

対訳付慣用表現の収集について

4B-9

田中 康仁

兵庫 大学

Email: yasuhito@humans-kc.hyogo-dai.ac.jp

0) はじめに

機械翻訳システムが多くのメーカーやソフトウェア会社から発表されている。しかし、品質はまだまであり、充分使えるというほどのものではない。この理由としては、「文は単語の集まりである」という考えから単語辞書は充実させたが、それ以外の単語と単語の結合したものについてはあまり充実していない。専門用語については少し集められている。

一方、「文は類似した文の学習から作られる」という考えに基づき、文の類似度からの翻訳が研究し実用化なされようとしている。これにも、どれだけ文の集合でよいのか？、どれだけ類似度で類似とみなすかなど問題は多い。そこで、筆者は単語と単語の結合したものにもっと関心をはらうべきだと考えている。建築でよく使われる中間部材や台所用品などの施設と同じ考えに立つ。

そこでここでは慣用表現に焦点をあてる。

1) 慣用表現の数量的分析

慣用表現（熟語）がどの程度使われているか数量的に分析した資料を調べてみた。旺文社刊の「英熟語ターゲット1000」には16年間の日本全国の大学入試に出された試験問題から慣用表現について調べている。その異なりの総数は14,368である。

1回	……	8,083	(56%)
2回	……	1,884	(13%)
3回	……	926	(6.4%)
4回	……	532	(3.7%)
5回	……	417	(2.9%)
		11,842	(82%)

出典「英熟語ターゲット1000」旺文社

大学入試はおもしろい対象である。これで全ての慣用表現を推定するのは大変であるが一つの目安になる。

How do we extract Idiomatic knowledge!!
Yasuhito Tanaka
Hyogo University

筆者の考えでは、この4~5倍程度は集めなければならないと思う。5倍とすると7万程度の慣用表現を集めなければならないことになる。

2) 慣用表現をどのように考えるか

慣用表現を集めた本を数冊調べその索引を見ると、次のようになっている。

- (i) 動詞句 ex. 動詞⊕前置詞のようなもの
例: look up
- (ii) 前置詞句
例: in front of
- (iii) 副詞句
例: as soon as
- (iv) be動詞⊕動詞句
- (v) 否定形notを含むもの
- (vi) その他

がある。このうち(i)、(ii)が多い。

3) 英語の前置詞に注目する。

英文と日本語文が対になったパラレル・コーパスを分析の対象と考える。しかし、パラレル・コーパスはなかなか入手できない。もし、これらが入手できるとすると、英文を中心に考える。まずKWICを作成する。次のようなものを作る。

英文の一部 (前)	前置詞	英文の一部 (後)	日本語(対訳)

このようなKWICを作成し、前置詞を含む右側で分類すると前置詞句の集めやすいKWICになる。

また、前置詞を含む左側にむかって分類すると動詞句が調べやすいことがわかる。人手による調査、エディターによる調査、自動(半自動)抽出等について考えねばならない。

4) 英語の前置詞

英語の前置詞を調べてみると3種類のものがある。これは首藤訓宏、中島武者(「前置詞・イディオムの英会話」-重要20前置詞で表現自在-ダイヤモンド社)から引用する。

前置詞を分類すると次のようになる。

1. Simple Prepositions「単純前置詞」
2. Compound Prepositions「複合前置詞」
3. Participial Prepositions「分詞前置詞」

これらは次のようになっている。

Simple Prepositions (23)

at	forth	on	till
but	from	out	to
by	in	over	up
down	like	per	via
* ere	of	save	with
for	off	through	

(ere[ɛə]「前に」の意だが、現在では詩などに使われるだけ)

Compound Prepositions (43)

abaft	amidst	besides	throughout
aboard	among	between	toward(s)
about	* anent	* betwixt	under
above	around	beyond	underneath
across	* aslant	despite	until
* adown	* athwart	except	unto
after	before	inside	upon
against	behind	into	* withal
* aloft	below	onto	within
along	beneath	outside	without
amid	beside	since	

Participial Prepositions (13)

* barring	during	past	respecting
bating	excepting	pending	saving
concerning	* notwithstanding	regarding	touching
considering			

(*barring=excepting「を除いて」。notwithstanding「にもかかわらず」)

これらのうち単純前置詞だけを調べればほぼ良いことが前述の旺文社の資料を調べる中でわかる。

慣用表現を抽出するには対訳コーパスを使い抽出する方法が考えられる。

もう一つの方法は辞書からの抽出である。

5) どのような辞書や本を対象とするか？

厚い本で、立派な慣用表現の辞書がある。しかし、これら本は良いものであるが次のような問題がある。

i) 著作権の問題

ii) 本ができるまでに長期間かかっているため少し、古くなっている。

iii) 本の著者によって慣用表現の選別がなされている。これは良い面もあるが、悪い面もある。

本からの引用には著作権の問題は常につきまとう重要な問題点であることは忘れることはできない。

次のような本や雑誌を対象とする。

高等学校の生徒が使用する参考書、雑誌、How to 物のような実用書の中から慣用表現(英語)と対訳を頭の中に入れ記憶し、その後、その内容を抽出する。

このような本は、厚い本を作成する前段階として作られたものや、商業主義的なおもしろさや実用主義を中心としたものである。しかし、英語という巨大な対象物を完全に調べつくすことは簡単ではない。しかし、多くの人の少しの知識をうまく集め、体系化すればよいのではないだろうかと考える。

例文は抽出しない。

例文は別の方法で作成したコーパスの中から抽出し、訳語の検証等に利用する。

6) 集められた慣用表現についてどのように標準化するか？

7) 慣用表現と訳しわけ

8) データの収集

色々な問題はあるが、ある程度のもを集めさらに、それらが有用であるかなども考えなければならぬ。この意味からも10冊程度の簡単な本や、雑誌からデータを入力し数千件程度のものを作成した。今後、さらにデータを増やし、標準化も行いながら整理したいと考えている。

9) おわりに

対訳付慣用表現の問題点を考えながらデータの収集を開始した。今後このデータが機械翻訳システムの品種向上に大きな役割をはたすことを期待している。

「量的変化は質的变化をもたらす」という哲学の言葉を信じ、実行してゆきたい。

10) 参考文献

1. 花木金吾「英熟語ターゲット1000」旺文社
ISBN 4-01-030754-4
2. 首藤訓宏、中島武「前置詞・イディオムの英会話」-重要20前置詞で表現自在-ダイヤモンド社 ISBN 4-478-98011