

英日機械翻訳システムにおける派生語推定とその訳語付与の改良

佐田 いち子 吉見 毅彦 九津見 毅 岡本 純子

シャープ株式会社 ソフト事業推進センター

{sata, yoshi, kutsumi, o-junko}@isl.nara.sharp.co.jp

1 はじめに

著者らの開発している『シャープ英日翻訳支援システム Power E/J』では、現在、改良の主眼として、辞書索きのできない(できても不適切になる)単語・熟語を少しでも減らすことに取り組んでいる。その手段の一つとして、接頭辞・接尾辞を検出することによって、そのままでは辞書に登録されていない派生語の辞書索きを可能にする処理の強化を行っている。この処理は従来から組み込まれているものであるが、この処理の適用方法をより適切にし、より自然な訳語とするための改良を行っているので報告する。

接尾辞	品詞	訳
dependent	ADJ	依存の 指向の
oriented	ADJ	
ility	NOUN	
ster	NOUN	
ness	NOUN	
ible	ADJ	
able	ADJ	
ive	ADJ	
ist	NOUN	
ing	VERB	
or	NOUN	された
er	NOUN	
ed	VERB,ADJ	

2 語基と接辞の認定

翻訳支援システムの辞書索き処理において、入力単語に対し、基本辞書・ユーザ辞書などの見出しおよびその変化形にマッチしなかった場合、派生語処理に移るが、そこでの接頭辞・接尾辞の認定は基本的には以下のように行う。

接頭辞 接頭辞テーブルを所定の順序(例: アルファベット順)に作っておいて、スキャンして最初にマッチしたものの

接尾辞 接尾辞テーブルを所定の順序(例: 第1キーを長さの長い順、第2キーをアルファベット逆順)に作っておいて、スキャンして最初にマッチしたものの

この点にまつわる各問題について、著者らの採った解決策と共に述べる。

接頭辞	訳
ex	前
non	
re	再
un	不

図 1: 接頭辞テーブルの例

図 2: 接尾辞テーブルの例

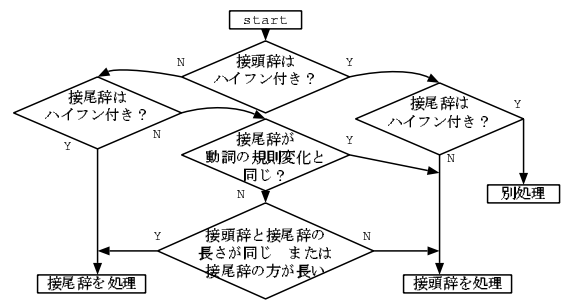


図 3: 接頭接尾辞ともマッチする場合の処理決定手順

2.1 接頭辞と接尾辞の競合

接頭辞と接尾辞の双方ともがマッチする場合の判定手順について、著者らは実験の結果、図3のように定めた。この判定の結果、語基として既知語が索かれなかった場合は、異なる接辞の処理に移る。図4では最終的に採用される接辞と語基に下線を引いているが、たとえば nonradioactive の例は、最初の判定では接尾辞 -ive が試されるが、nonradioact が未知語であると判明するので、最終的に接頭辞 non- が処理される。また、接尾辞が動詞の変化形と同じ形(-ing/-ed)だと接頭辞処理優先としているが、これは、接尾辞テーブ

ルの-ing/-ed は元来、非動詞から動詞を派生している場合を主に想定しているので、接頭辞処理をすれば、-ing/-ed を含めた語基が動詞の変化形とみなされる可能性が高いからである (図 4 の unborning, unseeded, untraded)。

2.2 接頭辞同士、接尾辞同士の競合

図 4 の assister の例では接尾辞 -er と -ster がマッチするが、このような場合は最初は長い方の接辞 (-ster) を採用する。これで語基として既知語が索かれなかったら短い方の接辞 (-er) の処理に移る。

2.3 同一の接辞に対する異なる語基の競合

派生語認定時の調整処理のため、同一の接辞に対して複数通りの語基が認定され、そのいずれもが既知語である場合が存在するが (図 4 の noter)、このような場合は語基の品詞を参照して不適な候補を落とす。

