日英機械翻訳用辞書データ記録仕様書

平成２７年４月

特　　許　　庁

目　次

1. 媒体とファイル記録方式

（１）媒体はＣＤ－Ｒ（ライトワンスＣＤ）である。

（２）ＣＤ-Ｒの物理フォーマット、論理フォーマットはＣＤ-ＲＯＭ公報仕様（意匠、商標、公開・国際商標、審決：第４版）準拠である。

（３）日英機械翻訳用辞書のデータは、日本語見出しと訳語（英語）、及びその属性情報等を記録するテキストファイルで構成される。

（４）テキストファイルで使用する文字コードはシフトＪＩＳである。

2. 各ファイルの概要

2.1. テキストファイル

著作権ファイルは独自フォーマットのテキストファイルである。

著作権ファイル以外のデータファイルは、アジア太平洋機械翻訳協会（ＡＡＭＴ）にて定義されているＵＰＦ（Universal PlatForm）形式で記述されている。

ＵＰＦ形式とは、異なった機械翻訳ソフト間でユーザ辞書を交換するための共通フォーマットであり、記述形式は複数のプラットフォームで可読性を保証するためにテキスト形式で記載されており、各々の辞書情報はＳＧＭＬのようなタグとその内容で表わされている。

3. CD―R のボリューム識別子

ボリューム識別子は次の通り作成されている。

①	②	③	④
ＪＰ	ＪＥ	２００５	００１

① 発行国識別　（“ＪＰ”固定）

② データ識別　（“ＪＥ”固定）

③ 作成年　（データ作成年の西暦年４桁）

④ 連番　（年毎に”００１”から始まる連続番号）

4. ファイル名称

各ファイルのファイル名は以下で作成されている。

- 「著作権ファイル」　　：　COPYRGHT
- 「日英機械翻訳用辞書ファイル（既存分）」　：　JPO_JE_DIC.UPF
- 「日英機械翻訳用辞書ファイル（追加分）」　：　JPO_JE_DIC_ADD.UPF
- 「修正情報ファイル」　　：　CORRECTION.TXT

