

# 翻訳者向けレファレンス・ツールにおける 「包括性」概念をめぐって

影浦峽<sup>†</sup>, 阿辺川武<sup>‡</sup>

<sup>†</sup> 東京大学大学院教育学研究科

<sup>‡</sup> 国立情報学研究所連想情報学研究開発センター

## 1 はじめに

辞書を典型とする言語レファレンス・ツールの品質を左右する要因の一つに集合としての見出し語の性格(収録語数、どのような語がカバーされているか、どのような取捨選択がなされているかなど)があるが、この点を巡る研究は、定義や訳語の妥当性などと比べて進んでおらず(Pearsall 2010)、標準的な辞書学の教科書にもこの点を巡るまとまった記述は存在しない(cf. Atkins & Rundell 2008; Sterkenburg 2003; Svensen 2009)。言語処理の分野でも最近になって辞書構築における見出し語集合の性質の重要性が指摘され出したが(佐藤 2010)、十分な研究の蓄積はない。図書館情報学分野では「蔵書(本の集合)は、それ自体、本に書かれている個別コンテンツの総体を超越するものである」(Sandstrum 2010)ことが広く認められており、それに対応して辞書は個々の見出し語の説明の総体ではなく、見出し語が集合としてまとまったところに個別の情報の和を超越する価値が生まれることは理解されてきたものの、蔵書構築研究に対応する集合としての見出し語構築の研究はほとんどない。

本研究では、筆者らが開発運用しているオンライン統合翻訳支援・ホスティングサイト「みんなの翻訳」(Utiyama, et. al. 2009)の枠組みを前提に、レファレンス・ツールが提供する見出し語(あるいはそれに相当する参照単位)集合の性質に求められる要件を広い意味での「包括性」(例えば限定性が積極的意義を持つ場合はそれも一つの「包括性」と捉える。後に述べるように「規範性」の概念が積極的に関わってくる)の観点から整理する。翻訳者を支援するためのレファレンス・ツールを高度化する方針については、既に影浦ら(2006)が論じているが、そこでは既存のレファレンス情報源を前提とし、そこを出発点として相対的に高度化する方針を整理しているのみであり、本来、レファレンス情報源に求められる要件はどのようなものかを目標の概念から整理しているわけではない。本研究は、既存のレファレンス・ツールの位置づけも含め、レファレンス情報源の「包括性」を巡る要件を翻訳者

の側から理念的に整理するものである。なお、見出し語集合の性質としては他に体系性なども考え得るが、「包括性」の概念は体系性などの内部構造に関わる概念に先立って整理されるべきものと我々は考えている。

## 2 問題の所在

IR研究の布置を参照しつつレファレンス・ツールの「包括性」を巡る問いの位置づけを明らかにしておこう。基本的に、IRの研究では対象となるデータベースは所与のものと見做されその存在を出発点とする(cf. 徳永 1999)。これに対し当たり前のことだがレファレンス・ツールの「包括性」はいわばデータベースに何がどこまで収録されているかに対応する。この対応にそって考えると2つの点が直ちに明らかになる。

1. トムソン社のデータベースのように選択的であるが故にこそ社会的な意味を持つ場合がある。レファレンス・ツールについては4節で「規範性」との関係でこれに似た状況を考えるが、制限性が包括性を生み出す場合と厳密に同じ状況は、翻訳の最終基準が翻訳文書にあるため、考えにくい。
2. IRではF値が意味を持つが、これは適合文献が代替可能であることを前提としている。すなわち、いくつかの文献が検索されなくても他の文献で代替されうることを前提としている。レファレンス・ツールにおいてこの前提は成立しない。「red」が辞書で見つからなかったのに「brown」で済ませることはできない。このことは、自動専門用語抽出において、抽出された結果を一つのレファレンス・ツールとして見る観点からは、再現率、精度、F値による評価が妥当でないことを意味する。

「はじめに」でも述べたように本研究では1.のような「制限的」な状況も「包括性」概念の射程に含める。2.に関連して、「包括性」の概念が、IRの標準評価概念とは異質であることを確認しておこう。

さて、以上を踏まえて「包括性」概念を考えるわけだが、純粹に事實的・經驗的な包括性を備えたレファ

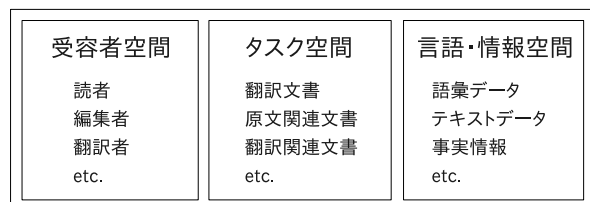


図 1. 翻訳者の意思決定に関与する空間

レンス・ツールは論理的に存在し得ない。「包括性」は従ってむしろ社会的な認識に定位する概念であるわけだが (cf. Luhmann 1973)、そのレベルでの包括性は、「そこになればそのレベルにおける情報確認行動は諦めてよい」という基準を担保する基本的性質と捉えることができる。例えば、図書館は社会的にそのような場として認識されてきたし、オンラインの世界では Google はそこで探して見つからなければ存在しないものと諦めてよいという社会的な認識を構築してきた。以下の整理はこれを基本的な視点に置く。

