

UTX 1.20仕様

目次

0.	概要	4
0.1	背景.....	4
0.2	目的.....	4
0.3	UTXの用途.....	5
0.4	国際規格の参照.....	5
0.5	改版履歴.....	6
1.	用語と定義	8
1.1	UTX用語集の種類.....	8
1.2	用語.....	8
1.3	項目.....	8
1.4	プロパティ名、プロパティ値、プロパティ項目.....	9
1.5	フィールド、フィールド値、およびフィールド項目.....	9
1.6	同義語.....	9
1.7	ユーザー ロール.....	9
1.8	翻訳方向.....	10
1.9	UTX対応アプリケーションとUTX変換ツール.....	10
2.	UTXファイルの構造	12
2.1	最もシンプルなUTXの例.....	13
2.2	文字コード.....	13
2.3	行.....	13
2.4	行コメント.....	13
3.	ヘッダー	14
3.1	言語タグ.....	14
3.2	UTXヘッダーの構造.....	14
-	3.2.1 用語集プロパティ	15
-	3.2.2 必須の用語集プロパティ	15

-	3.2.3	省略可能な用語集プロパティのリスト	15
-	3.2.4	UTX versionプロパティ	16
-	3.2.5	langプロパティ(言語宣言)	16
-	3.2.6	creation dateプロパティ	17
-	3.2.7	last modified dateプロパティ	17
-	3.2.8	glossary IDプロパティ	17
-	3.2.9	domainプロパティ	18
-	3.2.10	creatorプロパティ	18
-	3.2.11	glossary administratorプロパティ	18
-	3.2.12	copyrightプロパティ	18
-	3.2.13	licenseプロパティ	19
-	3.2.14	directionalityプロパティ	19
-	3.2.15	sortableプロパティ	19
-	3.2.16	glossary versionプロパティ	20
3.3		用語集詳細	20
3.4		フィールド定義	20
4.		フィールド定義と本文.....	21
4.1		フィールドのための言語タグ	21
4.2		term (src/tgt)フィールド	21
4.3		フィールドのリスト	22
4.4		posフィールド	22
-	4.4.1	posフィールドとそのフィールド項目	22
-	4.4.2	sentenceと特殊文字	23
4.5		term statusフィールド	24
-	4.5.1	provisional (暫定)	24
-	4.5.2	approved (承認)	24
-	4.5.3	空白用語ステータス	25
-	4.5.4	non-standard (非標準)	25
-	4.5.5	forbidden (禁止)	25
-	4.5.6	rejected (却下)	25

- 4.5.7	obsolete (廃止)	26
4.6	ユーザー定義フィールド	26
5.	高度な概念	27
5.1	一列用語ステータスと言語ごと用語ステータス	27
- 5.1.1	一列用語ステータス	27
- 5.1.2	言語ごと用語ステータス	28
- 5.1.3	MT辞書での用語ステータスの扱われ方	28
5.2	概念IDフィールド	30
5.3	用語集ID (glossary ID) フィールド	31
5.4	言語固有のフィールド	32
6.	多言語用語集	34
6.1	langプロパティ(言語宣言)	34
6.2	term (src/tgt)フィールド(多言語)のための言語タグ	34
7.	付録A : UTXコンテンツ ガイドライン	35
7.1	一般的なガイドライン	35
7.2	機械翻訳用途向けのガイドライン	36
7.3	英語特有のガイドライン	36
7.4	日本語の表記法サンプル	36
8.	付録B : UTX対応アプリケーションで推奨される実装	37

文書情報

著者：AAMT共有化・標準化ワーキンググループ：山本ゆうじ（秋桜舎）、村田 稔樹（沖電気工業会社）、Francis Bond（南洋理工大学）、大倉 清司（株式会社富士通研究所）、加藤マイケル孝仁（ジャパニーズ・グレイツ株式会社）、秋元 圭（合同会社ことばや）、高橋 博之（株式会社クロスランゲージ）、亀谷 展（株式会社サン・フレア）

ウェブサイト：<http://www.aamt.info/japanese/utx/>

状態：正式版

更新日付：2019年3月19日

文書の言語：日本語

使用許諾：CC0 1.0

免責事項：<http://aamt.info/japanese/utx/index.htm#disclaimer>を参照

0. 概要

0.1 背景

UTXは、当初、ルールベース翻訳ソフトのための標準化されたユーザー辞書形式として策定された。その後、UTXは、CAT（コンピューター支援翻訳）や自然言語処理を含む、より多くの分野で使用できる用語集形式として方向修正がされた。

翻訳対象の文書(原文文書)には、しばしば専門用語、人名、および地名が含まれる。この種の情報は、ルールベース翻訳ソフトの基本システム辞書では、不足、欠落しているか、不適切であることが多い。この場合、機械翻訳システムでは満足な翻訳結果は得られない。さらに、ユーザーによって作成された辞書（ユーザー辞書）の構造やファイル形式は、異なるMT（machine translation：機械翻訳）システム間でしばしば互換性がなく、辞書の共有と再利用が困難である。しかし、用語が吟味され、特定の分野で適切であれば、中核的な用語情報（原語および訳語）のみで、MTの適切性と正確さを向上するには十分であることは広く知られている。標準化された用語集形式の需要に応えるため、AAMT（Asia-Pacific Association for Machine Translation：アジア太平洋機械翻訳協会）<www.aamt.info>は、異なるMTシステム間で使用できる共有可能な辞書に関する、一連の仕様を策定した。1995年に、AAMTは、IPAの支援を受けてUPF（Universal PlatForm）と呼ばれる最初の仕様を策定した。

その後、機械翻訳の技術や利用方法のさまざまな変化を反映するために、2006年から新しい仕様の策定を開始した。新しい仕様は、2007年8月に「**UTX (Universal Terminology eXchange)**」と呼ぶことが正式に決定された。2009年に、AAMTは、単純なタブ区切り形式のUTXであるUTX-Simpleを策定した。より複雑な機能は、既存の用語集の仕様であるTBXとTBX-Basicで実現できることが分かったため、AAMTはUTX-Simpleの開発に重点を置くようになった。2011年4月に、UTX-Simpleから「-Simple」を外して、「UTX」と名称を変更した。

0.2 目的

UTXの目的は、非専門家を含む幅広い層のユーザーに、簡単に作成でき、簡単に使える用語集を作成するための一連のルールを提供することである。UTXは、高度な管理性や可逆変換よりも、使いやすさと簡潔さを重視している。UTX用語集は、人間にとっても読みやすい形式であるため、翻訳ソフト用途以外の一般的な用語集として使うこともできる。

翻訳ソフトの個人ユーザーがユーザー辞書を作成するとき、精度向上に役立つ辞書を作成するには多大の労力を要する。また、辞書形式が標準化されていないと、単純なプレーンテキストファイルであっても、共有や再利用が難しい。しかし、複数のユーザーがUTXのような単一の規格に基づくユーザー辞書を作成すれば、辞書を簡単に共有できる。これらの辞書は、さまざまなメーカーの翻訳ソフトなど、多様なツールで使える。

UTXの使用法としては、たとえば日々の翻訳業務で翻訳知識を共有するために、複数の情報源から用語集をすばやく作成する場合がある。また、UTXは広範囲の用途で、用語集を配布、共有、再

利用する際にも役立つ。

UTXはシンプルな構造であるため、複雑な用語管理を単純化できる。UTXは表構造であり、表計算アプリケーションで編集できる。多数の言語が関与する場合、またフィールドの数が多い複雑な用語管理については、TBX形式のほうが適切であることがある。

