慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	研究データベースの英語発信におけるニューラル機械翻訳の課題
Sub Title	What improvements are required for neural machine translation in translating research databases from Japanese to English?
Author	宮田-佐藤, 由紀子(Miyata-Sato, Yukiko) 田中, 泉吏(Tanaka, Senji) 金子, 洋之(Kaneko, Hiroyuki)
Publisher	慶應義塾大学日吉紀要刊行委員会
Publication year	2022
Jtitle	慶應義塾大学日吉紀要. 自然科学 (The Hiyoshi review of natural science). No.69 (2022. 10) ,p.33- 52
JaLC DOI	
Abstract	研究用データベースを英語化した状態で世界に向けて発信する際に，機械翻訳の利用は人手翻訳における時間や予算といった課題解決が期待できるが，翻訳の質を確保するための情報は少ない。我々は，細胞行動データベースの記述データを2種類のニューラル機械翻訳（NMT）によって日本語から英語に翻訳し，文章，文，専門用語における翻訳精度を調べた。いずれのNMTでも翻訳された文章の質は良いことが母語話者による校閲で指摘された。一方，専門用語の不適切な選択や原文の単語情報が抜けおちる訳抜け，意味の取り違えなどの誤訳が高頻度で生じていた。これらの誤訳は，母語話者による校閲でも修正することはできなかった。本論文では，これらの解析結果からNMTを利用した生命科学型のデータベースの現状を評価する。また，NMTの技術的な発展への期待に加え，用語集との連携，翻訳し易い処理，英文の質チェックなどを含んだ翻訳前後の処理なども提案する。そして，機械翻訳による学術研究データベースの自動英訳化に向けて必要な要件を明らかすることを目指す。 Machine translation tools can perform translation tasks faster and at lower costs than manual translators. However, information on the quality of the translation in research databases is limited. Therefore, to assess the effectiveness of machine translation, the descriptive data in the Cell Behavior Database was translated from Japanese to English using two neural machine translation services. Subsequently, both the translations were reviewed by native speakers, who examined the accuracy of the translation in terms of sentences, phrases, and terminology. Consequently, the quality of English sentence translation was rated as good. However, several problems such as inappropriate terminology, missing words, and mistranslations were observed. Even the native speakers could not determine the intended meaning and correct these errors. Therefore, these problems must be overcome to ensure reliable translation, for which machine translation and designing quality control systems for translated data, such as pre-and post-translation processing, should be further improved.
Notes	研究ノート
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN10079809-20221030-0033

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

Hiyoshi Review of Natural Science
Keio University No. 69, 33-52（2022）

研究データベースの英語発信におけるニューラル機械翻訳の課題

宮田一佐藤由紀子*・田中泉吏**・金子洋之***

What Improvements are Required for Neural Machine Translation in Translating Research Databases from Japanese to English?

Yukiko Miyata-sato, Senji Tanaka and Hiroyuki Kaneko

要旨

研究用データベースを英語化した状態で世界に向けて発信する際に，機械翻訳の利用は人手翻訳における時間や予算といった課題解決が期待できるが，翻訳の質を確保するための情報は少ない。我々は，細胞行動データベースの記述データを2種類のニューラル機械翻訳（NMT）によって日本語から英語に翻訳し，文章，文，専門用語における翻訳精度を調べた。いずれのNMTでも翻訳された文章の質は良いことが母語話者による校閲で指摘された。一方，専門用語の不適切な選択や原文の単語情報が抜けおちる訳抜け，意味の取り違えなどの誤訳が高頻度で生じていた。これらの誤訳は，母語話者による校閲でも修正することはできなかった。本論文では，これらの解析結果からNMTを利用した生命科学型のデータベースの現状を評価する。また，NMTの技術的な発展への期待に加え，用語集との連携，翻訳し易い処理，英文の質チェックなどを含んだ翻訳前後の処理なども提案する。そして，機械翻訳による学術研究データベースの自動英訳化に向けて必要な要件を明らかすることを目指す。