この判定の結果、接辞と語基が認定されるにもかかわらず、全ての候補を不適として、接辞処理をやめて全体を未知語と処理する場合もある (図 4 の mer)

単語	マッチする接辞	マッチする語基
exportable	ex-, -able	portable, export
nonradioactive	non-, -ive	radioactive, nonradioact
unborning	un-, -ing	boring, unbor, unbore
unseeded	un-, -ed	seeded, unseed
untraded	un-, -ed	traded, untrad, untrade
assister	-er, -ster	assi, assist, assiste
noter	-er	not, note
mer	-er	m, me

図 4: 接辞処理例

3 訳語生成

図 1・図 2 の接頭辞・接尾辞テーブルの訳語欄に訳語が存在している場合は、派生語の訳語生成方法は、原則として語基の訳語に接辞の訳語を接続したものである (状況により例外もある)。

接頭辞・接尾辞テーブルの訳語欄に訳語が存在していない接辞については、主に語基の訳語の形に合わせて、語基の変形のさせ方をきめ細かに定める方法をこのたび導入した。現在、十数種類の接辞について生成規則を導入している。このうち接頭辞 non- の規則を図 5 に、接尾辞 -ness の規則 (24 件のうちの一部分) を図 6 に示す。

語基の訳の形	変形のさせ方	処理結果
~名詞+『の』	『非』+語基の訳	nonacid 非酸性の, nonradioactive 非放射性的
~形容動詞 (『~的な』)	『非』+語基の訳	nonactive 非活動的な, nonorganic 非有機的な
~形容動詞 (『~的な』以外)	語基の訳の連用形+『ない』	nonimportant 重要でない, nonprecious 貴重でない
~形容詞	語基の訳の連用形+『ない』	nonreligious 信仰が厚くない, nonpossessive 所有欲が強くない
~格助詞+『ある』	『ある』を『ない』に置換	nonremunerative 利益がない
~動詞 (上記以外)	語基の訳の未然形+『ない』	nonabrasive すり減らさない, noninhabitable 住むことができない
~動詞+『た』	『た』を『て』に置換+『ない』	nonstable 安定していない, nonsegregated 分離されていない

図 5: 訳語生成規則の例 (non-)

語基の訳の形	変形のさせ方	処理結果
~『が』+形容詞 (『ない』以外)	~『の』+形容詞の語幹+『さ』	amiableness 感じのよさ
~形容詞 (『ない』以外)	~形容詞の語幹+『さ』	acrimoniousness 厳しさ
~『的な』的	『的な』的を『性』に置換	abidingness 永続性
~『{が の}』可能な』	~『の可能性』	medicableness 医療の可能性
~『{が の}』+形容動詞 (『可能な』『ような』以外)	~『の』+形容動詞の語幹+『さ』	expressiveness 表情の豊かさ
~『可能な』	~『可能性』	removableness 移動可能性
~形容動詞 (『可能な』『ような』以外)	~形容動詞の語幹+『さ』	abstruseness 難解さ
~『{名詞 サ変動詞語幹}+『できる』+『出来る』+『出来る』	『できる』+『出来る』を『可能性』に置換	adaptiveness 適応可能性
~『に』+サ変動詞	語基の訳+『こと』	laudableness 賞賛に値すること
サ変動詞 (上記以外)	サ変語幹	ubiquitousness 偏在

図 6: 訳語生成規則の例 (-ness, 抜粋)

4 おわりに

派生語処理の改善を進めていくと、実に多くの考慮すべき事情が存在することが判明してきており、ここに挙げたのはその一端にすぎない。(ここに挙げた他にも、訳語生成への語基の意味情報の利用などを行っている。)

今後は、辞書の充実を図りながらも、これらの問題に対処していき、形態素解析の能力自体を人間の持つ単語認識能力・造語能力に近づけていきたいと考えている。

5 謝辞

本研究にあたり日頃より御指導頂いているソフト事業推進センターの仲川與万所長、福持陽士副参事、並びに有益な示唆を頂いている英日翻訳関連グループの諸氏に深謝する。