべき適切な用語（訳語）を置換候補として示すこともできる。そのために、UTX 用

語管理では、各用語に用語ステータスを付与し、その用語ステータス情報をさまざまなシステムで共有できるように標準化している。

3 用語管理による翻訳品質の改善

次に、用語管理により、翻訳品質をどう改善できるかについて簡単に述べる。

一般向けの体系的翻訳での翻訳品質の基本的な評価としては、たとえば以下のような観点がある。

1. 正確さ
 - 誤訳、訳抜けがない
 - 誤字脱字がない
 - 用語が正しく訳されている
2. 文書品質
 - 読みやすい
 - 用語・表現・表記が統一されている
 - 翻訳資産が再利用できる（→次回の品質・効率向上）

翻訳発注企業・翻訳会社・社内翻訳のような立場の違いや、機械翻訳など他の要素が加わると、これらの観点にはまた変更が必要になるだろう。

特に、グローバル企業での翻訳品質は、一般的な翻訳品質とはいくつか異なる観点がある。たとえば、「各言語での自然さ」と「企業メッセージの一貫性」のバランスを取る必要がある。つまり、訳文が日本語なら、日本語として不自然でないようにする一方で、原文に含まれる企業としてのメッセージが確実に伝わるようにする必要がある。だが、用語のレベルだけを考えても、英語のある用語が、日本語での用語に常に一対一で対応するとは限らない。翻訳者が臨機応変に判断して適切な訳語を選ぶことも必要だが、その判断を翻訳者に常に押し付け、任せきりにすると、翻訳結果にばらつきがでる。そのような場合に、用語集で訳し分けを明確に規定しておけば、翻訳者の負担が減り、翻訳結果もより一貫したものにできる。各言語での用語の対応関係を探り、その用語の意味（概念）をすりあわせて調整し規定する作業は、「**概念調和**」とも呼べる。この概念調和は、たとえば英語の観点からのみ考えるのではなく、日本語のほうから

2 「特許版・産業日本語について」 <<http://japio-tjp.org/gaiyo.html>>



も考えて調整しないと、日本語翻訳の最終的な読者に正しいメッセージが伝わらない。三言語以上の多言語の場合は、それぞれの言語の組み合わせを多方向で考える必要がある。

このような概念調和が行われた用語データに基づいて翻訳や文書作成を行うと、読みやすい用語の提案、用語の正確な翻訳、用語の統一、翻訳資産の再利用、訳抜けや誤字脱字の防止など、翻訳品質をさまざまな面で向上できるうえに、翻訳効率も上げることができる。

4 TBXグループ(分科委員会3ワーキンググループ3) 会議への参加

前記の点を踏まえたうえで、筆者が、松江国際会議、特に TBX グループの会議に参加した詳細について説明する。

今回、AAMT として筆者が ISO 国際会議に参加した目的は、ISO/TC37 の国内委員としての責任と義務に加えて、以下である。

- TBX（後述）関連の会合に重点的に参加して、UTX と TBX の互換性向上について TBX グループのメンバーと協議する
- 海外の組織に UTX の利点を紹介してアピールする。パンフレットと UTX 要約仕様を配布し、ミニプレゼンテーションを行う

UTX は（少なくとも現時点では）ISO 標準規格ではないが、ISO 参加組織との情報交換は今後、UTX の普及に重要な意味を持つ。特に、筆者は、ISO/TC37 の国内委員ではあるが、ISO 国際会議への参加は初めてであり、多くの参加者と直接会って話し、信頼関係を構築することが重要であった。

今回、筆者は、SC3 の中でも、TBX³ を扱う、WG (working group) 3 の会議に主に参加した。ISO では、仕様が番号で呼ばれる他に、略語が非常に多く、部外者には非常に理解しづらい。SC3 ワーキンググループ3 の役割は厳密には TBX のみではないが、本稿では、便宜上「TBX グループ」と呼ぶことにする。他の分科委

3 TBX (TermBase eXchange) は、用語データの ISO 規格 (ISO 30042) である。詳細は <<http://www.tbxinfo.net/>> を参照。

員会では各国の利益がうまくかみ合わずに、緊張が高まることもあったようだが、TBX グループは、技術的な議論は活発にされたが、終始極めて友好的な雰囲気での会議が行われた。

