

# 引く辞書から飛び込んでくる辞書へ

## ー「用語集超活用ソフト HT<sup>3</sup>(イフイー-スリー)」対応の辞書づくりー

How to edit dictionaries for HT<sup>3</sup> (Hitachi Terminology Transformation Tool)

(株)日立国際ビジネス

開発グループ

富井 玲子

Reiko Tomii

HT<sup>3</sup>は、既存の用語集を100%活用して、技術・学術文書などの用語の確認や訳語の付与をするという「用語と訳語の統一用ソフト」である。ユーザーは、HT<sup>3</sup>を使う前に、HT<sup>3</sup>対応のユーザー辞書を準備する必要がある。HT<sup>3</sup>用ユーザー辞書は、辞書形式のExcel専用シートからデータベース形式のものに変換するため、これまでの電子辞書では不可能だった、既存の用語集データの流し込み、一覧表示、内容の変更などが、すべて可能となる。ここでは、HT<sup>3</sup>用ユーザー辞書を作成するための、データの入力とデータ入力後のチェック、辞書の登録とヒット(一致)の仕方を具体例をあげて紹介する。さらにHT<sup>3</sup>では、日本語/英語の見出し語だけでなく、同義語など見出し語と表現の異なった用語(異表記)も登録できる。この異表記の登録によって、より高いヒット率と実用性が格段に向上するのである。ここでは、「異表記」の考え方もあわせて紹介する。

### 1. はじめに

「用語集を完全活用して、翻訳をすすめたい」、「用語集が正確に使われたかをチェックしたい」、このような現場のニーズから生まれたのが、「用語集超活用ソフト HT<sup>3</sup>(イフイー-スリー)」である。HT<sup>3</sup>は、機械翻訳ソフトではない。文書中に存在する、用語集にある用語に対して、用語の確認や訳語の付与をするという、画期的なツールソフトである。文書中にある用語を用語集でいちいち引くのではなく、用語集の方から用語が文書中に飛びこんでくるのである。HT<sup>3</sup>によって、既存の用語集を100%活用し、翻訳をしたり、用語を統一したりすることが可能となる。

HT<sup>3</sup>の大きな特徴の一つが、日英/英日の見出し語の対訳語に加え、異表記を記入できるExcelの電子辞書形式の採用である。この辞書形式によって、これまでの電子辞書では不可能だった既存の

用語集データの流し込み、作り込んだユーザー辞書の一覧表示、内容の変更、グループ内共有の用語集の使用および優先順位をつけた複数の辞書の統合使用がすべて可能となった。原則としてHT<sup>3</sup>を使う場合は、ユーザー自身が辞書を作ることが必要だが、(財)日本規格協会(JSA)発行の「JIS工業用語大辞典 CD-ROM版 with HT<sup>3</sup>(Light)」が市販されている。

ここでは、HT<sup>3</sup>用ユーザー辞書(以降、「HT<sup>3</sup>辞書」と呼ぶ)の入力の仕方、データ入力後のチェック、辞書の登録とヒット(一致)の仕方、さらに異表記の考え方を紹介する。