5. ファイル構成

以下に各ファイルの格納構成を記す。

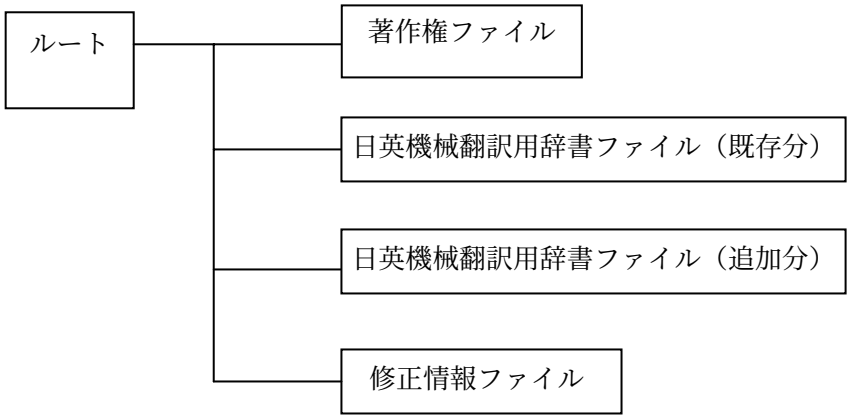


図１　ファイル構成

【ルート】ディレクトリ直下に、各ファイルを作成する。

「修正情報ファイル」は、「日英機械翻訳用辞書ファイル（既存分）」の中に修正内容が含まれる場合のみ作成される。

6. テキストファイルの詳細

6.1. 著作権ファイル（COPYRGHT）

6.1.1. 「COPYRGHT」の内容

提供媒体の著作権を記録したファイル。

6.1.2. 「COPYRGHT」のデータフォーマット

テキストファイル形式とする。フォーマットを下表に示す。

NO	項目	サイズ（byte）	内容例
1	著作権	28	Copyright△(C)△JPO△and△NCIPI△
2	提供媒体の作成年（西暦）	4	2005

（注）“△”はスペースを表す。

6.2. 日英機械翻訳用辞書ファイル（JPO_JE_DIC.UPF、JPO_JE_DIC_ADD.UPF）

6.2.1. 「JPO_JE_DIC.UPF」の内容

前回までに提供した日英機械翻訳用辞書の内容を全て記録したファイル。

6.2.2. 「JPO_JE_DIC_ADD.UPF」の内容

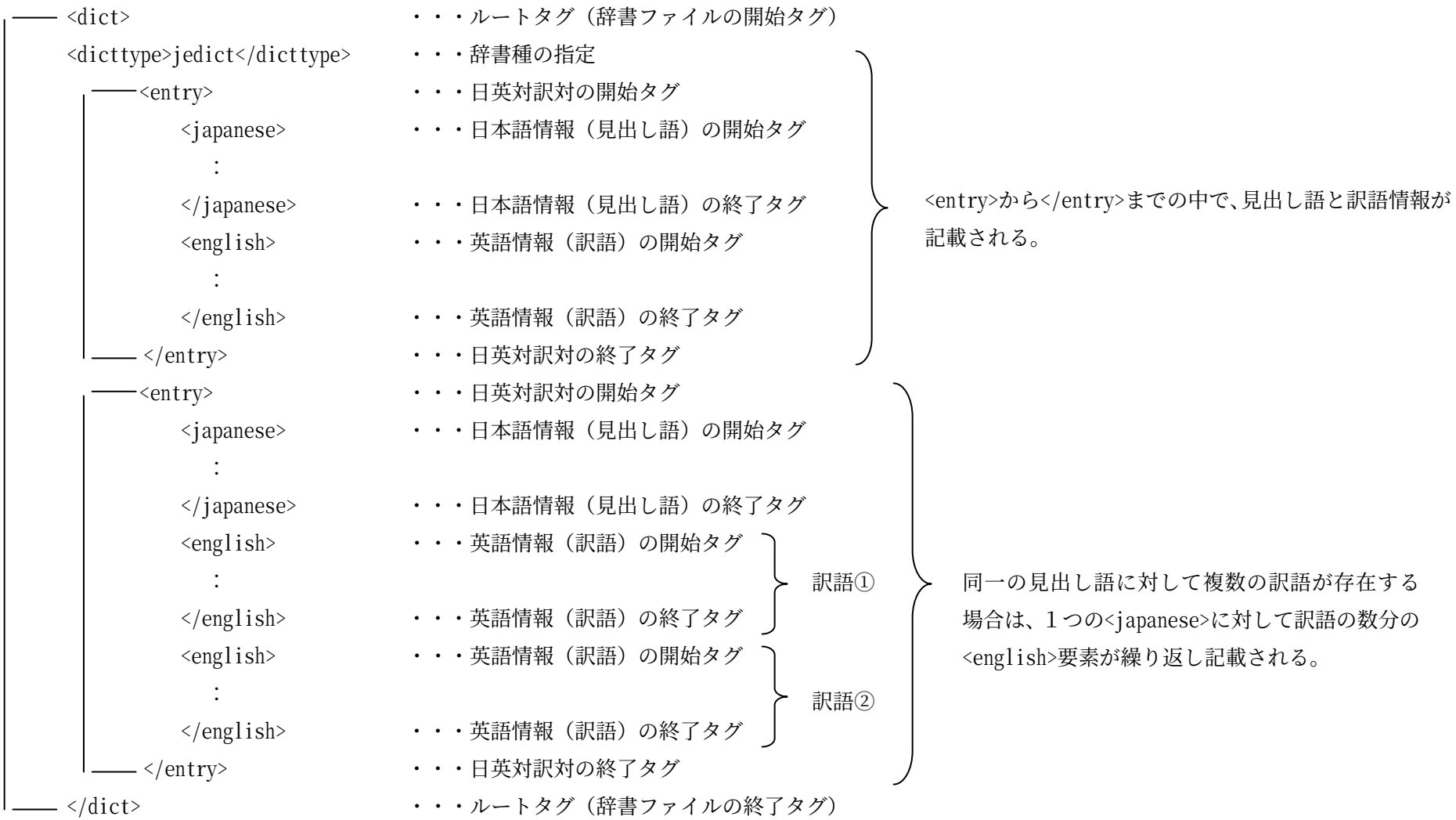
今回新たに追加提供する日英機械翻訳用辞書の内容を記録したファイル。

6.2.3. 日英機械翻訳用辞書ファイルのデータフォーマット

ＵＰＦ形式のテキストファイルとする。

6.2.4. 日英機械翻訳用辞書ファイルの論理構造

（固定値として設定されるタグ）



6.2.5. 日英機械翻訳用辞書ファイルの記述中で使用されるタグ

（1）基本言語変換標準

	UPF基本言語変換標準			
No	開始タグ	終了タグ	タグの意味	日英機械翻訳用辞書ファイルに出力される値
1	<dict>	</dict>	辞書ファイルの開始と終了	(開始・終了タグのため値は存在しない)
2	<dicttype>	</dicttype>	辞書種の指定	jedict （日英辞書を表す)
3	<entry>	</entry>	一つの対訳対の開始と終了	(開始・終了タグのため値は存在しない)
4	<japanese>	</japanese>	日本語の辞書	(開始・終了タグのため値は存在しない)
5	<jentry>	</jentry>	日本語見出し、または訳語	見出し語 (日本語) を記載
6	<jpos>	</jpos>	日本語の品詞	(名詞、動詞、形容詞、形容動詞、副詞、連体詞、単位、サ変名詞) のいずれか
7	<jnountype>	</jnountype>	名詞のタイプ	(普通名詞、固有名詞) の場合のみ出力