キーワード：　ニューラル機械翻訳，細胞行動データベース，研究データベース，生命科学，日英翻訳

*　慶應義塾大学自然科学研究教育センター（〒223-8521　横浜市港北区日吉4-1-1）：Research and education center for natural sciences, Keio University, 4-1-1, Hiyoshi, Kohoku-ku, Yokohama 223-8521, Japan. E-mail: yukiko_s@keio.jp
**　慶應義塾大学文学部（〒108-8345　東京都港区三田2-15-45）：Faculty of Letters, Keio University, 2-15-45, Mita, Minato-ku, Tokyo, 108-8345, Japan. E-mail: senji@keio.jp
***　慶應義塾大学自然科学研究教育センター（〒223-852　横浜市港北区日吉4-1-1）：Research and education center for natural sciences, Keio University, 4-1-1, Hiyoshi, Kohoku-ku, Yokohama 223-8521, Japan. E-mail: theband@z3.keio.jp

この著作物は著作権法によって保護されています／This content is protected by the Japanese Copyright Act

Summary――Machine translation tools can perform translation tasks faster and at lower costs than manual translators. However, information on the quality of the translation in research databases is limited. Therefore, to assess the effectiveness of machine translation, the descriptive data in the Cell Behavior Database was translated from Japanese to English using two neural machine translation services. Subsequently, both the translations were reviewed by native speakers, who examined the accuracy of the translation in terms of sentences, phrases, and terminology. Consequently, the quality of English sentence translation was rated as good. However, several problems such as inappropriate terminology, missing words, and mistranslations were observed. Even the native speakers could not determine the intended meaning and correct these errors. Therefore, these problems must be overcome to ensure reliable translation, for which machine translation and designing quality control systems for translated data, such as pre-and post-translation processing, should be further improved.

Keywords: Neural machine translation, Cell behavior database, Research database, Life science, Japanese-English translation

1. 背景

近年，多くの研究分野において世界中の研究者がデータベース（DB）を利用して，日々の研究を進めている。生命科学の分野では，米国主導による NCBI（統合 DB）[1], Pubmed（文献検索）[2]，BLAST（ホモロジー検索）[3] など，多様な研究支援データベースが共通言語である英語で発信されている。日本からも，KEGG（パスウェイ DB）[4] や DDBJ（塩基配列 DB）[5] など大規模プロジェクトのもとで開発された英語ベースのデータベースが発信されている。一方，比較的中規模のプロジェクトのもとで開発されたデータベースや研究室内部のデータをまとめたデータベースも存在し，これらは NDBC（https://biosciencedbc.jp）などの総合ポータルサイトを通じて広く利用可能となっている。

ただし，比較的中規模，また専門的な内容のデータベースは，日本語のみで発信されているものも多く，使用範囲が日本語を理解できる研究者に限られている。本来ならばデータベースを構築した研究者が英語化できることが望ましいが，その作業は多くの時間や費用を要することから，なかなか実現が難しいのが現状である。我々の研究グループでも，さまざまな細胞行動を収集したデータベース（Cell Behavior Database: CBDB）を構築しているが [6]，英語化までには至っていない。現実問題として，CBDB は単語や文章記載型であり，英語化へ向けては数十万語を翻訳する大規模プロジェクトとならざるを得ない。

近年，ニューラルネットワーク技術の発展に伴い，さまざまなニューラル機械翻訳（NMT）が開発されている。2016 年の Google の自然言語プログラムの発表 [7] 以降，その翻訳精度は日進月歩で上がっている。NMT を利用することで，人手による翻訳でかかる金銭的コストと時間的コストを大幅に減少できる [8]。これらの NMT の中には Application Programming Interface（API）を公開しているものもある。API を利用することで，Web 上の記事を自動

で翻訳することが可能となる。実際，多くの民間企業や個人のホームページでは，NMT の API を利用し，オリジナルの文章を翻訳して，多様な言語で発信している。現在日本語で公開されている研究データベースも，NMT の API を利用することで，英語や他の言語に自動変換させて発信できる可能性は極めて高い。

このような状況の下，生命科学の研究現場を鑑みると，NMT による翻訳の自動化は進んでいない。大きな理由の一つは，NMT に対する信頼性である。例えば，現段階で，英語から日本語への翻訳精度が上がっていることはある程度認知されてきているが，日本語から英語への翻訳精度に関する情報はほとんど無く，使用に耐えうるとのコンセンサスは形成されていない。また，研究データベースは，記述内容の誤訳による影響が極めて大きいにもかかわらず，専門用語などの誤訳の発生率についても不明である。