### 3 「包括性」に関わる空間

レファレンス・ツールの包括性を考える前提として、まず、翻訳者の正当な意思決定に関わる空間（「空間」という言葉は多少変だがとりあえず用いる）の整理を行なおう。暫定的に、関連する空間の配置を図 1 のように考えることができる。

翻訳プロセスの観点から、図 1 を補足しておこう。

1. これらの空間に対応してレファレンス・ツールやレファレンス情報源が存在する。例えば百科事典は「言語・情報空間」の中の「事実情報」のところにあり、英和辞典は「語彙データ」にある。翻訳メモリは「タスク空間」に「原文関連文書」と「翻訳関連文書」を縦断するかたちである。
2. 翻訳という行為自体は「タスク空間」に位置する。「タスク空間」は翻訳文書に応じて決まり、それに対応して、例えばもともと「語彙・情報空間」に位置づけられる情報の一部が「タスク空間」に位置づけ直されると考えることができる。
3. 最終的に翻訳は「受容者空間」における受容を先取りした意思決定行為である。「受容者空間」における社会的合意や慣例に応じて、「タスク空間」「言語・情報空間」に存在するレファレンス情報源の位置づけが定まる。
4. 「受容者空間」は未知の他者的な存在ではなく、むしろコミュニティ的なものとしてある。従って、あらかじめ固定的な「受容者空間」像を想定しそれに従った「タスク空間」「言語・情報空間」の編成をそれなりに考えることができる。

このような配置の中で、翻訳者は想定した受容をそれなりに見込むことができるよう、対応する正当な意思決定手続きを踏む。その過程で、レファレンス参照において情報源の「包括性」が正当な意思決定を支える情報源側の性質の重要な側面として関与することになる。厳密には、時代や情勢の変化のような時間軸も考慮する必要があるかも知れないが、暫定的にそれらは翻訳の時点において各空間の属性として考慮されるものと考え、ここでは扱わない。以下では、ここで述べた図式を前提としつつも、この図式を分析的に精緻化しそこからレファレンス・ツールの包括性要件を取り出すのではなく、この図式の中での翻訳プロセスを想定しつつそこから包括性要件を整理する。ただし紙幅の都合で翻訳プロセスの具体相は説明しない。また、参照情報の種類（意味を求める、表現を求めるなど）は異なる「包括性」のレベルと一定の相関関係を有するが、これについても紙幅の都合で個別に言及するにとどめ、体系的な対応付けを全面的に展開することはない。

### 4 「包括性」概念の整理

#### 4.1 基本的な分類

翻訳者にとってのレファレンス参照は、大きく「調べなくてはならない」場合（それに相当する知識を翻訳者自身が有している場合は調べなくてよいが、その場合は翻訳者が「知識を有していなくてはならない」と、「調べると好都合かもしれない」場合に区別できる。「調べなくてはならない」場合に関連するレファレンス資源は、規範性（条件付きで信頼性と言い換えることができるが真実性からくる信頼性とは異質であるため規範性という言葉がより妥当である）の概念と強く関わる。例えば、収録範囲は限定的であっても誰もが参照する専門用語対訳辞書ならば規範として参照しなくてはならない。逆に、規範的であるならば包括的でなくてもよい。従って、「調べなくてはならない」場合を巡る包括性は規範性とともに考えなくてはならない。以下の 3 つのクラスを区別することができる。

(A) 対タスク規範性／包括性：あるグループの文書を翻訳しているときに、一貫性を保つ、本来タスクに設定した具体的な基準を満たすなどの目的で絶対に参照しなくてはならない情報が存在する場合がある。例えば図書の翻訳における既訳部分の参照やアムネスティの翻訳におけるアムネスティ自身の関連既訳文書の参照など、通常は目標言語の表現に課される規範であることが多い。この場合、参照情報源の「包括性」を実現する範囲は「客観的」に決まる。ここでは限定的であることがそれ自体規範的であるとともにその規範のもとで包