8	<jinfl>	</jinfl>	日本語の活用	(一段、五段、カ変、サ変) のいずれか
9	<jcase>	</jcase>	取り得る日本語の格助詞	動詞・サ変名詞：(が、を、に、へ、と、で、から、より、まで) 形容詞・形容動詞：(が、を、に、と、から、で)　のいずれか
10	<sem>	</sem>	意味分類	(人、組織、その他の具体物、時間、場所、その他の抽象物、動物、植物、行為、属性) のいずれか
11	<english>	</english>	英語辞書	(開始・終了タグのため値は存在しない)
12	<eentry>	</eentry>	英語見出し、または訳語	訳語 (英語) を記載
13	<epos>	</epos>	英語の品詞	(noun、verb、adjective、adverb、determiner、unit) のいずれか
14	<enountype>	</enountype>	英語の名詞のタイプ	(commonnoun、propernoun) のいずれか
15	<enum>	</enum>	英語名詞の可算／不可算	(c、u) のいずれか
16	<epl>	</epl>	英語の複数形	(S、ES、IES、VES、U、O、P、複数形のスペル) のいずれか
17	<enumattribute>	</enumattribute>	英語の数属性	(singular、plural) の場合のみ出力
18	<edet>	</edet>	英語の冠詞指定	(the) の場合のみ出力
19	<eheadpron>	</eheadpron>	英語の先頭音	(vowel、consonant) のいずれか
20	<evpresent>	</evpresent>	英語動詞の三単現形	(S、ES、IES、不規則変化後のスペル) のいずれか
21	<evpast>	</evpast>	英語動詞の過去形	(ED、D、IED、KED、Z、不規則変化後のスペル) のいずれか
22	<evpp>	</evpp>	英語動詞の過去分詞形	(ED、D、IED、KED、Z、不規則変化後のスペル) のいずれか
23	<eving>	</eving>	英語動詞のｉｎｇ形	(ING、D、YING、Z、KING、不規則変化後のスペル) のいずれか
24	<ecomparative>	</ecomparative>	英語の比較級	(ER、R、M、I、F、N、不規則変化後のスペル) のいずれか
25	<esuperlative>	</esuperlative>	英語の最上級	(ER、R、M、I、F、N、不規則変化後のスペル) のいずれか

（補足）No13～25 の値については後述。<jpos>、<jcase>、<epos>の値については、資料の都合上、UPF 拡張言語標準で定義されている値も基本言語標準の欄に記載した。