0.3 UTXの用途

UTX形式は、翻訳ソフトのエンド ユーザーや翻訳者を念頭に設計されている。UTX用語集を作成、編集、および使用する際に、言語学、文法、XML、翻訳ソフトなどの特別な知識は必要ない。UTX用語集は、最小限の用語データがあれば作成できる。

UTX用語集は、どのような分野の翻訳でも使用できるが、たとえばICT、医学、法律、エンジニアリングなど、専門性の高い分野での効果が特に高い。分野は、「ICT分野でのPerl（スクリプト言語）」、「医学分野での心臓外科」などのように、専門性をより高くすることが理想的である。

注：複数の分野が翻訳プロジェクトに関わる場合、各分野で個々の用語集を作成することが推奨される。用語集の組み合わせを変更することによって、効率的にそれらを再利用し、別の目的に活用できる。

UTX用語集は、テキスト エディターや表計算アプリケーションで、直接、表示し、編集できる。また、用語チェックをする目的で、用語ツールからも使用できる。UTX用語集は、その用語集に基づく文書での用語の一貫性をもたらすため、用語の一貫性が求められる。

UTXは、専門性がない一般的・汎用的な内容の翻訳には適していない。UTXのフレームワークは、使用するMTシステムが充実したシステム辞書を備えていることを想定している。非専門用語は、特定分野内で特定の意味と訳語を持つときのみ、UTX用語集に含めるべきである。

0.4 国際規格の参照

この文書は、以下の国際規格を参照する。

言語タグ：IETF BCP 47, *Tags for Identifying Languages*

日付/時刻形式：ISO 8601:2004, *Data elements and interchange formats – Information interchange – Representation of dates and times*

ファイル符号化：ISO/IEC 10646:2014, *Information technology -- Universal Coded Character Set (UCS)*

用語集形式：ISO 30042:2008, *Systems to manage terminology, knowledge and content -- TermBase eXchange (TBX)*

翻訳メモリー形式：TMX 1.4b *Specification*

要件レベル：S. Bradner, *Key words for use in RFCs to Indicate Requirement Levels*, <http://www.ietf.org/rfc/rfc2119.txt> IETF (Internet Engineering Task Force) RFC 2119, March 1997.

0.5 改版履歴

UTX-Simple 1.00 (2009年11月10日)

- 初版。

UTX-Simple 1.10 (2010年11月22日)

- 全体的な改版。用語ステータス、辞書管理者、双方向、概念ID、辞書ID、および望ましいMT機能の項目を追加した。

UTX 1.11 (2011年5月25日)

- UTX-Simpleの名称を、UTXに変更した（「-Simple」を付けない）。
- 一部のトピックの説明を追加。

UTX 1.20 (2016年8月3日)

- 多言語用語集を作成できる。
- 語法：以後、「辞書」は「用語集」と呼ぶ。
- `dictionary ID`は`glossary ID`に変更され、テキスト文字列が使用できるようになった。
- 用語集にサブ用語集を含めることが可能になった。
- BOMをUTXファイルに追加する必要がある。
- 言語ごと用語ステータスが追加された。
- 用語ステータス：空白は「承認」と同一とみなされるようになった。
- 用語ステータス：却下／廃止用語ステータスを追加した。
- ヘッダー：`date created`が`creation date`に変更された。またシンプルな書式YYYY-MM-DDが明示的に使用できるようになった。
- ヘッダー：UTXバージョン以外のすべて用語集のプロパティーが、省略可能になった。
- 用語集プロパティーの追加：`last modified date`、`glossary administrator`、`domain`、`copyright`、`directionality`、`sortable`、`glossary version`。
- `bidirectional`プロパティーが削除され、`directionality`プロパティーがその機能を引き継いだ。
- 品詞フィールド項目の追加：`vt`、`vi`、`prenominal`。
- フィールド（列）の順序が自由になった。
- 概念IDが数値に限定されなくなった。

UTX 1.20 (2016年9月20日)

- 表現面での修正。

UTX 1.20 (2017年5月24日)

- 執筆メンバー変更。

UTX 1.20 (2018年2月16日)

- 使用許諾Creative Commons 4.0 BYと指定。

UTX 1.20 (2019年3月19日)

- 使用許諾CC0 1.0と指定。

1. 用語と定義

このセクションでは、UTXに関する重要な用語と定義を説明する。

1.1 UTX用語集の種類

UTX用語集は、特定の専門分野内での用語の集合である。人間とコンピューターの両方にとって可読性を併せ持つように設計されている。

UTX用語集は、翻訳ソフト（機械翻訳）の正確さと流暢さを向上するためのデータ辞書として使用できる。いくつかのMTシステムでは、2種類の辞書を使用する。**ユーザー辞書**とはユーザーによって作成された辞書であり、**システム辞書**とは、MTシステム内に組み込まれた辞書である。

用語集に含まれる言語数という観点からは、単一言語、二言語、または多言語のUTX用語集が作成できる。

単一言語用語集は、1つの言語のみの用語を含む。単一言語用語集は、用語チェック、あるいは、表現や技術用語の異表記の統一などに使用できる。（日本語の読みがななどの）発音情報を含む単一言語用語集は、音声認識および音声合成アプリケーションに使用できる。（日本語など）いくつかの言語では、そのような発音情報を言語入力システムに使用できる。

二言語用語集は、2つの言語の用語を含む。UTX用語集は、多くの場合、二言語である。別記ない限り、この仕様書では、UTX用語集が二言語であると仮定する。

多言語用語集は、3つ以上の言語の用語を含む。

注：UTX 1.11以前は、二言語および単一言語用語集のみを認めていた。UTX 1.20以後では、多言語用語集を作成できるようになった（「6. 多言語用語集」参照）。

1.2 用語

用語とは、原文言語あるいは訳文言語の見出し語である。UTX用語集内の用語は、辞書の見出し語のような単語の基本形（レンマ）である必要がある。「7. 付録A：UTXコンテンツ ガイドライン」も参照。

注：用語定義は、UTX用語集では省略可能である。

1.3 項目

項目とは、UTX用語集内の、1つまたは複数の用語および追加情報から構成される単位である。

1項目は1行に対応する。詳細は、「4 フィールド定義と本文」を参照。

注：UTX用語集がルールベース機械翻訳で使用される場合、「一語一義」の原則に従う必要がある。すなわち、1つの用語は、その辞書の特定分野においてはただ一つの意味（概念）しか持たないということである。

1.4 プロパティ名、プロパティ値、プロパティ項目

プロパティとは、対象物の特性である。別記なき場合、UTXでのプロパティは用語集プロパティを指す。

プロパティ値とは、あるプロパティに対して設定された値である。たとえば、ユーザーは、`glossary administrator`プロパティのプロパティ値を決定できる。

プロパティ項目とは、プロパティに対して、あらかじめ定義されたピックリスト項目である。たとえば、`uni`、`bi`、および`multi`は、`directionality`プロパティ値のプロパティ項目である。

1.5 フィールド、フィールド値、およびフィールド項目

フィールドとは、項目または用語の特性（たとえば品詞など）である。表形式では、1フィールドは1列に対応する。

フィールド値とは、あるフィールドに対して設定された値である。たとえば、ユーザーは、`glossary ID`フィールドのフィールド値を決定できる。

フィールド項目とは、フィールド値に対して、あらかじめ定義されたピックリスト項目である。たとえば、`noun`は、`pos`（品詞）フィールド値のフィールド項目である。

1.6 同義語

同義語とは、同じ概念（または意味）を持つ用語である。各々の同義語は独立した項目となり、さらに概念IDでグループ化される。同じ語の別表記は、同義語と同じ方法で扱われる。用語ステータスは、同義語や異表記の異なるステータスを示すために使用できる。詳細は、「5.2 概念IDフィールド」を参照。