### 2. 「HT<sup>3</sup>辞書」入力の実際

実際に「HT<sup>3</sup>辞書」を作成する方法を手順を追って説明しよう。

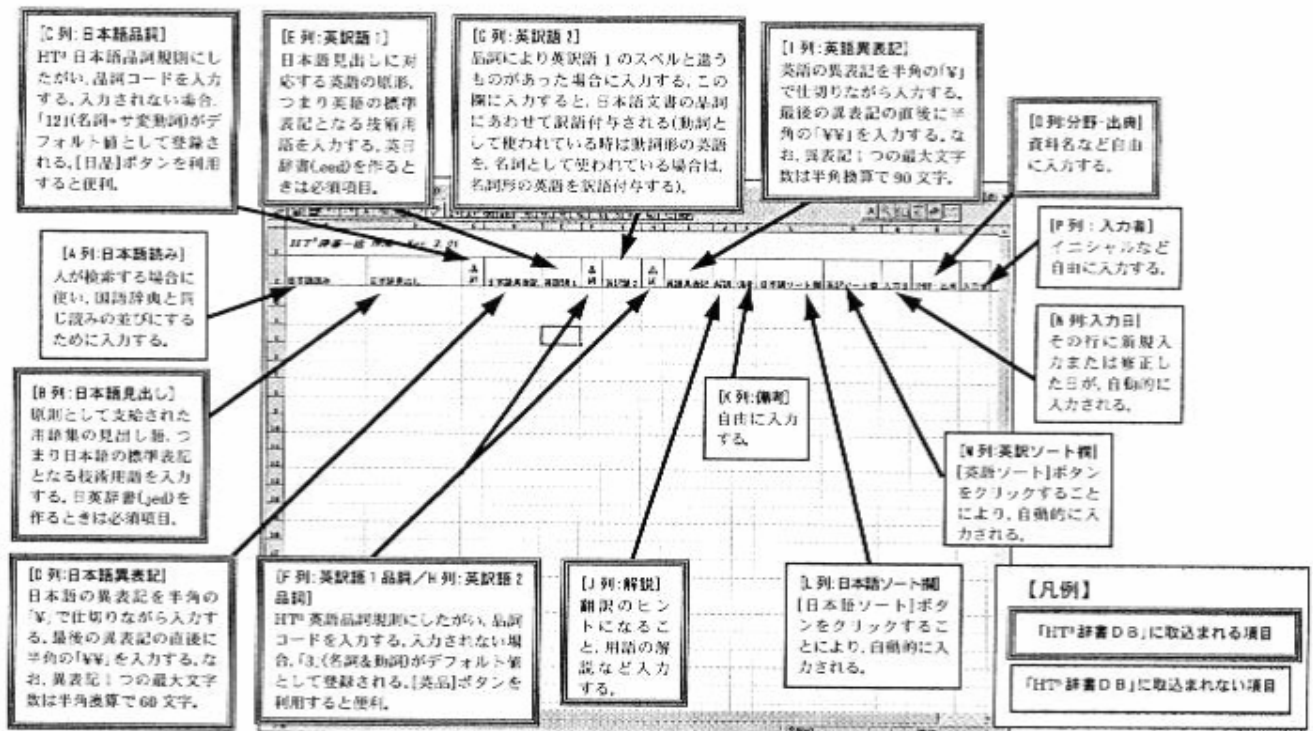


図1 「Excel 一括作成シート」画面の説明

「HT<sup>3</sup>辞書」は、Excel上で読み込んだHT<sup>3</sup>形式の用語集(以降、「Excel 一括作成シート」と呼ぶ)にデータを入力し、データベース形式のHT<sup>3</sup>用語辞書(以降、「HT<sup>3</sup>辞書DB」と呼ぶ)へ変換して作成する。実際の用語統一・訳語統一のための用語検索は、変換後の「HT<sup>3</sup>辞書DB」とWord文書との間で行なっている。複数の「HT<sup>3</sup>辞書DB」に優先順位をつけ、統合して使用すること:「辞書統合」も可能である。この場合、「A辞書」と「B辞書」両方で、同じ用語が登録されていた場合、優先される辞書の用語の方が統合後の「HT<sup>3</sup>辞書DB」に登録される。

#### ①「Excel 一括作成シート」への入力

HT<sup>3</sup>アイコングループの「辞書新規作成」アイコンをダブルクリックすると、HT<sup>3</sup>専用の「Excel 一括作成シート」が開く。入力要領にしたがって、この「Excel 一括作成シート」に入力していく(入力要領については、図1「Excel 一括作成シート」画面の説明参照)。