欧米では、翻訳の歴史が長いことに加えて、「翻訳学」(translation studies) が確立されており、実務と文芸のそれぞれで研究が進められている。一方、日本では翻訳は独立した専門分野として認識されておらず、特に実務翻訳を専攻とする学科は、大学や大学院ではまだ少ない。今回は、翻訳・用語を学問的に長年研究してきた海外研究者の方々と親しく話させていただき、非常に勉強になった。

2011 年、筆者はボストンでの LISA (Localization Industry Standards Association) 会議に参加した。このころから、TBX の策定に当初から深く関わってきたブリガム ヤング大学 Alan Melby 教授の多大な協力を得て、TBX と UTX という用語集形式の相互変換の環境整備が進められた。筆者と TBX グループとは、メールやウェブ会議で継続的に意見交換を行っていた。

TBX には、TBX-Default、TBX-Basic、TBX-Min の3つの公式な「方言 (dialects)」がある⁴。TBX-Default が本来の TBX であり、TBX-Basic は TBX-Default を簡略にした形式、TBX-Min はそれをさらに簡略化したものである。この他に、法律など各業界などで使われることを想定した方言（つまり、特定業界の要求を満たす TBX のバリエーション）もありうる。TBX-Min と UTX は、シンプルさと実用性を重視する点で非常に共通点が多い用語集形式だが、TBX-Min は XML、UTX はタブ区切りという違いがある。

TBX グループの会議では、2008 年に策定された現在の TBX 仕様をリフレッシュする「モダン化」についての議論が行われ、その中でも複雑なままの TBX-Default 仕様をシンプルにすることの重要性がたびたび言及されていた。具体的には、TBX 仕様について主に XML 形式の観点から、<ref>、<xref>、<hi> のようなタグの定義に関する議論が行われた。また descrip や termnote といったタグを整理することについての議論も行われた。

4 "Introduction to TermBase eXchange (TBX): TBX Dialects" <<http://www.tbxinfo.net/tbx-dialects/>>

今回の会議は、筆者にとっては、TBX そのものについて理解を深める機会でもあり、また ISO での仕様策定・意志決定過程を見るのはおおいに参考になった。TBX グループの会議では、プロジェクトで TBX 仕様書を見ながら議論を進め、変更履歴で変更を記録していた。これは、AAMT での UTX 策定と同様の方法なので、非常に親近感があった。仕様書の説明不足の点、表現の統一などが議論されていた。また、実際に編集過程を目にすることで学ぶ点もあった。"Note" というセクションを設けて、重要な情報とは別に付加的な情報をまとめる、という方法も参考になり、UTX 仕様にも取り入れていきたい。

筆者は、TBX グループの会議に物理的に参加することは初めてであったため、仕様のユーザーに近い観点から仕様を改善することに貢献できたものと思う。たとえば、TBX 仕様の中では、「データ カテゴリ」の概念の説明が不足しており、分かりにくいのではないかと、いった意見を述べた。また <dig> というタグが直感的に理解しづらい点も指摘した。

5 UTX のプレゼンテーション、UTX と TBX (ISO 規格) の関係

6月25日のTBXグループの会議で、筆者はUTX(図1)を紹介するプレゼンテーションを行った。これは、本来の議題にはなかったものの、筆者が依頼して、TBXメンバーに特別に時間を作っていただいで実現した。このプレゼンテーションでは、主に以下の点を説明した。

- 用語集形式 UTX 仕様の概略
- UTX 策定の背景・意図
- TBX との棲み分け
- 日本での翻訳ソフトの状況
- UTX が MT と関連が深いこと
- UTX 変換ツールによる形式変換

このプレゼンテーションにより、UTX の意義をしっかりアピールできたものとする。

通常、日本国内で UTX について発表する場合は、用語管理についての理解が浸透していないため、「なぜ翻訳で用語管理が重要か」という点から説明を要するところである。今回は、用語管理の専門家のみが聴衆であったため、そのような前置きを一切しなくてもよいのが新鮮に感じられた。一方で、日本の MT の状況について