（２）拡張言語変換標準

	UPF拡張言語変換標準			
No	開始タグ	終了タグ	タグの意味	日英機械翻訳用辞書ファイルに出力される値
26	<editor>	</editor>	作成者名	Japan Patent Office and National Center for Industrial Property Information and Training
27	<date>	</date>	作成日付	(データ作成日を記載)
28	<comment>	</comment>	注釈	(任意のコメントを必要に応じて記載)
29	<jadverbtype>	</jadverbtype>	日本語の副詞の用法	(形容詞・副詞修飾、数量修飾) の場合のみ出力
30	<eadjectivetype>	</eadjectivetype>	英語の形容詞のタイプ	(post-attributive) の場合のみ出力
31	<tagdefine>	</tagdefine>	新規タグ定義の開始	(開始・終了タグのため値は存在しない)
32	<tag_name>	</tag_name>	新規タグの名前	本ファイル (第1版仕様) では ehdwd を定義
33	<tag_descript>	</tag_descript>	新規タグの説明	本ファイル (第1版仕様) では ehdwd タグの説明を定義
34	<parent_tag>	</parent_tag>	新規タグの親のタグ名	本ファイル (第1版仕様) で使用する ehdwd タグの親タグは english
35	<value_sets>	</value_sets>	新規タグに対する値	本ファイル (第1版仕様) では ehdwd タグの値を定義
36	<value_sets_descript>	</value_sets_descript>	新規タグに対する値の説明	本ファイル (第1版仕様) では ehdwd タグの値の説明を定義
37	<tagdefine_comment>	</tagdefine_comment>	新規タグに対するコメント	本ファイル (第1版仕様) では ehdwd タグに対するコメントを定義

（補足）No30～37 の値については後述。

6.2.6. 日英機械翻訳用辞書ファイルの記述における補足事項

（1）項番１３ <epos>「英語の品詞」の値について

値	意味
noun	名詞を意味する。
verb	動詞を意味する。
adjective	形容詞を意味する。
adverb	副詞を意味する。
determiner	限定詞（冠詞も含む）を意味する。
unit	単位（数量詞）を意味する。

（2）項番１４ <enountype>「英語の名詞のタイプ」の値について

値	意味
commonnoun	普通名詞を意味する。
propernoun	固有名詞を意味する。

（3）項番１５ <enum>「英語名詞の可算／不可算」の値について

値	意味
c	可算名詞を意味する。
u	不可算名詞を意味する。

（4）項番１６ <epl>「英語の複数形」の値について

値	意味
S	複数形のとき、語尾に s が付加されることを意味する。
ＥＳ	複数形のとき、語尾に es が付加されることを意味する。
ＩＥＳ	複数形のとき、語尾の y が i に変えられ es が付加されることを意味する。
ＶＥＳ	複数形のとき、語尾の fe が v に変えられ es が付加されることを意味する。
U	単数複数同形名詞で、複数形のとき、変化しないことを意味する。 例）sleep
O	不可算名詞であり、複数形はないことを意味する。 例）furniture、information
P	複数扱いの集合名詞、または複数形を訳語としたもので、複数形のとき、変化しないことを意味する。 例）police、parts
（不規則変化後のスペル）	不規則変化を意味する。複数形のスペルが記載される。

（5）項番１７ <enumattribute>「英語の数属性」の値について

値	意味
singular	常に単数扱いを意味する。
plural	常に複数扱いを意味する。

（6）項番１８ <edet>「英語の冠詞指定」の値について

値	意味
the	冠詞が付くとき、定冠詞が付けられることを意味する。

（7）項番１９ <eheadpron>「英語の先頭音」の値について

値	意味
vowel	母音を意味する。
consonant	子音を意味する。

（8）項番２０ <evpresent>「英語動詞の三単現形」の値について

値	意味
Ｓ	三人称単数現在形のとき、語尾に ｓ が付加されることを意味する。
ＥＳ	三人称単数現在形のとき、語尾に es が付加されることを意味する。
ＩＥＳ	三人称単数現在形のとき、語尾の ｙ が ｉ に変えられ es が付加されることを意味する。
（不規則変化後のスペル）	不規則変化を意味する。実際のスペルが記載される。

（9）項番２１ <evpast>「英語動詞の過去形」の値について

値	意味
ＥＤ	過去形のとき、語尾に ed が付加されることを意味する。
Ｄ	過去形のとき、語尾に d が付加されることを意味する。
ＩＥＤ	過去形のとき、語尾の y が i に変えられ ed が付加されることを意味する。
ＫＥＤ	過去形のとき、語尾に ked が付加されることを意味する。
Ｚ	過去形のとき、語尾の子音字を重ね、ed を付加することを意味する。
（不規則変化後のスペル）	不規則変化を意味する。実際のスペルが記載される。