括的であることを意味する。従って、翻訳支援システムはその「客観的」範囲の情報を提供しなくてはならない。

(B) 領域規範性／包括性：専門用語の翻訳に典型的であるが、翻訳においては翻訳文書が属する領域での言語運用習慣／規範を参照しそれに従うことが必要である。これもやはり目標言語の表現に課される規範である。領域規範性のもとでの包括性を現実に完全に実現することはできない。従って、受容者空間における合意の観点から、既存の探索範囲の中で相対的に優位な包括性を確保する必要がある。ただし、このクラスでは包括性において劣っていても規範性に優れた情報源を参照する必要はあるため、規範性は包括性に優先すると言ってよい。このクラスに属する参照のもう一つの例として、目標言語を用いている国が批准した国際条約の公式訳をあげることができる。この例は、理論的には「包括性」の範囲が「客観的」に決まるという点で(A)に近いが、包括的な探索範囲を具体的に保証できるとは限らないこと、また、タスクに依存するのではなくタスクとは独立に領域に依存すること、相対的に優位な包括性の実現が現実的な問題となること、という点でこのクラスに属すると考えることができる。

(C) 受容者規範性／包括性 翻訳者の観点からは一般にピアの翻訳者や編集者などが「受容者」を代表する。(A)と(B)を除いた場合、このクラスの参照は「誤訳」や「かんちがい」を解消したり「よい訳」を作るために情報源を参照する必要があるもので、概念的には「調べると好都合かもしれない」場合に近い(翻訳者の既有知識との関係が現実的に異なる場合と考えればよい)。この場合、目標言語の表現に強い規範性はいかからないため、一定の質を保った情報源ならば相対的包括性を確保した情報源の参照が要求されることになる。例えば「みんなの翻訳」で英和辞典に三省堂『グランドコンサイス英和辞典』(三省堂 2001)を用いているのはこのためである。また、テキストデータの参照においてオンライン翻訳を想定すれば Google を検索エンジンとして用いることが求められるのも、受容者側の慣例・制度を踏まえた相対的包括性が Google で実現されているためと考えることができる。

これに対し、「調べると好都合かもしれない」場合の情報源参照は、できるだけ多様かつ広範囲にわたっていればよく、その場合の「広範囲」についても、「調べなくてはならない」場合のCのような慣例・制度の縛りはなくてよい。まさに偶然見つければ儲けものとい

う探索を翻訳者ができるかどうかという、むしろ翻訳者の個別技量に関わってくると考えることができる。従って、翻訳支援システムの側から包括性を有するレファレンス情報源の構築と提供を考えるならば、「調べなくてはならない」場合を基本的に想定すればよい。

## 4.2 クラス間の関係

4.1で導入した3つのクラスのうち、(A)は規範性と包括性をタスクにおいて現実に満たせるタイプの参照、(B)は規範性と包括性をともに実現することが好ましいものの両者が同時に実現していない場合まずは規範性が優先されるタイプの参照、(C)は規範性は緩く一定の質の閾値をクリアしていれば包括性が重視されるタイプの参照である。既に述べたように、強い規範は主に目標言語の表現にかかるなど、それぞれのクラスが強い相関関係を持つ情報探索の単位と探索情報のクラスは異なるが、翻訳者の情報探索としては必ずしも重なりのないものではない。従って、どの段階(下訳・修正・見直しなど)でどの探索が行なわれるかは別に、論理的関係としては、一般に、(A)→(B)→(C)という優先順位が成り立つ。すなわち、例えば“crimes against humanity”という句のレファレンス参照において(A)で確認が取れれば、(B)と(C)を参照する必要はなく、逆に(C)を探したとしても、(B)、(A)に存在するならば、それらを参照することは必須である。翻訳における理想は、従って、すべてのレファレンス情報源が(A)のクラスで構成されることにあると言えよう。また、規範性が包括性と関係する限り、規範性のレベルを逸脱した包括性はその規範性レベルにおいては評価されない。

## 4.3 応用の設計

以上から、「包括性」概念の周辺を手がかりに翻訳者向けレファレンス情報源を充実させるための方針を整理することができる。これらは、まとめてしまうと極めて常識的かつ凡庸なものであると言える。

相対包括性優位の実現：(B)と(C)それぞれのクラスの中で(場合によっては(A)においても)、既往のレファレンス情報源の収録範囲を広げること。

規範性上昇 (C)→(B)、(B)→(A)へと、上昇先ではカバーされていなかったレファレンス情報の規範性を上げること。

これに加えて、包括性とは直接関係ないが、実際的な応用システムの設計では、環境改善、すなわち、包括性や規範性の条件が同一である場合に、レファレンス情報源への参照手続をより楽で便利なものにも、有効な応用の設計方針となりうる。