1.7 ユーザー ロール

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。

1.8 翻訳方向

ある言語から別の言語への翻訳の方向（「翻訳方向」）は、一方向、双方向、多方向のいずれかである。この情報は、「3.2.14 directionalityプロパティ」で指定できる。

一方向用語集は、翻訳方向が主として原文言語から訳文言語への一方向となる用語集である。

例：一方向二言語日英UTX用語集

原文言語：日本語、訳文言語：英語

主な翻訳方向：日本語から英語

注：臨時の手段として、一方向用語集内の用語をエクスポートして逆方向で使用できる。この操作は、**逆方向エクスポート**と呼ばれる。この場合、原文言語が訳文言語になり、訳文言語が原文言語になる。完全な双方向用語集と比較すると、逆方向にした結果が徹底的に検討されていないため、逆方向エクスポートされた一方向用語集には問題が含まれることがある。

双方向用語集とは、二方向の翻訳で使用できるように作られた用語集である。一方の言語の用語はもう一方の言語に翻訳でき、その逆も可能である。

例：双方向二言語の日英UTX用語集

言語1：日本語、言語2：英語

翻訳方向：日本語↔英語

例：双方向多言語の英語・フランス語・ドイツ語用語集

言語1：英語、言語2：フランス語、言語3：ドイツ語

翻訳方向：英語↔フランス語、英語↔ドイツ語（ただしフランス語↔ドイツ語は不可）

多方向用語集とは、任意の翻訳方向の翻訳で使用できるように作られた多言語用語集である。

例：多方向多言語の英語・日本語・中国語用語集

言語1：英語、言語2：日本語、言語3：中国語

翻訳方向：上記の言語の任意の組み合わせ

1.9 UTX対応アプリケーションとUTX変換ツール

UTX対応アプリケーションとは、用語集をUTX形式で読み込み、または書き出しできるアプリケーションである。そのようなアプリケーションの例として、翻訳ソフト、用語集ツール、および

UTXエディターがある。

UTX対応アプリケーションのうち、**UTX変換ツール**とは、**UTX用語集**をその他のファイル形式に変換し、またその他のファイル形式を**UTX用語集**に変換するツールの総称である。

2. UTXファイルの構造

UTXファイルのファイル拡張子は".utx"である。

UTXファイルは以下の構造を持つ。

ヘッダー

ヘッダーは、以下の要素から構成される。

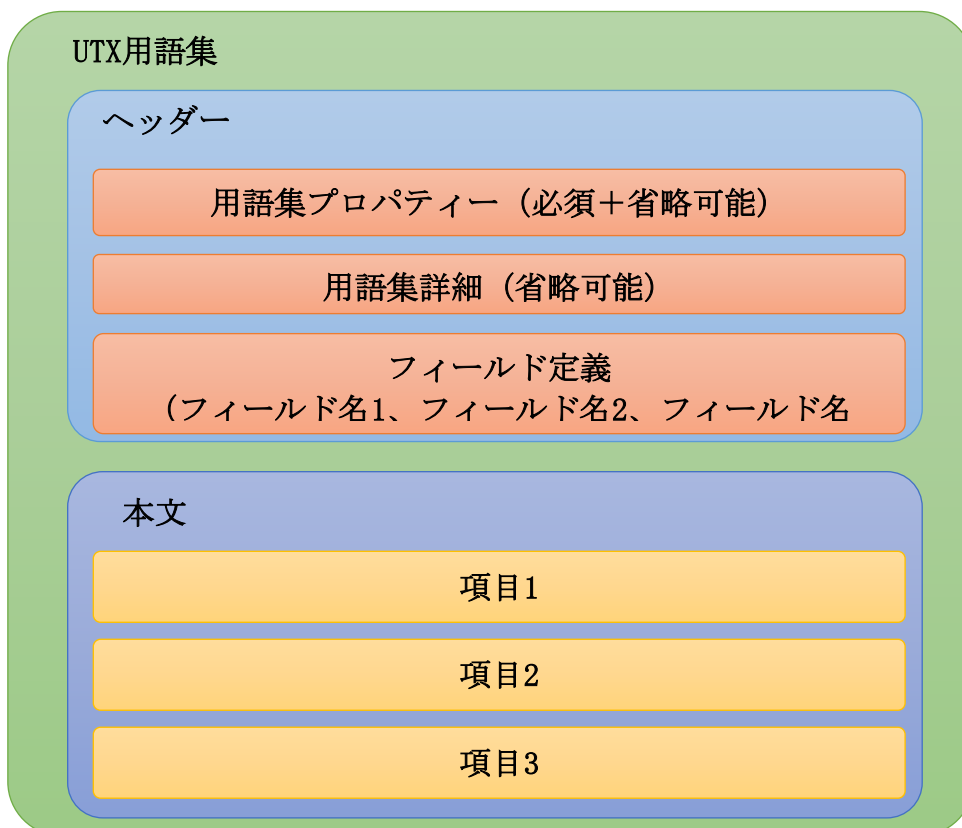
1. 用語集プロパティ。「3.2.2 必須の用語集プロパティ」と「3.2.3 省略可能な用語集プロパティのリスト」を参照。
2. 用語集詳細（省略可能）。「3.3 用語集詳細」を参照。
3. フィールド定義。「3.4 フィールド定義」を参照。

本文

項目から構成される本文（タブ区切りテキスト）。「4. フィールド定義と本文」を参照。

以下の図は、UTX用語集の構造を示す。構成要素は後節で詳述する。

図1：UTX用語集の構造



2.1 最もシンプルなUTXの例

以下は、最小の必要情報を備えた完全なUTX用語集の例である。各要素は、以下の章で後述する。

#UTX 1.20	
#term:en	term:ja
test	テスト

2.2 文字コード

UTXファイルの文字コードは、UTF-8 (BOM付き) である。

注：「BOMなし」であったUTX 1.11から変更されていることに注意。BOMは、Microsoft WindowsなどのOSでUTF-8エンコードを識別するために使用される特殊文字である。

2.3 行

UTXファイルで使用される改行コードは、"`\r\n`" (CR+LF) である。改行コードは、行の終端を表す特殊文字である。空行は認められない。

注：Microsoft Windows環境では、CR+LFは多くのテキスト エディターの既定の設定であり、通常は特別な注意は不要である。UNIXまたはMac OSなどのLFがよく使用される環境では、UTXファイルの作成者は、改行コードをCR+LFに変更する必要がある。

2.4 行コメント

(シャープ記号) で始まるすべての行は、行コメントとして扱われる。既存の項目も、行頭に"#"を入れることによりコメントアウトできる。たとえば、用語集を変換する際の技術的な問題を切り分ける目的で、特定の項目をコメントアウトできる。

注：行コメントを受け付けない形式にエクスポートする場合、UTX変換ツールあるいはUTX対応アプリケーションは行コメントを除外する必要がある。

行コメントを含むUTX用語集では、意図しない並べ替えを防ぐために、必要に応じてヘッダーで`sortable`プロパティを`false`に設定すべきである（「3.2.15 `sortable`プロパティ」を参照）。

3. ヘッダー

3.1 言語タグ

UTX用語集の内部で、言語を示す場合には、IETF BCP 47言語タグを使用する。これは、HTMLとXMLで使用される言語タグと同じである。<<http://www.w3.org/International/articles/language-tags/Overview.en.php>>を参照。言語タグは、言語を示すサブタグと、区別が必要な場合は、それに続く地域または文字種サブタグから構成される。地域差が重要でない場合、地域サブタグは省略する必要がある。