作成した「HT<sup>3</sup>辞書」をどのように使用するかに応じて、入力する欄が決まってくる。必ずしもすべての欄に入力する必要はない(日本語用語確認ま

たは英語用語確認のいずれかだけで使用する場合には、表1「Excel 一括作成シート」入力制限参照)。

この「Excel 一括作成シート」によって、既存の用語集データの流し込みや、「HT<sup>3</sup>辞書」の一覧表示、変更が可能となる。

#### ②「Excel 一括作成シート」の保存

「Excel 一括作成シート」へのデータの入力が終わったら、ファイル名をつけて、必ず保存する。これは今後の追加、訂正等を行なう原本となる。

#### ③「HT<sup>3</sup>辞書DB」への変換

「Excel 一括作成シート」を保存したら、専用ツールバーの[辞書変換]ボタンをクリックする。[辞書変換]ボタンをクリックすることによって、「Excel 一括作成シート」から「HT<sup>3</sup>辞書DB」へ変換が始まる。

「Excel 一括作成シート」の保存を確認するメッセージが表示されたら、「Excel 一括作成シート」を保存してあるときは[はい(Y)]を、保存していないときは[いいえ(N)]をクリックする。[いいえ(N)]をクリックした場合は、「Excel 一括作成シート」を保

表1 「Excel一括作成シート」入力制限

Excel該当列	項目名	最大文字数(半角換算)	文字種	日本語用語確認だけで使用する場合(英訳に使用しない)	英語用語確認だけで使用する場合(和訳に使用しない)
A列	日本語読み	255	ひら・全半角カタ・全半角英数	○	-
B列	日本語見出し	60	ひら・全半角カタ・全半角英数	◎	-
C列	日本語品詞	2	半角数字	○	-
D列	日本語異表記	255	ひら・全半角カタ・全半角英数	◎	-
E列	英訳語1	90	半角英数	-	◎
F列	英訳語1品詞	1	半角数字	-	○
G列	英訳語2	90	半角英数	-	○
H列	英訳語2品詞	1	半角数字	-	○
I列	英語異表記	255	半角英数	-	◎
J列	解説	255	ひら・全半角カタ・全半角英数	○	○
K列	備考	255	ひら・全半角カタ・全半角英数	-	-
L列	日本語ソート欄	255	ひら	〔日本語ソート〕ボタンによって自動入力	-
M列	英訳ソート欄	255	半角英数	-	〔英語ソート〕ボタンによって自動入力
N列	入力日	6	半角英数	自動入力	自動入力
O列	分野・出典	50	ひら・全半角カタ・全半角英数	-	-
P列	入力者	255	ひら・全半角カタ・全半角英数	-	-

◎：入力必須項目 ○：急ぎの場合省略可 空欄：入力自由 -：入力不要

存してから、再度[辞書変換]ボタンをクリックする。ひとつの「Excel一括作成シート」から、日英辞書(ファイル拡張子: jed)と英日辞書(同: eed)の2種類の「HT<sup>3</sup>辞書DB」へ同時に変換できる。画面のメッセージにしたがって、変換作業を進める。

まず、「HT<sup>3</sup>辞書DB」のファイル名を確認する。必要により日英辞書(jed)、英日辞書(eed)のファイル名を変更する。

新しいファイル名のときは「新規にファイルを作成しますか?」とメッセージが出るので[はい(Y)]をクリックする。既にある「HT<sup>3</sup>辞書DB」に追加するときはこのメッセージは表示されず、追加されるのは見出し語が異なるものだけになる。注意しなければならないのは、見出し語が同じで標準表記や訳語の異なるものが修正されない点である。「Excel一括作成シート」で既にある辞書の部分修正、追加、削除を行ったときは、一旦「HT<sup>3</sup>辞書DB」ファイル(jed または eed)を削除してから、新規に辞書変換をした方が良好だろう。

画面の[実行]ボタンをクリックし変換処理件数を確認する。既存の「HT<sup>3</sup>辞書DB」ファイルの場合は登録済レコード件数が表示されている。[開始]ボタンをクリックしメッセージボックスの[はい(Y)]をクリックすると変換処理が開始される。

