

翻訳者支援のための言語レファレンス・ツール高度化方針

影浦峽（東京大学大学院教育学研究科）
佐藤理史（名古屋大学大学院工学研究科）
竹内孔一（岡山大学大学院自然科学研究科）
宇津呂武仁（京都大学大学院情報学研究科）
辻慶太（筑波大学大学院図書館情報メディア研究科）
小山照夫（国立情報学研究所人間・社会情報研究系）

1 はじめに

我々は現在、主にオンラインの文書を翻訳し、オンラインで翻訳文書を公開しているボランティア翻訳者を支援するシステムを構築している。

高精度な全自動の機械翻訳を実現することの困難さが認識されたこともあって [1]、人間の翻訳者を支援する様々なシステムが提案され、実装され、提供されている。現実にも最もよく使われているのは、TRADOSをはじめとする [2]、翻訳者向けデータ管理システムであるが、これらは、どちらかという、コンテンツよりもコンテンツを管理するメカニズムの側面で翻訳者を支援するものである。一方、高度な処理メカニズムを備え、次の訳語候補を推定して表示するシステム等も提案されているが [3]、現実の翻訳においてはほとんど使われていない。これは、一つには翻訳者のコンピュータ環境にもよるが、もう一つには、翻訳者の要求とシステムが提供するものが一致していないことにもよる [4]¹。

ほとんどの翻訳者は、翻訳の際に唯一使えるレファレンス・ツール/メカニズムは、良質の辞書と図書館（オンラインの世界では Google も図書館に類するイメージととらえられている）であると語る [4]。また、高度な翻訳支援よりも、自分の環境におけるレファレンス・ツール/メカニズムの内容および操作性の充実に第一に望んでいる。もちろん、言語処理技術の活用や計算機技術一般の活用にかかわらずしも明るわけではない翻訳者もいるため、翻訳者の現在の要望を忠実に反映させるだけでは、言語処理技術の可能性を十全に展開すればありえるかもしれない翻訳者の潜在的要望を引き出し切れないことも考えられる。しかしながら一方

¹筆者らの聞き取り・アンケートでは、機械翻訳システムや翻訳支援システムは知らない翻訳者も多く、知っている翻訳者も全員が使わないという結果であった。これは、英国における Fulford らの調査とも一致している [5, 6]。

で、人間と機械の作業領域の切り分けについては、翻訳者を支援する以上、翻訳者の能力と判断力を重視するのが妥当であり、その点から、翻訳支援システムの高度化にあたっては、翻訳者の要望を十分に考慮する必要がある。

そこで本稿では、伝統的に翻訳者が使っているレファレンス・ツール/メカニズムと参照すべき言語単位との関係を、翻訳者への聞き取りも交えて整理し、それを起点としたレファレンス・ツール/メカニズムの高度化方針を定義する。なお、本稿では、基本的に英日翻訳を想定する。したがって、英語が原言語、日本語が目標言語である。

2 翻訳者のニーズとツール

多くの翻訳者が、自分が調べたい言語単位とレファレンス・ツール/メカニズムとの関係、そして自分が苦手とする言語単位について、いくつかのクラスを識別している。翻訳者によって細かい違いはあるものの²、言語単位の観点からは、翻訳者は概ね、(1) 一般語、(2) 熟語・慣用句、(3) 固有名、(4) 専門語、(5) 引用句、(6) 簡単な連語、の 6 種類を区別している。

それらに対して、(1) どんな情報を参照するのか、(2) どんなツール/メカニズムを利用するのか、を整理したものを、表 1 に示す。ここで「図書館」としてある部分は、分野のテキストや百科事典等、個人では所有しきれない様々な資料を参照することを意味する。なお、人名辞典や専門用語辞典は、翻訳者が自分の専門分野についてはそれなりに個人で所有しているが、多くの場合、図書館にある諸種の人名辞典・専門用語辞典を参照する必要が生ずるという点では、図書館の力