言語タグは、言語宣言での使用に加えて、あるフィールドが特定の言語に属することを示すために使用される。「4.3 フィールドのリスト」を参照。

例：言語タグ

言語	言語タグ	地域サブタグが付いた言語タグ	文字種サブタグが付いた言語タグ
中国語（簡体字、中華人民共和国）	zh	zh-CN	zh-Hans
中国語（繁体字、台湾）	zh	zh-TW	zh-Hant
英語（アメリカ）	en	en-US	
英語（イギリス）	en	en-GB	
日本語	ja	ja-JP	
韓国語	ko	ko-KR	

3.2 UTXヘッダーの構造

UTXヘッダーは、用語集全体に関する情報を示す部分である。UTXヘッダー内のすべての行は、"**#**"から始まる。これは、**UTXヘッダー**内のすべての行がコメントアウト行として扱われることを示す。

UTXヘッダーは、用語集プロパティの**1**行、さらにフィールド定義の**1**行の、少なくとも**2**行の必須の行を含む。用語集プロパティの行は、**2**行以上記述できる。

用語集の詳細を説明する目的で、**2**行の必須の行の間に、**1**行または複数行の行を追加できる。

例：UTXヘッダー

```
#<用語集プロパティ> 必須+省略可能
#<追加の用語集プロパティ> 省略可能
(#<用語集詳細> 省略可能)
```

(#<追加の用語集詳細> 省略可能)

.....

#<フィールド定義> 必須

3.2.1 用語集プロパティ

用語集プロパティは、用語集に含まれる言語、作成日、使用許諾などの用語集の諸特性を記述する。

用語集プロパティは、プロパティ名、コロン、半角スペース、およびプロパティ値から成る。

構文	例
<プロパティ名>: <プロパティ値>	copyright: AAMT (2016)

個々の用語集プロパティは、セミコロンと、それに続く1つの空白によって区切られる。

例：二言語UTXファイルのヘッダー

```
#UTX 1.20; lang: en/ja; creation date: 2016-04-15; copyright: AAMT (2016)
```

プロパティ値が未決定だが早期の決定が望ましい場合は、プロパティ項目"undetermined"を指定できる。

例

```
license: undetermined
```

3.2.2 必須の用語集プロパティ

名前	構文	例
UTX version	#UTX <version>	#UTX 1.20

3.2.3 省略可能な用語集プロパティのリスト

名前	構文	例
lang (言語宣言)	lang: <言語1>/<言語2>あるいは <src>:<原文言語>/<tgt>:<訳文言語>など	lang: en/jaあるいは lang: src:en/tgt:jaあるいは lang: src:en/tgt:ja/tgt:fr
	creation date: YYYY-MM-DD	creation date: 2016-04-10

creation date	creation date: YYYY-MM-DDThh:mm:ssTZD	creation date: 2016-04-10T12:34:56Z
last modified date	last modified date: YYYY-MM-DD	last modified date: 2016-05-10
	last modified date: YYYY-MM-DDThh:mm:ssTZD	last modified date: 2016-05-10T12:34:56Z
glossary ID	glossary ID: <文字列>	glossary ID: Brain surgery
domain	domain: <文字列>	domain: Aerospace
creator	creator: <文字列>	creator: Yamada Tarou
glossary administrator	glossary administrator: <文字列>	glossary administrator: Yamada Hanako
copyright	copyright: <文字列> (<年>)	copyright: AAMT (2016)
license	license: <文字列>	license: CC BY 4.0
directionality	directionality: <uni/bi/multi>	directionality: bi
sortable	sortable: <true/false>	sortable: true
glossary version	glossary version: <数値>	glossary version: 1.00

3.2.4 UTX versionプロパティ

UTX version (UTXバージョン) プロパティは、用語集のUTXバージョンを示す。"UTX"、半角スペース、"1.20"で示す。UTX versionは、唯一の必須の用語集プロパティである。

3.2.5 langプロパティ(言語宣言)

langプロパティは、言語宣言とも呼ばれ、用語集に含まれる用語の言語を示す。

二言語用語集の言語宣言は以下の形式で指定する。

構文	例
lang: <言語1>/<言語2>	lang: en/ja

各言語の役割は、src (原文言語)、およびtgt (訳文言語) を使用して明確にすることができる。

構文	例
lang: <src>:<原文言語>/<tgt>:<訳文言語>	lang: src:en/tgt:ja

注：言語宣言は省略可能である。用語集に含まれる言語は、term (src/tgt)フィールドでも示される（「4.2 term (src/tgt)フィールド」を参照）。

3.2.6 creation dateプロパティ

creation date (作成日時) プロパティは、用語集が最初に作成された日付 (および時刻) を示す。ISO 8601形式を使用する。

現地時間は、UTC (協定世界時) に時差を追記することにより示すことができる。

例 : creation dateプロパティ

日本標準時 (UTCプラス9時間) の2016年4月10日14:28は以下のように表される。

種類	構文	例
現地日付のみ (時刻なし)	YYYY-MM-DD	creation date: 2016-04-10
日付と時刻	YYYY-MM-DDThh:mm:ssTZD	creation date: 2016-04-10T05:28:00Z あるいは creation date: 2016-04-10T14:28:00+09:00

記号の意味は以下のとおりである。

YYYY=4桁の年

MM=2桁の月 (01=1月など)

DD=2桁の日 (01~31)

hh=2桁の時間 (00~23) (am/pmは使用不可)

mm=2桁の分 (00~59)

ss=2桁の秒 (00~59)

TZD=時間帯指定子 (UTCを使用する場合は「Z」、現地時間を使用する場合は+hh:mmあるいは-hh:mm)

<<http://www.w3.org/TR/NOTE-datetime>>も参照。

3.2.7 last modified dateプロパティ

last modified date (最終変更日時) プロパティは、用語集が最後に修正された日付 (および時刻) を示す。ISO 8601形式を使用する。現地時間は、UTCに時差を追記することにより示すことができる。

3.2.8 glossary IDプロパティ

glossary ID (用語集ID) プロパティは、用語集に対して一意の識別子として使用されるテキスト文字列である。

注 : glossary IDは、一般的に用語集の分野を表す、用語集の「名前」となりうる。

glossary IDプロパティは、用語集が単一分野からの用語を含んでいる場合に限り、使用される。用語集が複数の分野からの用語を含んでいる場合は、**glossary ID**プロパティの代わりに、各分野に**glossary ID**「フィールド」を割り当てる必要がある。「5.3 用語集ID (glossary ID) フィールド」を参照。

例

glossary ID: Rocket engine

3.2.9 domainプロパティ

domain (分野) プロパティは、用語集の分野を示すテキスト文字列である。複数の用語集をある分野にグループ化する必要がある場合、このプロパティを使用する。分野名として**glossary ID**を使用する場合、**domain**プロパティは不要である。

例

domain: Aerospace

3.2.10 creatorプロパティ

creator (作成者) プロパティは、用語集を作成した人の名前を示す。これは用語集を作成した個人である。個人名が望ましいが、部門やグループの名前を使用することもできる。

例

creator: Yamada Tarou

3.2.11 glossary administratorプロパティ

glossary administrator (用語集管理者) プロパティは、用語の承認の責任を負う個人の名前を示す。責任を明確にするために個人名とすることが望ましい。用語集に複数の用語提出者がいる場合、用語集管理者がだれか決定することが推奨される。用語集作成者が唯一の用語提出者である場合、用語集管理者は不要である。

例

glossary administrator: Yamada Hanako

3.2.12 copyrightプロパティ

copyright (著作権) プロパティは、用語集のコンテンツ全体に対して著作権を持つ個人や組織の名前、および著作権が効果を現した年を示す。このプロパティは省略可能であるが、著作権を明確にする目的で推奨される。

例

copyright: AAMT (2013)

