対訳辞書における非同義関係の記述とその利用

7F－1　三池　誠司　松川　智義　安達　久博　天野　真家

㈱日本電子化辞書研究所

1．はじめに

㈱日本電子化辞書研究所では，自然言語処理のための大規模な電子化辞書を開発している．本稿では，その中の対訳辞書における，同義関係にない訳語についての記述とその利用について報告する．

2．対訳辞書と訳語情報

電子化辞書は，図１に示すように構成されている（専門用語辞書と共起辞書を除く）[1]．

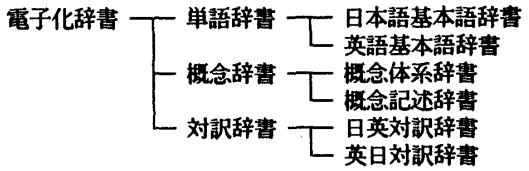

図１　電子化辞書の構成

対訳辞書内のデータは図２のような階層構造をなす．

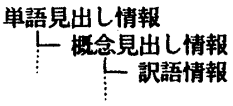

図２　対訳辞書の構造

対訳辞書は，単語辞書との間で「単語見出し」と「概念見出し」を共有している．概念見出しは単語見出しが持つ概念の説明文（語義に対応するもの）で，概念辞書への見出しでもある．図２の各階層では，上位の項目１つに対して１つあるいは複数の下位項目が記述されている．

3．訳語情報の内容

訳語情報には，訳語（対訳語の単語見出し），訳語区分，補足説明がある．

人間用の対訳辞書では，説明文的に訳語が記述されていることが多い．本対訳辞書では，自然言語処理の観点から，語の意味を示すための「説明文」としてではなく，言語表現中に表出される「訳語」として適切であることに注意を払って記述している[2]．また，人間用の対訳辞書では，見出し語と説明文中に現れる訳語との関係が必ずしも明確ではない．本対訳辞書では，見出し語と訳語の関係を，概念の包含関係によって，同義，類義，下位，上位の４種類に分類し，「訳語区分」として明示している．訳語区分が同義以外の場合に，概念見出しと訳語の概念の差を補正する情報として，「補足説明」を記述している．補足説明については後述する．

基本的に，１つの訳語区分を選択し訳語を記述している．複数の区分に訳語が存在する場合は同義を優先し，同義がなかった場合は，類義，下位，上位の順で最適と思われる訳語を記述している．

4．訳語情報における補足説明

訳語区分が同義以外の場合には，日本語と英語の概念の対応関係が一対一でないことを示している．補足説明は，その差を明確にするための補助的手段である．

(1) 訳語区分が下位の場合

一般に訳語は複数記述している．個々の訳語はそれぞれ異なった状況で使用されるものである．使われる状況を限定するために，修飾語句などを＜＞の中に記述している．以下の例では，左列が見出し語，右列が見出し語の特定の１つの概念見出しに対する訳語と補足説明である．

(例1)

麦	barley<general in Europe> wheat<to yield flour> rye<for livestock feed>
horizon	＜地面との＞地平線 ＜海面との＞水平線
eraser	＜鉛筆の＞消しゴム ＜チョークの＞黒板消し ＜インクの＞インク消し
Danish	＜言語では＞デンマーク語の ＜人では＞デンマーク人の ＜国では＞デンマークの
decoder	＜物では＞デコーダ ＜人では＞解読者

(2) 訳語区分が上位の場合

概念見出しの概念より訳語の概念が広いので，訳語の概念を限定するための修飾語句や共起語句を () の中に記述している．補足説明を訳文の生成に用いることで適切に翻訳できる場合があることから，補足説明を訳語の一部として利用しても自然であることに配慮した表現にしている．

(例2)

兄	(elder)brother
消しゴム	(pencil)eraser
黒板消し	(blackboard)eraser
インク消し	(ink)eraser
chairwoman	(女性の) 議長
cap	(つばのない) 帽子
hat	(つばのある) 帽子
kill	(語・パラグラフを) 削除する

(3) 訳語区分が類義の場合

下位の訳語として用いられる場面を限定する補足説明

Usage of the Description for Unequivalents in a Bilingual Dictionary
Seiji MIIKE, Tomoyoshi MATSUKAWA, Hisahiro ADACHI and Shin-ya AMANO
Japan Electronic Dictionary Research Institute, Ltd.

