情報処理学会 第 47 回全国大会講演論文集, Vol.3, pp.183-184, 1993.

英日ＭＴシステムＤＵＥＴ－Ｑｔ用２ヶ国語基本辞書の構築 －－ 高精度ＭＴを目指して －－

佐田 いち子

（シャープ株式会社　情報商品開発研究所）

E-mail address: sata@islix.sharp.co.jp

1　はじめに

本稿では、商用英日ＭＴシステム『ＤＵＥＴ－Ｑｔ』のための２ヶ国語基本辞書（見出し語総数：１９９３年８月現在、約８８，０００語）の主な特徴について概説する。今回、言語処理部における翻訳方式の改良［1］に伴い、当辞書の枠組を拡張して動詞型尤度情報，格パターン優先度情報，新方式共起処理情報，等の新規情報を追加すると共に、意味共起情報，品詞尤度情報，例外的変換規則，分野情報，等の拡充を行った。

2　尤度情報

ＤＵＥＴ－Ｑｔでは、構文的曖昧性解消のために、文法規則、及び辞書に記述された尤度情報等に基づく独自の優先解釈方式を採用している。辞書の尤度情報には、以下の２つがある。

2.1　多品詞語の尤度情報

ＭＴにとって問題とされるのは、原言語の１つの単語が目標言語において複数の品詞を持っているため、必然的に構文的曖昧性が増大するという点である。この問題に対処するために、辞書の同一見出し語の異品詞間における尤度情報が利用される。現在、全ての［見出し語＋品詞ペア］は、２段階、すなわち『NORMAL 品詞』と『MINOR 品詞』に分類されているが、今回の改良では、評価文を用いて更に当情報の充実を計った。

構文木優先度決定処理の一環として、構文解析時に『NORMAL 品詞』を用いた場合は高い点数、『MINOR 品詞』を用いた場合には低い点数が与えられる仕組みになっている。また、熟語（２単語以上からなる見出し語）が『MINOR 品詞』に分類されている場合には、文法規則に基づき、更に複雑な優先解釈処理が行われる。

2.2　動詞型の尤度情報

動詞は、主として A. S. Hornby の動詞型分類に基づき、約１２０種類の動詞型（DUET 独自）に分類されており、全ての動詞型は文法規則において構文的点数を定められているが、同構文的点数は、常に全ての単語に適用できるとは限らない。従って、辞書の見出し語単位で、動詞型の構文的点数を例外的に定めることが可能になっている。従来の『MINOR 型』と『NORMAL 型』の２値の尤度に加え、『MAJOR 型』を新規採用すると共に、『MINOR 型』と『MAJOR 型』については、その度合を数値で表示し、より木目細かい尤度の設定を可能とした。

3　意味共起情報と格パターン優先情報

意味共起情報は、構文優先解釈、及び語義選択に利用される。今回、新たな試みとして、英文コーパスから抽出した木目細かい意味制限情報を『自然言語』のまま記述し、『自然言語：意味コード』ペア・テーブル（既存名詞データに付与されている各訳語と意味属性情報から自動作成）を介して自動的に『意味コード』に変換するという方式を採用してみた。この結果、意味共起情報記述作業の大幅な効率ＵＰを計ることができた。

また、意味共起情報を記述した各格パターンには、『格パターン優先情報』を併せて付与し、意味的に好ましい解釈の優先を計った。『格パターン優先情報』としては、従来の『優先フラグ』（２値：優先する／しない）に加え、数値で序列を示した『格パターン優先度』を新たに用いた。

4　例外的変換規則

当辞書では、２ヶ国語辞書の利点をフルに生かし、より自然な訳語の生成を行うための様々な情報を記述することができる。今回は、評価文を用いて当情報の充実を計った。

4.1　言い換え翻訳

入力文が受動態であれば、通常、「れる／られる」を述語動詞の語尾に付加した訳語が生成されるが、場合によっては、当規則を適用すると、非常に難解な訳語になってしまうことがある。これを回避するために、例外的変換規則として『受動態専用訳語情報』を辞書に記述することができる。

また、入力文が能動態の場合であっても、一般の変換規則に基づいた生成結果よりも自然な訳語を出力するために、『能動態専用訳語情報』を辞書に記述することができる。

4.2　生成語順指定

通常、生成の際の語順は、深層格記述に基づく訳語生成規則よって決定されるが、例外的規則として訳語と共に、生成語順を辞書に指定することが可能である。

5　新方式共起処理情報

どのような言語においても、語と語が強く結び付く特定の組合せがあるといえよう。このような特殊な共起関係（＝連語）は、慣用句とはいくぶん異なるものだが、それらを他の言語に変換するのは、かなり難しい場合が多い。共起関係は、従来、（ａ）熟語見出しとして登録、（ｂ）一方の辞書に、共起の相手先（原言語）を指定する，等の方法で処理されていた。（ａ）に関する問題点は、そのような熟語登録が辞書のサイズを不必要に増大させること、また（ｂ）に関する問題点は、原言語を指定するだけでは目標言語の語義までは保証されないということである。これらの問題に対処するために、独自の『単独特殊共起コード手段』または『集合特殊共起コード手段』を用いて特定［単語：訳語］ペアと特定［単語：訳語］ペアとの共起処理を行う方式を採用した。

5.1　『単独特殊共起コード手段』を用いた共起処理

１つの［単語：訳語］ペアと、別の１つの［単語：訳語］ペアが連語をなす場合に、一方の単語辞書の特定訳語ページに、相手先の［単語：訳語］ペアを表す『単独特殊共起コード』を指定することにより、完全な連語訳の生成を可能にする。

5.2　『集合特殊共起コード手段』を用いた共起処理

１つの［単語：訳語］ペアと、複数個の［単語：訳語］ペアが各々連語をなす場合に、基軸となる方の単語辞書の特定訳語ページに、相手先の［単語：訳語］ペアの集合を表す『集合特殊共起コード』を指定することにより、各々の組合せにおいて、完全な連語訳の生成を可能にする。

6　おわりに

今回の改良では、意味共起情報の拡充に力を入れ、実際に当情報が、意味的曖昧性解消に有効であるという検証を行うこともできた。しかし、生きた『ことば』を有限個の『意味マーカ』で分類することには問題点が多いのも事実である。一貫性のある階層状意味体系を構築することは非常に難しく、また人手による意味の分類には、必ずゆらぎが生ずる。今後は、使用目的に合わせて複数の意味体系（必ずしも階層構造でなくてもよいだろう）を構築していくことも必要であろう。

今回も英文コーパスから意味共起情報を抽出することを試みたが、今後は更に大規模なコーパスの構築を行うと同時に、そのコーパスからＭＴに有効な情報を抽出し、辞書に追加していくことが課題である。

［参考文献］

［１］　福持陽士他：ルールベースの優先解釈と曖昧性解消，情処学会第４７回全国大会，１９９３．