²たとえばプロの翻訳者として人文系の本も多数訳出しているある翻訳者は、言語単位の分類よりも意味・表現・事実のいずれを求めるとかの分類がまずあると語った。

表 1: 言語単位と参照情報・ツール

言語単位	どんな情報を参照するか	どんなツールを参照するか
一般語	原言語の意味・対象言語の表現	辞書
熟語・慣用句	原言語の意味・対象言語の表現	辞書
固有名	対象言語の表現・事実情報	人名辞典等・図書館・Google
専門語	原言語の意味・事実情報・対象言語の表現	専門用語辞典・図書館・Google
引用句	既存の訳	図書館・Google
簡単な連語	対象言語の表現	図書館・Google

テゴリーで参照されるものにも含まれる。また、この中で、引用句と簡単な連語、とりわけ簡単な連語については、体系的な情報探索の方途がないと語る翻訳者が多かった。

個別のレファレンス・ツールと図書館（あるいはGoogle）のような総合的なレファレンス・メカニズムとは異なるが、これらを一樣なものにとらえるならば、ツールに対する翻訳者の満足度は、内容と活用（検索）機能の2点から、表2のように整理できる。

表 2: 言語単位とツールへの満足度

参照単位	ツールの内容	ツールの活用機能
一般語	○	○
熟語・慣用句	○	×
固有名	△	△
専門語	△	△
引用句	×	×
簡単な連語	×	×

一般語や熟語・慣用句については、各翻訳者がお気に入りですべて使っているいくつかの辞書で内容的には十分満足している。ただし、一般語については検索機能にも満足しているが、熟語・慣用句については、検索機能が不十分であるという状況にある。固有名と専門語については類似の状況にあり、既存の固有名辞典や専門用語辞典が充実している領域ではそれなりに使えるものの、全体としては、個別の辞典としても、入手できる辞典を集合としてみても、そもそも十分な内容を辞典が提供していないとみなされている。ツールの活用機能については、辞典レベルでは、技術的な側面よりも、むしろ、内容の不十分さから、「求めている情報があるのかないのか判断しにくい」部分が大きく影響しており、図書館における探索では、文献探索は体系的にできても、言語単位の探索は一定範囲のレファレンス・ツールでしかできないため、求める情報の発見に困難を引き起こしている。引用句や簡単な連語については、

探索そのものが体系化できないほど、ツールの拡充が立ち後れており、ほぼ翻訳者の「カン」に依存している状態と言ってよい。いずれにせよ、固有名・専門語・引用句・簡単な連語については、内容的な拡充が最重要であることを翻訳者は指摘している。熟語・慣用句については、検索機能が不十分ながらも、それなりにツールを有益に使いこなしていることから、何よりもまず、レファレンス・ツールの有効性は、内容とりわけ基本的な見出し語の十分さに依存することがわかる。

3 ツールの高度化方針

2節での整理から、ツールの高度化方針を、参照言語単位の種類ごとに定義することができる。ここでは、それを、可能な場合には翻訳者のより具体的な探索活動に重ねて技術的な方針の概略を示すかたちで整理する。

3.1 一般語

一般語の辞書については、冊子体および電子辞書に満足している結果が示されている。したがって、内容の観点からは、翻訳支援システムを構築提供する際に、既存の優れた辞書を利用できるかどうかの問題となる。これは、むしろビジネス・モデルの問題である。一方、オンライン文書の翻訳を考えるならば、「理解.com」のような自動的な辞書引き機能が考えられるが[7]、これについては、技術的な実現は容易であり、翻訳者の側からはインタフェースの使い勝手次第という声が多い。

3.2 熟語・慣用句

熟語・慣用句については、検索上の問題を指摘する翻訳者が多い。実際、比較的熟練した翻訳者でも、熟語・慣用句については相対的に苦手意識を抱いている場合が多く、頻繁に辞書を参照するが、その際、見出し語を

推定して引かなくてはならないことが大きな問題となる。実際、翻訳者は、熟語・慣用句を明確に認定する場合ばかりではなく、「この範囲のこのあたりは熟語・慣用句と考えないと意味が通じない」といったかたちで消極的に認定することも多い。