3.2.13 licenseプロパティ

license (使用許諾) プロパティは、用語集の使用許諾、つまりユーザーがこの用語集をどのように使用できるかを示す。使用許諾は、クリエイティブ コモンズ<<https://creativecommons.org>>、パブリック ドメイン、あるいはその他の形式で指定できる。このプロパティは省略可能であるが、使用許諾を明確にする目的で推奨される。

以下の例で、使用許諾は、クリエイティブ コモンズ4.0「表示」である。

例

license: CC BY 4.0

3.2.14 directionalityプロパティ

directionality (翻訳方向) プロパティは、用語集の翻訳方向を示す（「1.8 翻訳方向」を参照）。プロパティ項目**uni**は、用語集の項目が1つの翻訳方向で使用されるよう意図されていることを示す。**bi**は、用語ステータスで許可されている範囲で、用語集内の項目が、双方向の翻訳で使用できることを示す。**multi**は、用語ステータスで許可されている範囲で、用語集内のすべての項目が、どのような方向の翻訳でも使用できることを示す。プロパティ項目**multi**は、多言語用語集でのみ指定できる。

用語集の種類	directionalityプロパティのプロパティ項目
一言語用語集	N/A
二言語用語集	uniまたはbi
多言語用語集	uni、bi、またはmulti

注：UTX 1.11では、翻訳方向性を示す目的で**bidirectional**フラグが使用された。これは、UTX 1.20で多方向の翻訳方向性が導入されたことに伴い、使用されなくなった。

3.2.15 sortableプロパティ

sortable (並べ替え可能) プロパティは、1つまたは複数のフィールドをキーとして使用して、用語集の項目を並べ替えできることを示す。プロパティ値はブール値 (**true/false**) である。項目の順序を維持するには、このプロパティ値を**false**に指定する。

例

sortable: false

注：`sortable`プロパティを`false`と指定することによって、用語集の最後に新しい項目を追加している場合、ユーザーが項目の追加順を簡易的に追跡記録できる。また、`"sortable: false"`と指定することで、ある項目がコメントアウトされた場合でも、用語集内でのその項目の位置を維持できる。

3.2.16 `glossary version`プロパティ

`glossary version`（用語集バージョン）プロパティは、用語集のバージョンを示す。これにより、用語集の改版を追跡記録できる。UTX形式のバージョンと混同しないように注意すること。

例

```
glossary version: 1.02
```

3.3 用語集詳細

用語集詳細は、用語集の詳細情報、追加の法的通知、および免責条項などの情報を示す、1行または複数行のテキスト行である。用語集詳細の先頭には「#」を置く。

例

- # 用語集の詳細情報をここに追加できる。
- # たとえば、「`readme`」のような用語集全体に関する情報である。
- # 免責条項も記載できる。

.....

3.4 フィールド定義

UTXヘッダーの最後の行は、「#」で始まり、一連のフィールド定義（あるいは列定義）を含む。フィールド定義と本文は密接に関連するため、詳細は「4. フィールド定義と本文」で説明する。

4. フィールド定義と本文

UTX用語集の本文は、1行1項目の集合から構成される。行内では、各要素（表計算での「セル」）はタブ文字によって区切られる。

4.1 フィールドのための言語タグ

言語固有のフィールドは、言語タグによって示される。言語タグのコロンの後にはスペースを入れない。

例：フィールドのための言語タグ

言語固有のフィールド	例
英語の用語（原文言語）	<code>src:en</code>
日本語の用語（訳文言語）	<code>tgt:ja</code>
日本語用語ステータス	<code>term status:ja</code>
中国語（簡体字）用語（訳文言語）	<code>tgt:zh-CN</code>
中国語（簡体字）用語ステータス	<code>term status:zh-CN</code>
日本語の用語に関するコメント（日本語の項目のためのユーザー定義フィールド）	<code>x-comment:ja</code>

注：「`x-comment`」のようなユーザー定義フィールドが項目全体に関わる場合は、言語タグを必要としない。しかし、特定の言語の用語に関係するユーザー定義フィールドは、言語固有のフィールドを必要とすることがある。たとえば、日本語の訳語のみに関するコメントがある場合、言語タグなしの`x-comment`フィールドではなく`x-comment:ja`フィールドを設けるべきである。

フィールド値の記述に使用される言語は、言語タグによって示された言語であってもなくてもよい。たとえば、`x-comment:ja`フィールドのコメントを英語で書くこともできる。

4.2 term (src/tgt)フィールド

用語のフィールドを示すために、`term`、`src`、あるいは`tgt`フィールドを使用する。これらのフィールドの言語は、言語タグによって示される。これらのフィールドは、用語集の言語宣言と一致する必要がある。

`term`フィールドは、一言語用語集や、原文言語と訳文言語の役割を区別しなくてよい場合に使用する。

`src`フィールドは、1つ以上の原文言語を指定する場合に使用する。

tgtフィールドは、1つ以上の訳文言語を指定する場合に使用する。

例については、以下の表を参照。言語タグの前のコロンの後に半角スペースは入れない。

例：用語集の種類と翻訳方向の表記法

多言語用語集の例については、「6. 多言語用語集」を参照。

用語集の種類	翻訳方向表記法の例	詳細
一言語	<code>term:en</code>	英語がこの用語集で使われる唯一の言語である。
二言語	<code>src:en tgt:ja</code>	原文言語が英語で、訳文言語が日本語である。
二言語	<code>term:en term:ja</code>	原文言語／訳文言語の区別が不要の場合。

4.3 フィールドのリスト

フィールド名	構文/フィールド項目/フィールド値	例
term (src/tgt)	term:<言語> src:<言語> tgt:<言語>など	<code>src:en tgt:ja</code>
pos	noun/properNoun/verb/vt/vi/prenominal/adjective/ adverb/sentence	noun
concept ID	<数値>	45
term status	blank/provisional/approved/ non-standard/forbidden/rejected/obsolete	approved
glossary ID	<文字列>	Brain surgery

4.4 posフィールド

4.4.1 posフィールドとそのフィールド項目

pos (part of speech) フィールドは品詞を示す。このフィールドが言語タグを伴わない場合、その品詞は項目内のすべての言語の用語に適用されるものとみなされる。さらに、`pos:<言語タグ>`の形式で、特定の言語の品詞を示すこともできる（コロンの後にスペースを入れない）。posフィールドは省略可能である（UTX 1.11からの変更点）。posフィールド値は空白のままでもよい。

以下のposフィールド項目が定義されている。

posフィールド項目	詳細
noun	名詞

properNoun	固有名詞
verb	動詞
vt	他動詞
vi	自動詞
adjective	形容詞
prenominal	連体修飾語
adverb	副詞
sentence	文

nounとproperNounは、それぞれ名詞と固有名詞を指定する。固有名詞には、人名や地名などがある。

verbは動詞を指定する。vtは他動詞を指定する。viは自動詞を指定する。vtとviは、その区別が必要な場合に指定できる。vt、vi、およびverbは混在できる。

例：verb、vt、およびvi

term:en	term:ja	pos
receive	受信する	verb
acquire	取得する	vt
listen	リッスン状態になる	vi

adjectiveは形容詞を指定する。prenominalは、連体修飾語（prenominal modifier）を指定する。

注：連体修飾語は、形容詞に似ているが扱いが異なる。たとえば、日本語の連体詞（「円錐形の」など）は、活用しない連体修飾語である。いくつかのMTシステムでは、連体修飾語と形容詞を区別する必要がある。

adverbは副詞を指定する。副詞句も含まれる。

4.4.2 sentenceと特殊文字

sentenceは、「用語」が文であることを示す特殊なposフィールド項目である。

注：sentenceは必要な場合のみに使う。sentenceは、たとえば、文として現れるユーザー インターフェイスのメッセージに使用する。対訳文のペアの項目は、原則として、用語集にではなく、TMXなど翻訳メモリ用のデータに記述すべきである。文を品詞の一種として扱わないMTシステム用にUTX用語集がエクスポートされる場合、sentence項目は名詞として扱うことができる。

sentence項目では、いくつかの特殊文字を使用できる。これらはsentence項目でのみ使用される。

文字	表記法
タブ文字	\t
改行文字	\n
リテラル エスケープ文字 (バックスラッシュ (円記号) そのものを表す場合)	\\

4.5 term statusフィールド

term status (用語ステータス) は、用語の状態を示すフィールドである。空白、provisional、approved、non-standard、forbidden、rejected、あるいはobsoleteの7つの用語ステータスがある。用語集管理者およびその代理者のみが、用語ステータスの値を変更できる。