本研究では研究用データベースの自動翻訳に向けて必要となる NMT の要件を調査した。API を一般公開している 2 種類の NMT で翻訳し，単語，文，文章レベルでその精度や誤訳の傾向を解析した。その結果，現段階では NMT のみで研究データベースを自動英訳するのは未だ現実的ではなく，専門用語の翻訳精度の向上のみならず，文の誤訳を無くすことが課題であることが見えてきた。用語集との連携，翻訳しやすい処理，英文の質チェックなどを含んだ翻訳前後のシステムを運用することを提案したい。

2. 材料と方法

慶應義塾大学の細胞行動データベース（Cell Behavior Database: CBDB）について検証した。データ番号 1–85，87–100，102–110 のデータ（補足資料 1–16）を，2 種類のニューラル機械翻訳システム（Google Neural Machine Translation; GNMT［9］と DeepL’s NMT system; DeepL［10］）で日本語から英語に翻訳した（2022 年 2 月 6 日）。次に英語を母語とする話者（株式会社 Editage）にネイティブチェックを依頼して翻訳された英語文章の違和感や文法上の誤りを調べた。また，筆者らによって翻訳された英語データを，日本語の原文と比較しながら誤訳がないかを目視で調べた。

3. 結果

3–1　細胞行動データベース

最初に，本調査で使用した細胞行動データベース（Cell Behavior Database: CBDB）の概要を示す。CBDB は，「細胞がある一定の条件下でどういった行動を取るか」というフォーマットで細胞行動を収集し，高次生命現象を理解することを目的としている［11］。

図 1 は CBDB 内のデータ例を示している。データは行動の主体となる「細胞・組織・器官」，主体の状況（「生物名」と「発生ステージ」），主体に与えられる条件（～の状況下で，以降「入力」と表記する），主体がとる行動（～をする，以降「出力」と表記する），該当する細胞行動

ID	0001
現象名	食作用
細胞・組織・器官	ケラチノサイト
生物種	ヒョウガエル
N/E	N
発生ステージ	成体
～の状況下で	表皮域のメラノサイトに接する
～をする	メラノサイトの樹状突起を食べる
説明	動物の皮膚には、メラノサイトと呼ばれる黒色の色素細胞が分布している。真皮域のメラノサイトは、自身の細胞内でのメラノソームの凝集と拡散により、体色変化を引き起こす。一方、表皮域でのメラノサイトはその樹状突起を周りのケラチノサイトに食作用させることにより皮膚を着色する。日焼けによる皮膚の黒化も同様のメカニズムで生じる。図は、ヒョウガエルの脊側の皮膚において、メラノソームが、樹状突起を伸ばしたメラノサイトからケラチノサイトに移動していることを示している。この状態は表皮性メラニン単位と呼ばれる。
文献	小比賀正敬 1982.「色素細胞－この特異な集団」 講談社サイエンティフィク, pp. 44-47.

図１　細胞行動データベースのデータの構造

を示す図表とその解説文（「説明」），細胞行動と関連する生命現象（「現象名」）の７つの項目から構成され，それぞれの項目には該当する記述データが格納されている。

各記述データにおいて，細胞，生物名，発生ステージ，現象名は専門用語（単語），入力と出力は文，説明は文章のデータで記載されている。

3-2　文章（説明文）の翻訳

まず，CBDB における「説明」の項目について調査した。「説明」は，各データに含まれる細胞行動を比較的長い文章（100 単語程度）で記述したものである。できるだけ異なった内容の文章を 20 件（補足資料 14–16）選んだ後，2 種のニューラル機械翻訳（NMT）システム（GNMT と DeepL）で翻訳した。次に英語を母語とする話者にネイティブチェックを依頼し，翻訳文のみを読んだ際の文章の違和感や文法上の誤りなどを確認した。GNMT および DeepL いずれの NMT においても，20 件の文章中，言語，文法，明瞭さ，読みやすさに関する修正が 12 箇所あった。しかし，校閲者のコメントは，“Please note that owing to the original language being good, this document did not require extensive revisions” であり，これらの修正箇所は小さな訂正項目で，NMT による翻訳文は母語話者にとって，十分に意味の通じるものであることがわかった。

