http://www.aamt.info/japanese/utx/

AAMT（アジア太平洋機械翻訳協会）機械翻訳課題調査委員会

ユーザー辞書共通フォーマット UTX

Universal Terminology eXchange

「翻訳ソフトは役に立たない」

そう思っていませんか？

UTXの利点

- 翻訳ソフトの精度を大きく向上
- 翻訳資産を共有・再利用
- 単語を調べる時間と労力を削減

概要

UTX（Universal Terminology eXchange）とは、アジア太平洋機械翻訳協会（AAMT）が策定している、翻訳ソフトのユーザー辞書共通フォーマットです。2009年に、ユーザー辞書の標準化を目指すオープンなUTX仕様のひとつとして、シンプルなタブ区切り形式であるUTXの仕様が策定されました。AAMTは、機械翻訳の研究開発者、製造販売者、利用者の三者から構成される団体です（機械翻訳は、翻訳ソフトの核となる技術です）。この仕様は、「UTX-Simple」と呼ばれていましたが、2011年4月以降、単に「UTX」と呼ぶことになりました。

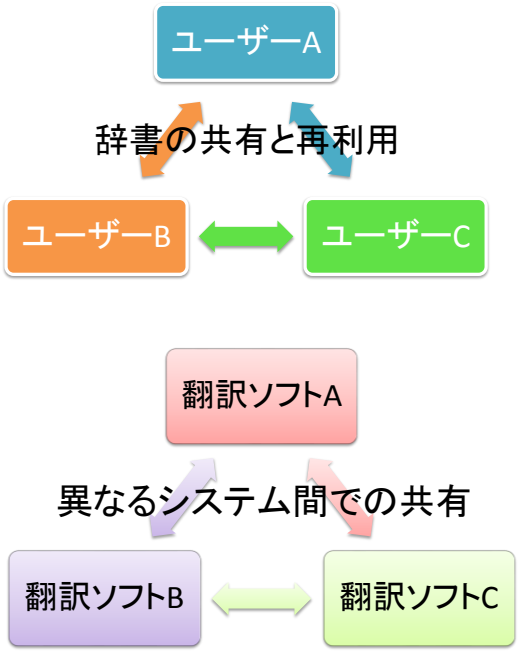
特徴

UTXは、対訳形式の用語の知識を辞書として共有することで、翻訳ソフトの翻訳精度を大きく向上します。ユーザーの視点から、シンプルで「**作りやすく、使いやすい**」ことを目指しています。ユーザーは、同じUTX辞書を、そのまま、あるいは各形式に変換して、さまざまな翻訳ソフトや翻訳支援ツールで使えます。また、UTX辞書を、翻訳ソフト以外で、独立した用語集としても使うこともできます。

なぜ使うのか

UTXを使えば、翻訳ソフトのユーザー辞書を簡単に共有・再利用できます。「翻訳ソフトは変な訳ばかり出す」と思われていませんか？　翻訳ソフトがうまく訳せないのは、ある語や句をどう訳すべきかという**翻訳知識が不足している**からです。UTXは、翻訳知識をユーザー辞書として補うことで、翻訳ソフトの翻訳精度を大きく改善できます。

また、これまで翻訳ソフトのユーザーは、細切れのユーザー辞書を個人で作成しても、有効活用できませんでした。単純なテキスト ファイルでも、形式が揃っていないと共有や再利用は困難です。ウェブサイトでもさまざまな用語集が公開されていますが、実際には各種ソフトウェアですぐに活用できる形式ではなく、手間のかかる修正が必要です。しかし、UTXのような単一の規格を使えば、さまざまなメーカーの翻訳ソフトや用語管理・検索ツールで広く共有し、すばやく再利用できるようになります。



だれが使うのか

翻訳ソフトのユーザーや、翻訳者が使うことを想定しています。作成や使用にあたって、文法、言語学、翻訳ソフトなどの高度な専門知識は不要です。複数形、名詞や動詞など品詞の区別など、最低限の情報のみで作れます。

どのような分野で使うのか

IT、医療、法律、工学など、一定の専門性がある分野の翻訳であれば、どのような分野でも使えます。

どんな語を含むのか

UTX辞書は、製品・部品名、病名、薬品名、法律名など、**特定分野の専門用語**や、人名、地名、施設名などの**固有名詞**のみを含みます。多くの場合は名詞、特に複合名詞がほとんどです。たとえば「XML declaration」のような語は、辞書に登録することではじめて「XML宣言」などと正しく訳せます。「window=窓」のような基本的な語彙は、翻訳ソフトのシステム辞書に含まれているため含めません。翻訳ソフトの購入時に付属しない、きめ細かい対訳の情報を集約して共有・再利用することにより、翻訳精度を

向上できます。

文（センテンス）については、一種の「単語」として扱うのが適切な場合にのみ含めることができます。ただし、原則としてUTXは、翻訳メモリー（文単位の対訳データベース）とは区別して使います。

多言語に対応しているか

UTXの文字コードは、Unicode（UTF-8、BOMなし）で、UTF-8で扱える言語はすべて扱うことができます。基本的には、1つのUTX辞書に、起点言語A（原文）から目標言語B（訳文）への方向の訳が含まれます。各項目には、「暫定」「禁止」「承認」「非標準」（provisional、forbidden、approved、non-standard）のいずれかのステータスを指定できます。「承認」ステータスを持つ語は逆方向の翻訳でも使用できます。

どうやって作るのか

UTX辞書は、**テキスト エディターや表計算ソフトで簡単に作成･編集･表示**できます。さまざまな形式とUTX形式との相互変換を行うツールも開発中です。

どうやって使うのか

UTX辞書は、簡単に変換して各種ツールにインポートできます。翻訳メモリー ツールOmegaT、用語検索ソフトApSIC Xbenchなどのツールでは、ほぼそのまま使えます。