（１０）項番２２ <evpp>「英語動詞の過去分詞形」の値について

値	意味
ＥＤ	過去分詞形のとき、語尾に ed が付加されることを意味する。
Ｄ	過去分詞形のとき、語尾に d が付加されることを意味する。
ＩＥＤ	過去分詞形のとき、語尾の y が i に変えられ ed が付加されることを意味する。
ＫＥＤ	過去分詞形のとき、語尾に ked が付加されることを意味する。
Ｚ	過去分詞形のとき、語尾の子音字を重ね、ed を付加することを意味する。
（不規則変化後のスペル）	不規則変化を意味する。実際のスペルが記載される。

（１１）項番２３ <eving>「英語動詞の ing 形」の値について

値	意味
ＩＮＧ	現在分詞形のとき、語尾に ing が付加されることを意味する。
Ｄ	現在分詞形のとき、語尾の e が除かれ ing が付加されることを意味する。
ＹＩＮＧ	現在分詞形のとき、語尾の ie が除かれ ying が付加されることを意味する。
Ｚ	現在分詞形のとき、語尾の子音字が重ねられ ing が付加されることを意味する。
ＫＩＮＧ	現在分詞形のとき、語尾に king が付加されることを意味する。
（不規則変化後のスペル）	不規則変化を意味する。実際のスペルが記載される。

（１２）項番２４ <ecomparative>「英語の比較級」の値について

値	意味
ＥＲ	比較級のとき、語尾に er が付加されることを意味する。
Ｒ	比較級のとき、語尾に r が付加されることを意味する。
Ｍ	比較級のとき、前に more が付加されることを意味する。
Ｉ	比較級のとき、語尾の y が i に変えられ er が付加されることを意味する。
Ｆ	比較級のとき、語尾の子音字が重ねられ er が付加されることを意味する。
Ｎ	比較級のとき、変化しないことを意味する。
（不規則変化後のスペル）	不規則変化を意味する。実際のスペルが記載される。

（１３）項番２５ <esuperlative>「英語の最上級」の値について

値	意味
ＥＲ	最上級のとき、語尾に est が付加されることを意味する。
Ｒ	最上級のとき、語尾に st が付加されることを意味する。
Ｍ	最上級のとき、前に most が付加されることを意味する。
Ｉ	最上級のとき、語尾の y が i に変えられ est が付加されることを意味する。
Ｆ	最上級のとき、語尾の子音字が重ねられ est が付加されることを意味する。
Ｎ	最上級のとき、変化しないことを意味する。
（不規則変化後のスペル）	不規則変化を意味する。実際のスペルが記載される。

（１４）項番３０ <eadjectivetype>「英語の形容詞のタイプ」の値について

値	意味
post-attributive	後置修飾を意味する。

（１５）項番３１～３７について

訳語（英語）が複合語（単語が複数から成り立つ用語）の場合、活用対象となる単語が何番目に位置するかの情報が必要となる。
そこで、日英機械翻訳用辞書では、複合語における活用対象語の情報を新規タグを用いて表現する。

新規タグの定義内容

```
<tagdefine>
<tag_name>ehdwd</tag_name >
<tag_descript>複合語の中で活用対象となる単語が何番目に出力されているかを数字で記載</tag_descript>
<parent_tag>english</parent_tag>
<value_sets>1-39</value_sets >
<value_sets_descript>半角文字で１つの数字のみ記載される</value_sets_descript>
< tagdefine_comment>日英機械翻訳用辞書新規タグ< /tagdefine_comment>
</tagdefine>
```

（１６）項番５ <jentry> と、項番１２ <eentry> の値の補足事項

<jentry>の値として“※”が記述されている場合、“※”の部分は変数扱いとみなし、数字の羅列をそのまま読み込み訳語として反映させる用語を意味しています。この場合、<eentry>の中では“※”に対応する訳語を“<n>”（n は数字で※の出現順に１から付番される）として記述しています。

“※”を記述する場合の例

原文：実開平１５－１２３４５
訳文：Publication of unexamined utility model application Heisei 15-12345