（＜＞で表示）と，上位の訳語としてその概念を限定する補足説明（（）で表示）を併用して記述している．

(例3)

abbey	＜教会で＞（以前修道院だった）大教会 ＜邸宅で＞（以前修道院だった）大邸宅

5．補足説明の機械翻訳への利用例

一般に機械翻訳などにおいて同義関係でない訳語を生成するには困難を伴う．補足説明はそのような場合に有用な情報を提供し，次のような利用が可能である．

(1) 下位についての補足説明
補足説明と訳語のペアを表示しユーザの選択により半自動的に適切な訳語を選択する，あるいは，補足説明中の語句が文脈に出現していることの検査などにより訳語選択を行うなどの利用が可能である．

(2) 上位についての補足説明
例えば「I have two caps and five hats.」という文では，「つばのない帽子を2つと，つばのある帽子を5つ持っている．」と訳す必要がある．このとき，例2の「cap」と「hat」の補足説明をそのまま利用することで対処できる．

6．対訳辞書から概念辞書へのフィードバック

前節では，自然言語処理への利用という観点から補足説明をみた．一方，以下の2点に関し，現在開発中の概念辞書を充実させることを目的に補足説明を用いることが可能である．

6．1　概念の切り分け

英日対訳辞書の下位区分の補足説明に関して次のようなデータが得られた．なお概念体系とは，概念体系辞書において，概念の結び付きの観点から概念を分類し，分類の各集合に分類項目名というラベルをつけたものである．

①補足説明は「自立語＋助詞」のパターンが約 95%である．
②補足説明は延べ約 2,200個あり，異なり数は約 640である．
③自立語が概念体系における概念分類項目名を含む補足説明の異なり数は約 160個（異なり数比 25%）である．

③に該当するものは，例1の「Danish」や「decoder」の補足説明である．①～③から，「下位区分における補足説明は，重複するものが多く，また当該の訳語が属する概念分類項目を示唆するものが多い．」という結果が得られた．したがって，それらの概念（概念見出し）を分割の対象とすることができる．
概念見出しはトップダウンに設定し，概念辞書の構築の過程で概念見出しの統合化と安定化を行う．上述の補足説明は，分割すべき概念見出しの集合を与えることにより概念見出しの安定化に寄与する．また，補足説明の表記によって，分割後の概念見出しが概念体系のどの分類項目に属するかを知ることができる．

6．2　2言語間の概念の対応関係の記述

特に上位区分では，英語と日本語間における概念の結び付きの違いを示唆する補足説明がみられる．例えば，例2における「kill」の概念見出し（語・パラグラフを削除する）をc#kill，訳語の「削除する」の該当する概念見出しをc#sakujo，「語」と「パラグラフ」の該当する概念見出しをそれぞれc#go，c#paraとすると，それらの関係は概念記述辞書において図3のように示されることになる．（概念記述辞書では，概念間の関係が概念見出しと関係子（例えばobject）で記述される．）この場合訳語区分が上位として示されることは，c#sakujoがc#killを包含することであり，c#sakujoに結び付く概念の数がc#killのそれより多いことを意味する．つまり，c#sakujoに関係子objectで結び付く概念はc#go，c#paraなどに限らないということである．

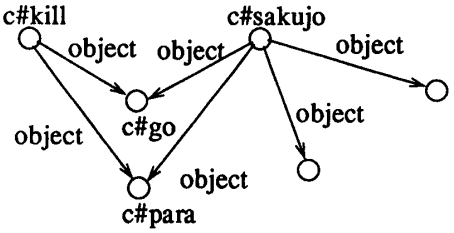

図3　概念記述辞書における概念の関係の記述

図3において重要なことは，c#goとc#paraがc#killとc#sakujoに共通に結び付く概念であることである．概念記述辞書のデータはテキストの分析によって得る．したがって，同義の概念見出しは概念関係のデータを共有できるが，同義以外の概念の対応関係は，日本語・英語両方のテキストの分析，あるいは日英対訳テキストの分析を待たねば得られない．このテキストの分析に対し，補足説明はその関係の一部（同義以外の概念見出しがどの概念見出しに共通に結び付くか）を最初に与えるものである．

7．おわりに

本対訳辞書では，説明文と訳語を峻別して訳語表記を規格化し，訳語区分を設けている．さらに，概念見出しと訳語の概念に一対一の対応がない場合に，その差を補う目的で補足説明を記述している．補足説明は，機械翻訳に必要とされる情報を含んでいる．
さらに補足説明の中には，単語見出しの概念の切り分けを示唆するものと，2言語間の一対一に対応しない概念に介在する概念を示すものがある．前者は概念見出しの安定化に寄与し，後者は概念辞書に含まれるべき2言語間の概念の結び付きの一部を与える．補足説明から得られるこれらの情報の抽出と類別には人手を要するが，本文で示したデータなどから一部機械的な前処理が可能であるとみられる．

【謝辞】
本研究の機会を与えて頂いた横井所長，並びに貴重な意見・コメントを頂いた研究員の皆様に感謝いたします．

【参考文献】
[1] 日本電子化辞書研究所，対訳辞書（第1版）TR-014，日本電子化辞書研究所，1989
[2] 熊野他，自然言語処理のための対訳辞書における訳語選定，情報処理学会第35回全国大会1S-3，1987
[3] 日本電子化辞書研究所，概念辞書（第2版）TR-012，日本電子化辞書研究所，1989