例：用語ステータス フィールド

#src:ja	tgt:en	term status:ja	term status:en
プラグイン	plug-in	approved	approved
プラグイン	plugin		non-standard
アドオン	add-on	provisional	

注：用語集に用語ステータス フィールドがない場合、すべての項目は承認済み (approved) とみなされる。

4.5.1 provisional (暫定)

provisional (暫定) 用語ステータスは、用語提出者によりその項目が提案されたが、用語集管理者によってまだ承認されていないことを示す。暫定ステータスは一時的なものであり、用語管理者は、「承認」などの用語ステータスをすみやかに決定すべきである。

注：用語集管理者は、その用語を用語集から除外 (削除) するか、別の用語集に移動することもできる。

4.5.2 approved (承認)

approved (承認) 用語ステータスは、その項目が特定用語集 (分野) に属することが用語集管理者により承認済みであることを示す。承認ステータスは、文脈が適切な場合は、その用語 (訳語) を最優先で使用する必要があることを示す。用語に同義語や異表記がある場合 (たとえば"plug-in"と"plugin")、これらのうちの一つのみを承認ステータスとする必要がある。

ある言語の承認語は、別の言語の承認語とペアになる。複数項目であっても品詞が異なる場合は、異なる用語とみなされる。たとえば、"plot"という語には名詞と動詞があるため、それぞれに承認ステータスを付けられる。

4.5.3 空白用語ステータス

用語ステータスが空白の場合、承認とみなされる（UTX 1.11からの変更点）。非標準、禁止、却下、および廃止用語（後述）とペアになる用語の用語ステータスが、空白となっている場合、暗黙の承認ステータスを意味する。「4.5 term status フィールド」の冒頭の例を参照。

4.5.4 non-standard（非標準）

non-standard（非標準）用語ステータスは、一群の同義語や異表記の中で、優先度が低い用語を示す。

注：用語集管理者が、ある用語が特定の用語集で適切か否か決定する。したがって、この基準は、別の用語集管理者が管理する用語集では異なることがある。

4.5.5 forbidden（禁止）

forbidden（禁止）用語ステータスは、その用語（訳語）が使用されてはならないことを示す。用語が禁止されるのは、訳として不適切というだけでなく、最終結果の文書の文脈でその語が使われることが不適切とみなされる場合である。

禁止語は、非標準語と異なり、訳語候補として示されてはならない。

注：用語は、言語的観点、社会的観点、用語管理の観点、企業のブランド イメージやその他の観点から明示的に禁止されることがある。

UTX 1.11までは、訳語のみを禁止と示すことができた。UTX 1.20では、（原語を含め）任意の用語を禁止できる。

禁止語は、用語チェックのためにUTX用語集からエクスポートできる。この情報に基づいて、翻訳ツールの機能や用語チェック ツールで、不適切な用語が含まれていないか、翻訳ファイルをチェックできる。

4.5.6 rejected（却下）

rejected（却下）用語ステータスは、その用語を特定の用語集に含めるべきではないことを示す。却下語は、記録のために用語集に維持するか、別のリストへ移動するか、後で削除することができる。

4.5.7 obsolete (廃止)

obsolete (廃止) 用語ステータスは、その用語 (訳語) が、以前は使用されていたが、現在は使用してはならないことを示す。廃止語は、記録のために用語集に維持するか、別のリストへ移動するか、後で削除することができる。

4.6 ユーザ一定義フィールド

UTX用語集には、任意の数のユーザ一定義フィールドおよびそれらのフィールド項目を追加できる。

言語固有のフィールドについては、その言語を示すために言語タグを使用する。言語タグのコロンの後にはスペースを入れない。

構文	例
ユーザ一定義フィールド:<言語>	x-termUsage:en

5. 高度な概念

5.1 一列用語ステータスと言語ごと用語ステータス

用語ステータスには、一列用語ステータスと言語ごと用語ステータスの2種類がある。用語集では、これらのいずれかを使用して用語ステータスを示せる。

5.1.1 一列用語ステータス

用語ステータスは1つのフィールドで指定できる。この方法で示された用語ステータスは一列用語ステータスと呼ばれる。UTX 1.11まで、用語ステータスは一列用語ステータスしかなかった。

項目内の一列用語ステータスは、用語ステータスの種類によって、暗黙的に原語、訳語、あるいはその両方の情報を示す。各用語ステータスの対象語（フィールドの対象である語）を、以下の例で示す。

例：英日用語集の一列用語ステータス：

src:en	tgt:ja	term status	(用語ステータスの指示語)
outlet	コンセント	approved	原語と訳語の両方
outlet	アウトレット	forbidden	訳語
plugin	プラグイン	provisional	原語と訳語の両方
power point	コンセント	non-standard	原語

一列用語ステータスは、原語と訳語のそれぞれのステータスではなく、原語と訳語間の関係によって決定される。

一列用語ステータスでは、以下の規則が適用される。

- 「承認 (approved)」は、常に原語および訳語のペアに対して適用される。
- 「非標準 (non-standard)」は、原語の同義語や異表記の中で、標準的に使用されるべき用語以外の用語を示すために適用される。(したがって、非標準語は、逆方向エクスポートされたときに、訳語としては使用されてはならない。)
- 「禁止 (forbidden)」は、訳語のみに適用される。
- 「暫定 (provisional)」は、常に原語および訳語のペアに対して適用される。

利点：用語ステータスをシンプルな一つのフィールドで表せる。用語集の翻訳方向が一方向であり、逆方向には使用しない場合に適している。

欠点：用語ステータスの指示対象が混乱を招くことがある。さらに、一列用語ステータスでは、逆翻訳方向で用語がどのように使用されるべきか十分に示せない。

5.1.2 言語ごと用語ステータス

言語ごと用語ステータスは、2言語のペアについてではなく、各言語ごとの用語についての用語ステータスを表す。たとえば、二言語一方向用語集で使用する場合、原文言語用と訳文言語用にそれぞれ、計2つの用語ステータス列を必要とする。

構文	例
term status:<language tag>	term status:ja

言語タグが指定されない場合、用語ステータスは一列用語ステータス（UTX 1.11スタイル）として扱われる。

注：言語ごと用語ステータスは、二言語双方向用語集と多言語用語集を扱えるようにするために、UTX 1.20で導入された。

5.1.3 MT辞書での用語ステータスの扱われ方

UTX用語集からMT用の辞書をエクスポートする場合、言語ごと用語ステータスに基づいていくつかの規則が適用される。UTX用語集は、一言語、二言語、多言語用語集となることができ、さらに翻訳方向は、一方向、双方向、多方向となりうる。だが、MT辞書は、通常、二言語一方向のみである。以下の例は、1つの二言語双方向UTXが、日英・英日の2つのMT辞書にどのようにエクスポートされるかを示す。

以下は双方向UTX用語集である。この用語集では、原語としての「操作」は承認語であるため、「アクション」より優先度が高い。

例1：非標準語を含むUTX用語集

#src:ja	tgt:en	term status:ja	term status:en
操作	action	approved	approved
アクション	action	non-standard	