変換処理終了時に表示される[OK]ボタンをクリックすると、「HT<sup>3</sup>辞書DB」への変換が完了する。

### 3. 「異表記」とは何か? ～異表記の考え方

標準表記は、辞書の見出し語として登録されている用語をさす。異表記は標準表記の同義語や別の表現をさす。異表記には、業界スラングや自家製用語の異表記もあれば、長音記号の有無、拗音の大小、送りがなの有無のような「言葉遊び」的な異表記もある。

たとえば「フロッピーディスク」を表現する場合、他に「ディスケット」、「フロッピー」、「FD」などいろいろな表現がある。どの語を使用しても「フロ

#### 異表記の例

##### 【例1 日本語】

標準表記: 「置換え」の場合  
異表記: 「置き換え、置換、置き替え、入れ替え、差替え、リブレース、リブレイス」など

##### 【例2 英語】

標準表記: 「floppy disk」の場合  
異表記: 「diskette、floppy-disk、FD、F/D、archive、disk」など

ッピーディスク」に統一したいとき「フロッピーディスク」が標準表記、その他の表現(「ディスケット」、「フロッピー」、「FD」など)が異表記となる。HT<sup>3</sup>では、一つの標準表記につい

て最大 255 半角文字分(Excel7.0 の1セルへの入力限界) まで、複数個の異表記を登録できる。どれ

表2 異表記の作り方

日本語	英語
1. 同義語、慣用語	1. 同義語、慣用語
2. 表記違い(省略形、漢字/かな表記、和語/カナ表記、送り仮名等)	2. 表記違い(省略形、英米語の綴り違い、記号/スペルアウト等)
3. 英訳語を参考にしたもの	3. スペース、ハイフン等
4. 格助詞入り、中点、スペース等	4. 前置詞入り/倒置

を標準表記とするかはユーザー自身で決定する。

このような異表記を「HT<sup>3</sup> 辞書 DB」に登録することによって、ヒット率を格段に向上させることが可能になる。

例えば、これまでの電子辞書では、

日本語見出し	英訳語
相手 CPU	remote CPU
フロッピーディスク	floppy disk

のように、日本語/英語の見出し語のみの登録であるため、「相手の CPU からフロッピーの内容を…」という文章には全くヒットしない。

しかし、「HT<sup>3</sup> 辞書 DB」に、

日本語見出し	日本語異表記	英訳語	英語異表記
相手 CPU	相手の CPU リモート CPU 相手の中央処理装置	remote CPU	remote central processing unit
フロッピーディスク	ディスケット フロッピー FD F/D アーカイブデスク フレキシブルデスク アーカイブディスク ット	floppy disk	diskette floppy-disk FD F/D archive disk flexible disk archive diskette

のように、日本語/英語の異表記を豊富に登録することにより、同じ文章でも

「相手の CPU から フロッピー の内容を…」とヒットするようになり、実用性が増加する。

HT<sup>3</sup> では「用語確認」の場合、色と下線の違いによって、標準表記/異表記、全半角や活用による違いなどが一目で判別できる。

これまでの電子辞書は一つの日本語用語に対して一つの英訳語だったが、チェックしたい文章の言葉に相当のばらつきがあっても、「HT<sup>3</sup> 辞書 DB」に異表記を丁寧に作り込むことで、より高いヒット率と実用性が実現できるのである。

#### 4. 辞書入力後のチェック

「Excel 一括作成シート」が、「HT<sup>3</sup> 辞書」データに正確に変換されるためには、「Excel 一括作成シート」のデータが所定の要領にしたがって正確に入力されていないと、あるべき情報(データ)が、正しい場所(セル)に、あるべき状態(形式)で入力されていることがポイントである。そこで、辞書変換を行なう前に、「Excel 一括作成シート」の正確性をチェックすることが、あとで良い結果をもたらすことになる。

チェックには、

- ① HT<sup>3</sup> 専用ツールボタンを使ったチェック(未入力セル、重複語)
- ② Excel の機能を使ったチェック(スペルチェック、オートフィルタ、文字列検索など)