したがって、熟語・慣用句については、既存の高品質辞書のエントリーとテキストとの（異形を含めた）自動的なマッチングが大きな課題となっている。なお、このとき、少なからぬ翻訳者が指摘するのは、自動的に辞書引きができる際、ただ一つの正解を提示するのではなく、いくつかの候補を適切に提示するようなものが望ましいという点である。通常、翻訳者は、最終的には翻訳の正解とは判断されなくても正解であることの確信度を高めるために可能な解釈を複数検討する。熟練した翻訳者であればあるほど、検討した範囲が、誤訳をなくす判断に結びつくことが多いようである。こうした、翻訳者が検討する候補範囲から大きく外れない複数の候補をうまく提示することが、熟語・慣用句検索の高度化の具体的な目標になる [8]。

3.3 固有名

固有名は、第一に、レファレンス・ツールとして内容的拡充が求められている。その際の内容的拡充は、何よりも第一に、エントリー量の拡充すなわち規模の拡大である。翻訳者は、人名に関する古典的なレファレンスとしては、いくつかの人名辞典および図書館の各国概説テキストの巻末索引等を参照することが多い。オンライン上での参照は、原綴から音表記を類推して、Google を参照し、周辺情報から該当する情報かどうかを判断するのが標準的な翻訳者の手続きである。したがって、固有名辞典を拡充させるとともに、翻字規則等を用いた Google による固有名対訳検索システムを実装することが有益である [9]。組織名や地名等、人名以外の固有名については、音の対応でカタカナ表記を作る場合には、人名に準じて考えることができる。そうでない固有名の翻訳は、翻訳者による情報探索も、翻訳援助のあり方も、以下で説明する専門語に準じて考えることができる部分が多い。

3.4 専門語

専門語が置かれている状況は、固有名と似ており、何よりも、見出し語の拡大という意味での内容的拡充が求められている。翻訳者は、古典的なレファレンスとしては、いくつかの専門用語辞典および図書館でテキストの巻末索引等を参照することが多い。オンライン

上での参照は、原語から語構成要素の訳を手がかりに Google を参照するなどの手法を用いている。したがって、専門用語辞典を拡充させるとともに、ために、語構成要素の訳関係なども積極的に利用した用語収集のメカニズムを考慮することが有益である。これを実現するために必要な諸手法の開発は進められており [10, 11]、出発点として参考にすることができる。

3.5 引用句

引用句については、現れうるものの全体を探すというよりも、翻訳者が過去にどこかで見たものを探すといった探索が多い。冊子体の世界でもオンラインの世界でも同様である。これは、最初からテーマや分野で探索空間がしばられているということでもある。したがって、引用句を網羅的に収集した普遍的な図書館を作るという構想もありうるが、翻訳者が訳している文書と関連した文書ですでに訳された文書の集合をオンラインで求め、それを対象範囲として引用句を検索するシステムを実現すれば、かなりの実効性をもつであろうことが期待される [12]。オンラインで公開されている翻訳文書を自動的に収集するシステムについては、すでに実装ができています。

3.6 簡単な連語

簡単な連語とは、たとえば、“successful vote”のように、誰にでも意味はわかるが、いざ翻訳しようとするとき日本語で自然な表現を探すことが難しいものであり、翻訳者の熟練度に大きく関わっている。これについては、翻訳者のニーズは高いものの、翻訳者自身も情報探索パターンを体系化できておらず、つかみどころがないというのが実状である。引用句のように分野を限定してもそれなりの有効性を保てるような言語単位でもないし、熟語・慣用句のように基本的に語彙化された有限のものでもない。レファレンス・ツールの高度化目標としては、現在のところ、明確な定式化ができていない部分である。

4 おわりに

本稿では、翻訳者の実作業や要求を考慮しつつ、レファレンス・ツールの高度化という観点から、翻訳者を支援するために必要となる基本的な要件を整理してきた。その結果、翻訳者がそれなりに区別している (1) 一般語、(2) 熟語・慣用句、(3) 固有名、(4) 専門語、(5)

引用句、(6) 簡単な連語、の6種類のうち、(6) 簡単な連語、を除く5つのクラスについては、具体的なレファレンス・ツールの高度化方針を、実現可能な目標として定めることができることが明確になった。