上記のような原文に対して、上記訳文を出力したい場合、

見出し語：　<jentry>実開平※－※</jentry>
訳語　　：　<eentry> Publication of unexamined utility model application Heisei <1>-<2></eentry>

と定義することで、原文中、※の部分がどのような数字であっても、それをそのまま訳文に反映できる、という意味合いになります。

6.2.7. 日英機械翻訳用辞書ファイルの内容サンプル

```
<dict>
<editor>Japan Patent Office and National Center for Industrial Property Information and Training</editor>
<date>2005.03.31</date>
<dicttype>jedict</dicttype>
<tagdefine>
<tag_name>ehdwd</tag_name >
<tag_descript>複合語の中で活用対象となる単語が何番目に出力されているかを数字で記載</tag_descript>
<parent_tag>english</parent_tag>
<value_sets>1-39</value_sets >
<value_sets_descript>半角文字で１つの数字のみ記載される</value_sets_descript>
< tagdefine_comment>日英機械翻訳用辞書新規タグ< /tagdefine_comment>
</tagdefine>
  <entry>
    <japanese>
      <jentry>【優先権主張番号】</jentry>
      <jpos>名詞</jpos>
      <jnountype>普通名詞</jnountype>
    </japanese>
    <english>
      <eentry>[Application number of the priority]</eentry>
      <epos>noun</epos>
      <enountype>commonnoun</enountype>
      <enum>u</enum>
      <epl>0</epl>
      <enumattribute>singular</enumattribute>
      <eheadpron>vowel</eheadpron>
    </english>
  </entry>
```

（つづく）

（注）サンプル記述では、タグ構造（階層）を判り易くするために先頭にスペースを補記してあるが、実データ上は先頭スペースがなく記述されている。
なお、<entry>から</entry>までの記述も、実データでは改行なしで記述されている。

（つづき）

```
<entry>
  <japanese>
    <jentry>歯軸</jentry>
    <jpos>名詞</jpos>
    <jnountype>普通名詞</jnountype>
  </japanese>
  <english>
    <eentry>tooth axis</eentry>
    <epos>noun</epos>
    <enountype>commonnoun</enountype>
    <enum>c</enum>
    <epl>axes</epl>
    <eheadpron>consonant</eheadpron>
    <ehdwd>2</ehdwd>
  </english>
</entry>
</dict>
```

訳語が複合語の例

6.3. 修正情報ファイル（CORRECTION.TXT）

6.3.1. 「CORRECTION.TXT」の内容

「JPO_JE_DIC.UPF」（日英機械翻訳用辞書ファイル（既存分））の中の修正情報を記録したファイル。

6.3.2. 「CORRECTION.TXT」のデータフォーマット

テキストファイル形式とする。

6.3.3. 「CORRECTION.TXT」の論理構造

行頭が<!correct-unit>で始まる行から</!correct-unit>の行の前までが、１つの修正情報の単位とする。
＃で始まる行はコメント行とする。（“変更情報”または“削除情報”などを記載する）
<!correct-info>で始まる行から次の<!correct-info>か</!correct-unit>の行の前までが、具体的な修正内容とする。

“変更情報”の場合は、“変更前の登録内容”を<!correct-info>Delete</!correct-info>の次の行に記載し、“変更後の登録内容”を<!correct-info>Add</!correct-info>の次の行に記載する。
登録内容のうち、一部の属性のみを追加・削除した場合は“変更情報”として表す。

“削除情報”の場合は、“削除した登録内容”を<!correct-info>Delete</!correct-info>の次の行に記載する。

次ページに「CORRECTION.TXT」の論理構造例を示す。

```
<!correct-unit>-------------------------------------------------------------
# 変更情報
<!correct-info>Delete</!correct-info>
    <entry>
        （変更前の登録内容）
    </entry>
<!correct-info>Add</!correct-info>
    <entry>
        （変更後の登録内容）
    </entry>
</!correct-unit>
<!correct-unit>-------------------------------------------------------------
# 削除情報
<!correct-info>Delete</!correct-info>
    <entry>
        （削除した登録内容）
    </entry>
</!correct-unit>
```

<entry>から</entry>までの記述内容は、前述の「日英機械翻訳用辞書ファイル」に準じる。