オープンソース開発者・翻訳者の皆様へ―UTX形式で用語を公開・共有しませんか？

対訳用語集を、UTX形式にして、相互に公開・共有することで、ソフトウェアを、すばやく正確に多言語化し、世界中の人に使ってもらうことができます。

UTXメーリング リスト

どなたでもUTXの議論にご参加頂けます。

「UTXメーリング リスト」でウェブを検索してください。

参考資料

- 大倉清司他(2008)「共有ユーザー辞書仕様UTXの現状と今後の展開」言語処理学会第13回年次大会（東京）
- Francis Bond et al. (2009) “Sharing User Dictionaries Across Multiple Systems with UTX-S” in Second International Workshop on Intercultural Collaboration (IWIC2009), Stanford

実例

（IT分野の辞書サンプル）

#UTX 1.11; en-US/ja-JP; 2010-05-15T10:00:00Z+09:00; copyright: AAMT (2011); license: CC-by 3.0				
#src	tgt	src:pos	term status	src:plural
early adopter	アーリー アダプター	noun	approved	early adopters
fast	高速な	adjective	provisional	
optional	省略可能な	adjective	approved	
optional	オプショナルな	adjective	forbidden	
save	保存する	verb	approved	

#はコメント行を表す。

●1行目（コメント行）：辞書の基本情報。各項目はセミコロンと半角スペースで区切る。

#UTX-S <バージョン番号>; <起点言語>/<目標言語>; <最終更新日付時間>; copyright: <著作権者>; license: <ライセンス>; <追加の情報>（必要に応じて追加）

●2行目（コメント行）：属性の規定。各項目はタブ区切り。

上記の例では

#<原語> <訳語> <原語品詞> <用語ステータス> <原語複数形>

●3行目以降が、実際の項目。各項目はタブ区切り。

UTX辞書（用語集）作りのポイント

●辞書の分野を定義する

●固有名詞以外は大文字で始めない

●原形を記載する(市販辞書の見出しの形式)

●原語、訳語以外の情報はコメント欄に記す

●原語に対応する訳語は、最良のものを１つだけ選ぶ

免責事項

UTX、UTX-Simple、UTX-XMLの仕様（以下、これらを総称して「UTX仕様」という）またはUTX仕様に基づいて作成された辞書（以下、「UTX辞書」という）を利用した場合は、以下の事項に同意したものとみなされます。本事項のいずれかが無効または強制不能とされた場合、そのことはいかなる意味でも他の条項の有効性または強制可能性に影響を与えないものとします。

1. AAMT および AAMT 参加者から UTX 辞書および関連ツールの作成者へ

(1)UTX仕様は公開されており、どなたでもご使用頂けます。ただし、AAMTおよびAAMT参加者はUTX仕様に関する権利を放棄しておらず、どなたであってもUTX仕様を改変して公開することはできません。

(2)UTX辞書の作成に際して、AAMTおよびAAMT参加者はUTX仕様を現状有姿のまま提供するものであり、UTX仕様に関する一切の事柄を保証しません。UTX仕様およびUTX辞書は、UTX辞書の作成者各位の責任においてご使用ください。

(3) AAMTおよびAAMT参加者は、UTX辞書の作成者がUTX仕様やUTX辞書を使用した結果（権利侵害の有無・訳語の正確性・妥当性・品質を含むがこれに限らない）に関して、一切の責任を負いません。

(4)AAMTおよびAAMT参加者は、UTX辞書の作成者が作成した辞書の著作権の正当性について確認をせず、保証もしません。従って、UTX辞書の作成者が当該UTX辞書に関するデータについて適切な著作権を保有していない場合、法律的な問題が発生しても、UTX辞書の作成者の責任となります。

(5)AAMTおよびAAMT参加者は、UTX辞書の作成者に、適切な著作権を行使できる場合に限り、商業使用を含め、UTX辞書の使用者へのUTX辞書のライセンス条件を定めることを認めます。ただし、UTX辞書の作成者は、UTX辞書の基盤となるデータの著作権について、個別にデータの提供元に確認する義務があります。

(6)AAMTおよびAAMT参加者は、UTX辞書に関する各種ツールの作成者に対して、当該ツールの使用結果についてなんらの保証もしません。

2. UTX 辞書の作成者から UTX 辞書の使用者へ

UTX辞書の使用者は、UTX辞書を、UTX辞書の作成者が定めるライセンス条件に応じて使用できます。UTX辞書のライセンスは辞書によってそれぞれ異なりますので、UTX辞書を構成するファイルの先頭部分に含まれるライセンス条件をご確認ください。

3. AAMT および AAMT 参加者から UTX 辞書の使用者へ

AAMTおよびAAMT参加者は、AAMTおよびAAMT参加者は、UTX辞書の使用者がUTX仕様やUTX辞書を使用した結果（権利侵害の有無・訳語の正確性・妥当性・品質を含むがこれに限らない）に関して、一切の責任を負いません。UTX辞書の作成者との間で解決をお願いします。

機械翻訳課題調査委員会

共有化･標準化ワーキンググループ メンバー（順不同）

山本ゆうじ（リーダー）	秋桜舎
伊藤肇	株式会社インターグループ
村田稔樹	沖電気工業株式会社
Francis Bond	南洋理工大学（シンガポール）
島津美和子	東芝ソリューション株式会社
大倉清司	株式会社富士通研究所
加藤マイケル孝仁	ラーニング コンサルタント

http://www.aamt.info/japanese/utx/

問い合わせ先：aamt-info@aamt.info

2011年5月版