我々は、現在、これらのモジュールについて分担で開発を進めると同時に、インタフェースを含めた総合的なオンライン翻訳者支援システムの構築を進めている。

なお、いささか先走りになるが、翻訳者への聞き取り調査をしていて、翻訳者が参照するデータについて、明らかになった点がある。それは、翻訳者は、宙に浮いた構造的に参考になる例を求めているのではなく、具体的な言語表現に対して、これまで具体的に特定の場所で特定のかたちで与えられた表現や情報を求めているという点である [13]。その意味で、レファレンス・ツールの基本となるのは、翻訳メモリーが提示するような構造的に似通ったあるいは一部の表現が語彙的に共通した「類例」ではなく、探索の対象となっている単位そのものについての具体的な先行存在である。既存の高品質辞書と図書館が基本的な参照の枠組みとなっているのは、まさにこれがゆえであり、Google も、一つの歴史的な世界として、「そこになければあきらめる」という位置づけがなされている限りにおいて、参照に足るものと見なされる。

この観点からは、内容の拡充が求められている固有名や専門語については、それなりに明確に規定された範囲の中で、歴史的に表現されてきた固有名や専門語を包括的にカバーすること、あるいはその範囲を前提として、その任意の下位集合ではなく、一定の有用な方針のもとでの体系的かつ一貫した語彙の収集と提示が求められることになる。この点で、レファレンス・ツールの高度化の観点から見た固有名検索や専門語収集は、個別の要素の妥当性判断にもとづく再現率と精度では評価できない性質が本質的に求められることになる。そのような性質をどう評価していくかについても、個別のモジュール開発において、より明晰に定式化して配慮して行く予定である。

謝辞

本研究の一部は、日本学術振興会科学研究費補助金基盤 (A) 「翻訳者を支援するオンライン多言語レファレンス・ツールの構築」(課題番号 17200018) の支援を得て行われた。

参考文献

- [1] Kay, M. 1997. "The proper place of men and machines in language translation," *Machine Translation*, 12(1), p. 3–23.
- [2] TRADOS. <http://www.trados.com/> 翻訳支援一般については、Bowker, L. 2002. *Computer-aided Translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press.
- [3] Civera, J. et. al. 2004. "A syntactic pattern recognition approach to computer assisted translation," Fred, A. et. al. eds. *Structural, Syntactic, and Statistical Pattern Recognition*. Berlin: Springer. p. 207–215.
- [4] 専門翻訳者・ボランティア翻訳者8名への第一著者による聞き取り。また、アンケート調査も行い12名から基本的な翻訳プロセスにおけるレファレンス・ツールの位置づけについて回答を得ている。
- [5] Fulford, H. and Granell Zafra, J. 2004. "The uptake of online tools and web-based language resources by freelance translators: implications for translator training, professional development, and research," Second International Workshop on Language Resources for Translation Work, Research and Training. p. 37–44.
- [6] Fulford, H. 2001. "Translation tools: An exploratory study for their adoption by UK freelance translators," *Machine Translation*, 16(3), p. 219–232.
- [7] 理解.com <http://www.rikai.com>
- [8] 金平昂, 豊島実和, 竹内孔一, 影浦峯. 2006. 「英語イディオムの異形を整理する」 言語処理学会第12回年次大会発表論文集.
- [9] 辻慶太, 佐藤理史, 影浦峯. 2005. 「対訳人名検索における翻字・サーチエンジンの有効性評価」 言語処理学会第11回年次大会発表論文集. p. 352–355.
- [10] 佐々木靖弘, 佐藤理史, 宇津呂武仁. 2005. 「ウェブを利用した専門用語集の自動編集」 言語処理学会第11回年次大会発表論文集. p. 895–898.
- [11] 外池昌嗣, 宇津呂武仁, 佐藤理史. 2006. 「ウェブと要素合成法を用いた専門用語訳語推定」 言語処理学会第12回年次大会発表論文集.
- [12] 品川哲也, 森辰則, 影浦峯. 2006. 「オンライン対訳文書対からのテキスト領域抽出とアライメント」 言語処理学会第12回年次大会発表論文集.
- [13] 影浦峯. 2006. 「人間の翻訳におけるコーパスの位置づけ」 言語処理学会第12回年次大会発表論文集.