がある。

参考までに主なチェックポイントを表3に示す。

#### 5. 辞書登録とヒットの仕方

最後に、作り込んだ「HT<sup>3</sup> 辞書」がどのように活かされるのかを確認するために、ヒットの仕方について述べる。

HT<sup>3</sup> では、日本語文で1文字ごと、英文で1単語ごとの前方最長一致方式を採用している。つまり、文章の前の方から最も長く一致した用語がヒットする。辞書の登録の仕方によっては、意図しないヒットになる場合もあるため、注意が必要であるがほとんどのものは、これで解決できる。1文字の用語や一般的な単語も勿論登録できるが、意図しないヒット(例えば「行-line」を登録すると、「発行する」の「行」で意味のないヒットをしてしまう)を生む可能性が高いので、お勧めしない。あくまでも、技術・学術用語の登録にとどめておいた方が良からう。

表3 チェックポイント一覧

項目名	主なチェック内容
日本語読み	入力文字種の統一(かな・カナ・記), 不要なスペース
日本語見出し	未入力セル, 重複語, 一語のバイト数, 不要なスペース, 使用禁止記号(?, *, の使用), ( ) 付き用語
日本語表表記	重複語, 一語のバイト数, 不要なスペース, 使用禁止記号(?, *, の使用), ¥/¥
英訳語1	未入力セル, 重複語, 一語のバイト数, 不要なスペース, 使用禁止記号(?, *, の使用), ( ) 付き用語, ミススペル, 対訳英語がない用語
英訳語2	重複語, 一語のバイト数, 不要なスペース, 使用禁止記号(?, *, の使用), ミススペル
英語表表記	重複語, 一語のバイト数, 不要なスペース, 使用禁止記号(?, *, の使用), ミススペル, ¥/¥, 英米つづり
品詞	該当数字以外の使用, 入力もれ

### 前方最長一致によるヒットの仕方の実際

なお, ヒットの仕方は, 実線(付きイタリック文字)部分がヒットした用語, 破線部分が無視される用語である。

#### (例) リモート CPU 制御コマンド

辞書に登録した用語

ヒットの仕方

①リモート	→ <u>リモート CPU 制御 コマンド</u> ①      ②      ③      ④
②CPU	
③制御	
④コマンド	

\* 全部ヒット

①リモート CPU 制御	→ <u>リモート CPU 制御</u> コマンド ①      ②      ③
②CPU 制御	
③制御コマンド	

\* ①のみヒット, ②と③は無視される。

#### (例) remote CPU control command

辞書に登録した用語

ヒットの仕方

①remote CPU	→ <u>remote CPU control command</u> ①      ②      ③
②control	
③command	

\* 全部ヒット

①remote CPU	→ <u>remote CPU control command</u> ①      ②      ③
②CPU control	
③control command	

\* ①と③はヒット, ②は無視される。

## 6. 終わりに

「HT<sup>3</sup> 辞書 DB」の特性をつかんで, 「HT<sup>3</sup> 辞書」を構築すれば, 既存の用語集を確実に活かして, 翻訳やチェックの作業をすることができる。

「HT<sup>3</sup> 辞書」の作成は, 根気と時間がかかるが, かけた分だけ, 実際の作業に用語集を活かせる。

既存の用語集をうまく使いこなせないとお嘆きになる前に, HT<sup>3</sup> を利用して, 用語集を存分に活用する醍醐味を味わって頂きたい。

## 参考文献

貝島 良太 1996 技術用語統一ソフト HT<sup>3</sup>(用語統一・訳語付与ソフトの開発) テクニカルコミュニケーションシンポジウム'96 論文集, 123-127

貝島 良太 1997 用語と訳語の統一用ソフト HT<sup>3</sup>の開発 テクニカルコミュニケーションシンポジウム'97 論文集, 153-157

澤田 位 1997 JIS 工業用語大辞典 CD-ROM 版の開発について テクニカルコミュニケーションシンポジウム'97 論文集, 146-149