## 5 包括性と規範性

最後に、相対包括性優位の実現と規範性上昇を何に関連付けて設計することが有用か、可能か、という課題が残る。「世論操作」的な「受容者空間」の操作を別とすれば、タスク空間に応じた動的編成を考えるのか、タスクの外に存在する言語・情報空間の社会的編成に応じて考えるか、の選択であると言ってよい。既に述べたように、翻訳における理想は、いわばすべてのレファレンス情報源が(A)のクラスで構成されることにある。実際、出版翻訳において理想的な翻訳者が行なう「下調べ」は、(C)や(B)に属する情報をすべて(A)のレベルに一旦集約する操作と考えることもできよう。そして(A)のクラスはタスクに依存したものなのだから、理屈としては、タスク空間に応じた動的な応用システムを設計し開発するに越したことはないと言えそうである。

しかしながら、実際のシステム設計においては、タスクに応じたレファレンス情報資源の動的構築が技術的に可能かどうかという点に加え、上記(C)から(A)のステップにおいて人間の意思決定として翻訳者が行なう必要がある行為が存在するのか、それはどのようなものを巡る課題が残る。その点からは、相対包括性優位の実現と規範性上昇をすべて自動で行なうのではなく、とりわけ規範性上昇に相当する人間の確認プロセスを支援するシステムの設計という選択肢が当然存在する。現在のところ、システム側は候補と規範性を検証するために必要な情報を提供し、それを一定数あるいは一定資格の翻訳者が採用したときに、システム側で規範性を一段階上昇させたレファレンス情報源に組み込むというサイクルを一つの可能性として考えている。この問題は、我々が開発・運用しているオンライン翻訳者支援サイト「みんなの翻訳」に固有の課題であると同時に、オンライン翻訳支援一般の課題であり、さらに人間の意思決定が—それが「知的」なものであれそうでないものであれ人間が行なうことに社会的な信頼を担保する根拠がある場合に—関わる行為とシステムがどのように関わるかをめぐるより普遍的な課題でもある。

## 6 おわりに

「みんなの翻訳」では、これまで、レファレンス情報源の構築と提供に際し、同一クラス内の比較包括性優位の実現と使い勝手の改善に重点を置き、一定の成果をあげてきた(cf. Abekawa & Kageura 2009 など)。現在、「包括性」と「規範性」の概念を中心に、その先のレファレンス情報源構築・提供の研究を進めており、本発表はその途上での試行錯誤を概念的に整理したものである。今後は、ここで述べた枠組みを参照し

つつ、具体的な参照情報ごとに、真に有用なレファレンス情報源の(自動)構築と提供とを戦略的に進める予定である。

## 謝辞

本研究は、日本学術振興会科学研究費補助金基盤(A)「包括的な翻訳情報資源を実現する統合翻訳支援サイトの構築」(課題番号 00211152)および国立情報学研究所 2010 年度共同研究「異種情報源の特性を考慮した、実用的な専門用語対訳辞書の構築と活用」の支援を得て行なわれた。

## 参考文献

- Abekawa, T. and Kageura, K. (2009) “QRpotato: A system that exhaustively collects bilingual technical term pairs from the web,” *IUCS 2009*, p. 115–119.
- Atkins, S. & Rundell, M. (2008) *The Oxford Guide to Practical Lexicography*. Oxford: Oxford UP.
- 影浦峽・佐藤理史・竹内孔一・宇津呂武仁・辻慶太・小山照夫 (2006) 「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」言語処理学会第 12 回年次大会論文集, p. 707–710.
- Luhmann, N. (1973) *Vertrauen: ein Mechanismus der Reduktion sozialer Komplexität* [大庭健・正村俊之訳 (1990) 『信頼—社会的な複雑性の縮減メカニズム』勁草書房.]
- Pearsall, J. (2010) オックスフォード大学出版局辞書部門部長 Pearsall, J. 博士との Euralex 2010 会議における会話.
- Sandström, J. (2010). “Saving the Warburg Library”. <http://centeredlibrarian.blogspot.com/2010/09/saving-warburg-library.html>
- 佐藤理史 (2010) 「辞書自動編纂のためのテクノロジー」『言語処理技術の深化と理論・応用の新展開』シンポジウム.
- Sterkenburg, P. (2003) *A Practical Guide to Lexicography*. Amsterdam: John Benjamins.
- Svensen, B. (2009) *A Handbook of Lexicography*. Cambridge: Cambridge UP.
- 徳永健伸 (1999) 『情報検索と言語処理』東大出版会.
- Utiyama, M. et. al. (2009) “Hosting volunteer translators,” *MT Summit XII*.
- 三省堂編集所編 (2001) 『グランドコンサイス英和辞典』, 東京: 三省堂.