エクスポートされた日英MT辞書では、原語が「操作」と「アクション」のどちらであっても「action」という訳語に翻訳しなければならないので、両方とも含まれる。以下の例を参照。

例2：例1からエクスポートされたMT辞書（日英）

日本語	英語	優先度
操作	action	N/A
アクション	action	N/A

この場合、訳語は「action」しかないなので、優先度はない。

この用語集から英日MT辞書が逆方向エクスポートされる場合は、先の例とは異なる。元のUTX用語集で「操作」が承認語だったため、用語ペア「action／操作」は優先度がより高い。また元のUTX用語集で「アクション」が非標準語だったため、用語ペア「action／アクション」の優先度は低くなる。MTシステムは、この優先度を反映することが望ましい。MTシステムが優先度を持つ用語を識別できない場合、「action／アクション」はエクスポートされるべきではない。

例3：例1からエクスポートされたMT辞書（英日）

英語	日本語	優先度
action	操作	高
action	アクション	低

次の例は、禁止語を含むUTX用語集である。

例4：禁止語を含むUTX用語集

#src:en	tgt:ja	term status:en	term status:ja
configuration	構成	approved	approved
configuration	コンフィグレーション		forbidden

上記の用語集から英日MT辞書をエクスポートする場合を考える。これは以下のようになる。

例5：例4からエクスポートされたMT辞書（英日）

英語	日本語	優先度
configuration	構成	N/A

「コンフィグレーション」は、オリジナルのUTX用語集で禁止されているため、エクスポートされた辞書に含まれていない。

次に、上記の用語集から日英MT辞書を逆方向エクスポートする場合を考える。これは以下のようになる。

例6：例4からエクスポートされたMT辞書（日英）

日本語	英語	優先度
構成	configuration	N/A
コンフィグレーション	configuration	N/A

非標準ステータスの例と同様に、抽出されたMT辞書では、原語が「構成」と「コンフィグレーション」のどちらであっても「configuration」という訳語に翻訳しなければならないので、両方が抽出される。この場合、訳語は「configuration」しかないなので、優先度はない。

用語集には、ある語のスペルミスが（スペルミスであっても）翻訳できるように、原語として意図的に登録されていることがある。この場合は、逆方向エクスポートされたときにこの語が訳語として使用されないように、非標準ではなく、禁止ステータスを使用する必要がある。

例7：意図的なスペルミスを含むUTX用語集

#src:en	tgt:ja	term status:en	term status:ja
configuration	構成	approved	approved
configulation	構成	forbidden	

却下あるいは廃止用語ステータスは、禁止と同じ方法で扱われる。エクスポートされたMT辞書に暫定語を含めるか除外するかの判断は、用語集管理者の方針による。

5.2 概念IDフィールド

複数の同義語および異表記が同じグループに属することを示す目的で、「概念ID」を割り当てることができる。そのようなグループは「概念グループ」と呼ばれる。1つの原語に対し、1つの訳語しかない場合、この用語の概念IDは不要である。

概念IDには、数値あるいはテキスト文字列を使用できる。複数の用語集を統合するとき、同一の概念IDを持つ項目はその用語集IDで区別できる（「5.3 用語集ID (glossary ID) フィールド」を参照）。

UTX 1.11までは、複数原語と単一訳語（多対一関係）からなる項目グループがある場合、概念IDが必要だった。このシナリオに加えて、UTX 1.20では、単一原語と複数訳語（一对多関係）、また、複数原語と複数訳語（多対多関係）が可能である。

以下の例は、概念ID、概念グループ、および用語ステータスの関係について示している。

表 1：例（英語から日本語）

項番	src:en	tgt:ja	term status	concept ID
1	outlet	コンセント	approved	1
2	outlet	アウトレット	forbidden	1
3	power point	コンセント	non-standard	1

4	PowerPoint	PowerPoint	approved	
5	plugin	プラグイン	approved	2
6	plug-in	プラグイン	non-standard	2
7	outlet store	アウトレット ストア	approved	
8	AAMT	AAMT	approved	3
9	Asia-Pacific Association for Machine Translation	アジア太平洋機械翻訳協会	approved	3

(ここでの項番は、説明の目的でのみ使用している。)

項番1、2、3の用語は、同じ概念を示している。5と6、8と9もそれぞれ同じ概念である。したがって、それらは、(赤線で囲まれた)概念グループにそれぞれ属する。概念グループは、概念IDによって識別できる。概念IDが3である項目を含む概念グループは、「概念グループ3」と呼ぶことができる。

概念IDは、たとえば、頭字語と完全な綴りを1つの概念グループとしてまとめる場合に使用される。表1では、用語8 "AAMT"は頭字語であり、用語9 "Asia-Pacific Association for Machine Translation"は、完全な綴りの形である。これらは同じことを意味しており、したがって、両方に同じ概念ID、3が付けられる。同じ概念IDを持つ項目は、通常、少なくとも1つの原語あるいは訳語を共有する。しかし、概念が共有されていても、"AAMT"と"Asia-Pacific Association for Machine Translation"のように、2項目に共通の用語がないこともありうる。

概念グループ内で、ある用語が、別の言語では複数の用語に対応する場合、これらの複数用語のうち1つだけに承認語ステータスが付けられる。つまり、概念グループ1および2の中には、それぞれ承認語が1つのみある。

概念グループ内の複数項目に共通の原語および訳語がない場合、これら複数項目内の用語のそれぞれに承認ステータスを付けることができる。たとえば、概念グループ3は2つの承認語を含んでいる。

複数の用語の品詞が異なれば、共通の原語／訳語があっても、異なる概念IDを持つ。

5.3 用語集ID (glossary ID) フィールド

用語集IDフィールド (UTX 1.11以前は「辞書ID」) は、用語集の一意の識別子である。テキスト文字列により記述される。用語集IDフィールドは、複数の用語集を統合するとき、それぞれの用語集からの用語を区別するために必要となる。

UTX 1.20では、ある用語集内に複数の「サブ用語集」を含めることができる (UTX 1.11では、

「統合」する際に、用語集内の分野の区別は失われていた)。したがって、単一のUTX用語集内に、それぞれ異なる分野を持つ複数のサブ用語集を含めることができる。用語集IDは、用語集内のサブ用語集の名前とみなされる。

2つの用語集を統合する場合、たとえば、異なる用語集からの2つの項目が同じ概念IDを共有していることがある（「5.2 概念IDフィールド」を参照）。この場合、以下のように、異なる概念が同じ概念IDに偶然にグループ化されることがある。

元用語集	項目	concept ID
用語集A	outlet	76
用語集B	instantiate	76

この場合、用語集IDを使用すれば、同じ概念IDを持つ項目を識別できる。

元用語集	項目	concept ID	glossary ID
用語集A	outlet	76	Electronics
用語集B	instantiate	76	IT

5.4 言語固有のフィールド

動詞の活用形など、特定言語に固有の属性フィールドは、以下のように記述される。

<フィールド名>:<言語タグ> (コロンの後にスペースを入れない)

以下のフィールドがあらかじめ定義されている。

英語用の定義済みフィールド名

フィールド	詳細	例
plural	複数形	instances
3sp	三人称単数形	transmit
past	過去形	transmitted
presp	現在分詞形	transmitting
pastp	過去分詞形	transmitted
comparative	比較級	more opaque
superlative	最上級	most opaque

例：英語の動詞活用のための定義済みフィールド

#UTX 1.20; lang: en/ja; creation date: 2016-04-15T10:00:00+09:00; copyright: AAMT (2016); license: CC BY 4.0					
#src:ja	tgt:en	pos:en	plural:en	past:en	superlative:en
アーリー アダプター	early adopter	noun	early adopters		
手段	means	noun	-		
白濁した	opaque	adjective			most opaque
保存する	keep	verb		kept	