	A	B	C	D	E	F
1	#UTX 1.11; en-US/ja-JP; 2011-04-19T19:00:00+09:00; copyright: AAMT (2014); license: CC-BY 3.0					
2	#description: This is a sample dictionary for AAMT-related terminology. It is not an official dictionary.					
3	#src	tgt	src:pos	term status	comment	concept ID
4	entry	項目	noun	approved		1
5	entry	エントリー	noun	forbidden		1
6	dictionary	辞書	noun	approved		2
7	dictionary	ディクショナリー	noun	forbidden		2
8	merge	統合する	verb	approved	辞書について。	3
9	merge	マージする	verb	forbidden		3
10	optional	省略可能	adjective	approved		4
11	optional	オプション	adjective	forbidden		4
12	blank	空白	adjective	approved		5
13	blank	ブランク	adjective	forbidden		5
14	unique	一意の	adjective	approved		6
15	unique	ユニークな	adjective	forbidden		6
36	column definition	列定義	noun	provisional	field definitionのほうが望ましい。	
37	body	本文	noun			
38	glossary	用語集	noun	approved		7
39	glossary	グロッサリー	noun	forbidden		7

図1 Excel上のUTX用語集

は海外に知られていないため、説明が必要となった。たとえば、日本ではルールベース機械翻訳ソフトのパッケージが現在でも販売されており、ルールベースがすっかり下火になった欧米とは異なる事情であることなどを説明した。

プレゼンテーション後の質疑応答でも活発な質問があった。たとえば UTX がコンマ区切りではなく、タブ区切りである理由を問われた。これに対して、コンマ区切りの欠点を避けるための、検討を重ねた結果に基づく決定であるということをお返事した。

また、用語の「概念」に関する TBX と UTX の姿勢の違いを説明した。TBX は、いわば「**概念指向**」に基づく用語集形式である。TBX では、ある用語が意味している内容、つまり各言語に共通する概念が先にあり、その概念が各言語で用語として表されたときの関係を構造として表現できる。これは用語データを論理的に整理できる利点がある一方で、構造が複雑な入れ子状になっており、表計算のような形式で簡単に編集することができない。

一方で、UTX は実用面を優先した「**項目指向**」であるといえる。UTX では、翻訳者（あるいは機械翻訳システム）が最も必要とする最低限の情報、つまり「原語と、それに明白に対応する訳語」が最大の関心事である。用語データが概念に基づく構造を持っているか、ということは二次的な問題である。現場の翻訳者は、厳しい時間制限の中で作業しており、専門用語の概念としての位置付けに、いつまでも思い煩ってはられない。限られた時間で作業を進めるには、理屈よりも、実際の訳語としてそのまま今すぐ、自信を持って使えるかを知りたいのである。UTX は、表形式で簡単に編集できることを優先しているため、複雑な用語構造を扱うには適さない。そのような用途では TBX を使うことで棲み分けができる。筆者は、UTX ではシンプルな情報のみを必要とするシステム向けであるので、重要な情報のみに焦点を当て、些末な情報は切り捨てることを説明した。

また、TBX も UTX もさまざまな用語形式の相互変換に使える。だが TBX では特に「さまざまな形式の細かい部分についても確実に対応できるようにする」という方向性がある。国際標準という観点からあらゆる状況を想定する、という姿勢は理解できる。だが、時として

そのような姿勢は仕様を必要以上に複雑にしてしまう。その点では、UTX にはシンプルに作成・管理できる利点がある。UTX は、用語の専門家、つまりターミノロジストや用語管理者でない人や、XML の知識がない人がゼロから作成できる。つまり、XML ベースで機能豊富ではあるが、複雑な TBX と、シンプルな UTX が相互補完できる。

用語形式の相互変換では、UTX では情報維持よりも、実用性を優先している。一度ある形式に変換した後で完全に元の形式に戻せるような変換、いわゆる「ラウンドトリップ」変換が必要な場合は、豊富な機能を持つ TBX が適している。だが、実際には、性質の違うシステムでは、すべての情報を取り込めるわけではない。ある情報が適切に変換できないからといって、その処理に時間をかけすぎることはいできない。本質的な情報を優先したいときは UTX のほうが使いやすい。このような使い分けを広めることで、TBX と UTX の双方にとって有益な関係が築ける。

なお、UTX は、TBX との相互変換が可能である。UTX に関する変換ツールは、公式変換ツールを含め、以下で紹介している（公式変換ツールは、ソースコードを含め無料で公開されている）。

<<http://www.aamt.info/japanese/utx/tools.htm>>

当初から UTX は、ISO とその標準化の恩恵をこれまで受けてきた。IETF (Internet Engineering Task Force) による BCP 47 の言語タグ (ja, en など。ISO 標準を含む) の他に、ISO の日付形式に基づくなど、ISO に準拠する部分を含んでいる。また、UTX ファイルのエンコード方式である UTF-8 も ISO 標準でもある。このような標準化の恩恵は、関係者の真剣な討議（時にはユーモアもまじえながら）に基づくものであることが、今回の会議参加で改めて実感できた。

ISO では、標準化を目指す観点から、UTX で使用されている属性名をデータ カテゴリとして記載している。既存の仕様で使われるデータの性質を記載することで、属性名の乱立を防ぎ、互換性を高める意味がある。現行より古いバージョンの UTX の属性名であるが、<<http://www.isocat.org/rest/dcs/364>> にデータ カテゴリが記載されている。つまり、UTX の側から

も ISO に協力できることはあるし、またそうすべきと改めて考えさせられた。

6 会議参加者

会議参加者についても状況を報告する。海外からの参加者については、女性の IT 関連専門家・研究者の参加者が男性と比べて極端に少ないことはないようであった。日本でも女性の IT 関連専門家・研究者は増えつつあるようだが、海外の現状との差が感じられた。

また、会議全体を通じた感想として、参加者間のコミュニケーションが重要であると感じた。ふだんはメールやウェブ会議でやり取りし、実際に会うのは数か月から数年に一度であることもあるが、実際に顔を合わせることで信頼関係が生まれ、また強まる。

ISO 会議は英語で行われるが、たとえば総会での決議などは、公式言語であるフランス語でも記載が行われた。また、会議中にフランス語話者の懇親会が開かれ、筆者も参加した。これも参加者どうしの距離を縮め、円滑に会議を進める意味があった。

参加者はこの業界数十年のベテランの方たちが多い。本会議では、筆者は Christian Galinski 博士を始めとして ISO 規格を作ってきた多くの方々と新たに面識を得ることができた。斯界のベテラン、いわば専門家の中の専門家であるにもかかわらず、非常に気さくでありながら、高い技術や学識はもちろんのこと、先見の明があり、枠に囚われずに、物事を幅広い視野で捉えておられた方々であった。

TBX グループでは、アメリカ、ドイツなど欧米からの参加が多く、ネイティブや準ネイティブの会議は、専門用語を多用するうえに、非常にペースが速い。欧米以外からの参加は筆者一人であった。欧米以外の視点を取り入れるには、より多様な国からの参加者が TBX グループに参加することが望ましいと感じる。ネイティブや準ネイティブの英語での会議を一方向的に聞くだけでなく、主体的に参加して、専門的な議論を行うには、高度な英語力、たとえば TOEFL スピーキング スコアで 24 点以上は必要と推測される。ただし非ネイティブの割合が多いワーキンググループでは、状況が異なるかもしれな

い。

ISO/TC37 には、日本からさらに参加者が増える必要があると強く感じた。理想的には、メーカー企業からの、英語が流暢で、技術知識の造詣が深く、なおかつコミュニケーション・交渉能力に優れた人材である。

ISO では特定の国や特定企業・組織の利益のみに偏らず、フェアであろうとする精神はあるが、良くも悪くもお節節をすることはない。利害関係者（ステークホルダー）として組織が規格策定に関心があるならば、積極的に関わる必要がある。そうすることで初めて意見が反映される。ある組織が、ISO 活動そのものから距離を置いたままでなにか発言しても、その声が届くことはないだろう。