フィールド値は、混乱を避けるために完全形を常に記述すべきである。たとえば、複数形を示すためには、「s」のみではなく「oranges」と書く。このフィールド値は省略可能であり、すべての項目に網羅的に付ける必要はない。項目は空白のままでもよい。

ハイフン"-"は、あるフィールド項目が適用外であることを明示的に示すために使用する。たとえば、英語の単語"information"には複数形がないため、これを明示的に示す場合はplural:enに"-"を使用する。

6. 多言語用語集

UTX 1.20以降のUTX用語集では多言語用語集をサポートする。

多言語UTXファイル用のヘッダーの例

```
#UTX 1.20; lang: en-US/ja/fr-FR/ko; creation date: 2016-04-15; copyright: AAMT (2016);  
license: CC-by 4.0; glossary ID: Rocket engine
```

6.1 langプロパティ(言語宣言)

ヘッダー内の言語宣言では、用語集内の言語と、それらの言語の原文言語や訳文言語としての役割（必要な場合）を指定するために言語タグを使用する。

二言語および多言語用語集では、原文言語と訳文言語を、**src**と**tgt**で指定できる。**src**と**tgt**が指定されない場合、言語は原文や訳文の区別なしに等しく扱われる。

注：これは、UTX 1.11からの変更点である。UTX 1.11では、指定された最初の言語が原文言語であり、第2の言語が訳文言語である。

例：言語宣言

用語集の種類	言語	例
一言語用語集	(単一言語)	lang: en
二言語用語集1	<原文言語>/<訳文言語>	lang: src:en/tgt:ja
二言語用語集2	<言語1>/<言語2>	lang: en/ja
多言語用語集1	<原文言語>/<訳文言語1>/<訳文言語2>.....	lang: src:en-US/tgt:ja/tgt:zh-CN
多言語用語集2	<言語1>/<言語2>/<言語3>.....	lang: en-US/ja/zh-CN

6.2 term (src/tgt)フィールド(多言語)のための言語タグ

多言語用語集内のterm (src/tgt)フィールドのための言語タグは、言語宣言と同様の方法で指定する。

翻訳方向表記法の例	詳細
<code>src:en tgt:ja tgt:fr</code>	1つの原文言語と2つの訳文言語。
<code>src:en src:fr tgt:ja</code>	2つの原文言語と1つの訳文言語。

7. 付録A : UTXコンテンツ ガイドライン

このガイドラインでは、効果的なUTX用語集を作成するためのベスト プラクティスと推奨事項を示す。

7.1 一般的なガイドライン

1. 一般的に、UTX用語集は特定分野の専門用語のみを含む必要がある。UTX用語集には、一般的な意味しかない一般的な語を含めるべきでない。ある分野で特別の意味を持つ場合には、一般的な語を含めることもある。

適切な例 : XML declaration / XML宣言

適切な例 : window / ウィンドウ

不適切な例 : window / 窓

2. 単語は基本形で記述する（市販辞書の見出しの形式のように、名詞なら単数形、動詞なら基本形）。

適切な例 : define, flow

不適切な例 : defined, flows

3. 英数字は、全角ではなく半角文字を使用する。
4. 項目内には、三点リーダー（...）、「～」や他の文字によって示される変数（文脈によって入れ替えられる語）を含めない。

不適切な例 : "prefer ... to ..."

5. コメント専用でないフィールドには、コメントを含めない。コメントを追加するには、コメント用フィールドを定義するか、"#で始まるコメント行に記述する。
6. フィールドあるいはフィールド項目に接頭辞「x-」を追加すると、それらがユーザー定義であることを明確にできる。このような接頭辞により、ユーザー定義のフィールドまたはフィールド項目を、UTX仕様に今後追加される可能性があるフィールドまたはフィールド項目と区別できる。

例（ユーザー定義フィールド） : x-serialNumber

例（posフィールド内のユーザー定義フィールド項目） : x-preposition

7.2 機械翻訳用途向けのガイドライン

1. 用語集の分野を明確に定義する。

例1 : domain: Medical electronics

例2 : domain: Legal

2. 各項目では、1つの原語（承認語）に対して、1つの最適な訳語（承認語）を選択して追加する。人間翻訳者であれば、複数の代替の非標準語の中から選ぶことができる。しかし、多くのMTシステムでは1つの最適な訳語だけを使用できる。残りの非標準語はまったく使用されないか、手動操作によってのみ参照されうる。
3. 特定のMTシステムにのみ関係する用語は含めない。単にMTシステムの中で特定の問題あるいは制限を処理するためのみに、用語が追加されることがある。このような情報は、他のシステムでは必要ない可能性がある。このような用語が必要である場合は、これらを管理するために個別の用語集を作成する。

7.3 英語特有のガイドライン

1. 固有名詞を除き、1文字目は常に小文字とする。
2. 冠詞（a、an、the）は、それが固有名詞の一部でない場合は記載しない。

7.4 日本語の表記法サンプル

用語の表記法は、用語集内で一貫している必要がある。以下は表記法の例である。

1. 日本語として登録される英数字はすべて半角にする。
2. 半角カタカナや機種依存文字は使用しない。
3. サ変動詞は「する」で終わる。例：強調する
4. 形容動詞はadjectiveとして示す。
5. 形容動詞は「な」で終わる。例：静かな
6. 音引きは省略しない。例：ユーザー、セキュリティー、コミュニティー
7. カタカナ複合語などの語の区切りを示す中黒や半角スペースは省略しない。
8. すでに定着しているものを除き、意味があいまいなカタカナ語は避ける。

8. 付録B：UTX対応アプリケーションで推奨される実装

UTXを最大限に利用するため、MTソフトウェアおよび用語ツールなどのUTX対応のアプリケーションやUTX変換ツールには、以下の機能を実装することを推奨する。これらは必要条件ではなく、提案である。

(一般ユーザー向け) MTシステムでは、以下の機能の実装を推奨する。

1. 非専門用語用の高品質なシステム辞書。
2. 複合語の承認語を最優先として使用／インポート／エクスポートできること。
3. 用語の優先度のバランスを保つ機構を持っている。特に、優先度の高い短い用語は、システム辞書内の慣用表現より優先されるべきではない。たとえば、インポートされた項目「figure/図形」が承認語（すなわち優先）であっても、システム辞書の中にすでにある「figure of speech/比喩」という慣用表現より優先されるべきではない。
4. 代替の選択肢として（承認語の優先度より下げて）非標準語を表示する。
5. 名詞の単数形から複数形を推測できる。
6. 基本的な動詞の活用を推測できる。
7. 可能なかぎり、不足している情報を補足できる。
8. 同時に複数のユーザー辞書を使用できる。
9. 禁止語を使用しないようにすることができる。

UTX対応アプリケーションは、禁止語の使用を抑制する機能を持つことが望ましい。そのような機能がない場合、禁止語が使用されないよう事前に除外する必要がある。用語が、ある用語集では禁止されず、別の用語集で禁止されていることがある。UTX対応アプリケーションは、そのような競合を検出する仕組みを持つことが望ましい。

UTX変換ツールは、UTX用語集を変換するとき、（原文または訳文言語の）暫定語を含む項目を除外できるようにすべきである。

専門的翻訳のための翻訳支援アプリケーションは、高度な用語機能を持つことがある。これらのシステムでは、上記に加えて以下の実装を推奨する。

1. 複数の辞書セットを切り替え、管理できる仕組み。
2. 禁止語を検出する（可能であれば置換する）仕組み。