今回、日本が会議開催国であったことは大きな意味を持つと感じた。この会議では、(AAMT 前会長でもある) 井佐原均先生のご尽力と、松江市のご協力により、隅々まで行き届いた歓迎、充実したおもてなしが行われた。参加者からは、非常に満足したとの声が多く聞かれ、決議でも感謝の意が表明された。ISO の規格に関わる影響としては、文化的な価値観の違いも大きい。その違いをどの程度意識し理解しているかは、会議参加者で個人差がある。たとえば、日本文化に理解があるかないか、好感があるかないかで会議の行方が変わることはおおいにありうる。海外の参加者に対して、日本の立場を説明し、ふだんから理解を得る努力が必要であろう。



写真 2 総会会場

7 総会と決議

会議最終日には、各分科委員会総会と全体総会が行われた（写真 2）。TBX グループでの議論は非常に内容が濃かったため、TBX グループ以外の会議は、総会の時に短時間だけ見学できた。各分科委員会の総会では、（TBX グループを含む）SC3 全体では十数名、最大の SC5 で 30 名以上が参加していた。SC3 分科委員会総会では、Melby 教授が、XLIFF⁵などを策定した OASIS を ISO のリエゾン組織⁶として招待することを提案した。XLIFF ファイルには、用語部分が含まれる。筆者は、XLIFF に加えて DITA も含めることを提案した。また、Melby 教授は、AAMT もリエゾン組織として招待することを提案し、決議として承認された。筆者は、UTX については AAMT が ISO と今後、どう関わっていくかについて考えていたのだが、ISO でのやり方に慣れていないので、どうしたものか悩んでいた。そのとき、Melby 教授のこの提案が助け船となったわけである。もちろん、この決議はあくまで「リエゾン組織とし

5 XLIFF (XML Localization Interchange File Format) は、翻訳ツールで使われる、XML ベースのファイル形式である。詳細は <<http://www.oasis-open.org/committees/xliff/documents/xliff-specification.htm>>。

6 国内委員会とは別に、複数の国の利益に関わる、独立した組織として ISO に参加する組織。詳細は <http://www.iso.org/iso/home/store/publication_item.htm?pid=PUB100270>。

での招待」であるから、AAMT がリエゾン組織として参加するかどうかは、AAMT 自体の決定を待つこととなる。AAMT は今回の会議に一度参加したきりでなく、今後とも、話し合いに参加する道筋が見えたということである。なお、この決議では、AAMT が日本の利益を代表するだけの団体ではなく、南北アメリカの AMTA、ヨーロッパの EMTA と並んで、アジア太平洋地域を代表する国際的な側面を持つ団体であることが評価された。AAMT には、今後とも、日本以外の地域からのメンバーの参加が期待される。

8 まとめ

松江国際会議は、全体として参加者の満足度が高いように見受けられ、成功をおさめたものと思われる。

筆者としては、今回の松江国際会議で、AAMT から参加した目的である、UTX の利点のアピールを、プレゼンテーションなどを通して行うことができた。また、専門家との意見交換を積極的に行い、すべての議論の根底となる信頼関係を築けたものと思う。

来年度の ISO 国際会議は、コペンハーゲンで開催される（期間：6/26～7/1）。可能であれば、コペンハーゲン会議の参加を目指して、UTX グループ リーダーとして、今後とも TBX グループとの協議を続け、UTX の互換性を向上し、さらに UTX の普及を推進していきたい。

日本の各業界では、専門用語そのものや用語データ形式の標準化、用語管理、用語データの蓄積と再利用、体系的翻訳といった領域では、翻訳先進国の後塵を拝している状況である。だが2015年には、特許庁により約220万語の中日対訳辞書データがUTX形式で作成され、研究向けにALAGIN言語資源・音声資源サイトで無償公開された⁷。このような用語データが、今後さらに各分野・各言語で、UTXのような共有・再利用がしやすい標準形式で作られ、一般に広く公開されれば、特許はもちろん、あらゆる分野での文書や翻訳の品質と効率を大きく向上できる。

7 「用語集形式 UTX」 <<http://www.aamt.info/japanese/utx/achievements.htm>>。