表１　データベースの文章に対するニューラル機械翻訳の誤訳の例

原文
動物の皮膚には、メラノサイトと呼ばれる黒色の色素細胞が分布している。真皮域のメラノサイトは、自身の細胞内でのメラノソームの凝集と拡散により、体色変化を引き起こす。一方、表皮域でのメラノサイトはその樹状突起を周りのケラチノサイトに食作用させることにより皮膚を着色する。日焼けによる皮膚の黒化も同様のメカニズムで生じる。図は、ヒョウガエルの**脊側の皮膚**において、メラノソームが、樹状突起を伸ばしたメラノサイトからケラチノサイトに移動していることを示している。この状態は表皮性メラニン単位と呼ばれる。
GNMT
Black pigment cells called melanocytes are distributed on the skin of animals. Melanocytes in the dermis cause body color changes by agglutination and diffusion of melanosomes in their cells. On the other hand, melanocytes in the epidermal region color the skin by phagocytosis of the dendrites on the surrounding keratinocytes. Blackening of the skin due to sunburn occurs by a similar mechanism. The figure shows that in the **spinal skin** of the northern leopard frog, melanosomes are migrating from melanocytes with extended dendrites to keratinocytes. This condition is called the epidermal melanin unit.
DeepL
Black pigmented cells called melanocytes are distributed in the skin of animals. Melanocytes in the dermis cause body color change by aggregation and diffusion of melanosomes within their own cells. On the other hand, melanocytes in the epidermis color the skin by phagocytosing the surrounding keratinocytes with their dendrites. The darkening of the skin due to sunburn occurs by the same mechanism. The figure shows that melanosomes migrate from melanocytes with extended dendrites to keratinocytes in the skin of the **spinal side** of the leopard frog. This condition is called epidermal melanin unit.

このことから，専門的な内容の日本語の文章であっても，今回使用した２種類の NMT は十分に意味が通じる英文を作成できること，ただし，英語を母語とする話者にとっては文法や自然さに関して多少の違和感が残る場合があることが示唆された。

3–3　NMT の正確さ

ネイティブチェックはあくまで翻訳文の文法上の正しさの確認であり，NMT が原文の内容を正しく翻訳できているかまでは確認できない。そこで，NMT に翻訳された英文を日本語の原文と目視で比較し，誤訳がないかを調べた。

GNMT および DeepL のいずれにおいても，20 件中 18 件の文章において何らかの誤訳が認められた。9 割の文章に誤りが発生していたことになる。1 つのデータに複数の誤訳が認められた場合もあり，これまでに 30 箇所の誤訳をみつけている。適切な専門用語に翻訳できていないケースが最も多く（表１太字），原文の要素が含まれない訳抜けや逆に原文にない要素が追加されたケース，原文の意味を取り違えていたケースなどが認められた（表１下線部）。これら誤訳された箇所のうち，ネイティブチェックによって指摘された改善箇所と共通していたのは１箇所だけであった。

このように，現在の NMT で研究データベースを翻訳すると，専門用語を中心に高頻度で誤訳が生じること，これらの誤訳に対して英語が母語の話者でも間違いに気がつくことが困難であることが示唆された。

3–4　専門用語（単語）の翻訳

一口に専門用語といってもさまざまな種類がある。そこで，専門用語の種類に応じて NMT の精度を調べることとした。前述したように CBDB では，細胞・組織・器官，生物名，発生表記，および，生命現象名に分かれて専門用語が分類されている（図 1）。まず，108 件のデータから専門用語を抜き出し，翻訳をした。重複している語句の訳については，同じ語句はすべて同じ英語に訳されていたので，1 つにまとめて整理した。最終的に，細胞名 69 種類，生物名 49 種類，発生ステージ 30 種類，現象名 71 種類で，合計 219 種類の専門用語について解析した。

表 2 は，誤訳された専門用語と NMT システムの関係を示している。いずれの項目においても，NMT システムの種類に関係なく誤訳される語句があった一方で，どちらか一方の NMT システムのときのみ誤訳される語句もあった。各項目の翻訳精度を評価するため，常に正確に訳された語句を 1 点，片方の翻訳ツールでのみ正確に訳された語句を 0.5 点，どちらのツールでも誤訳された語句を 0 点としてその平均点を計算した。その結果，生物名が 0.76 点で最も高く，細胞名と現象名は同程度の精度（0.68 点）で，発生ステージは 0.42 点と低かった。

このように，生物系の専門用語と一口にいっても，項目によって精度に違いがあることがわかった。発生ステージの表記のように，一般的な文章ではあまり使用されない特殊な語句に関しては，適切な語句の選択が極めて難しいことが示唆された。一方，一般的な文章でしばしば認められる生物名，細胞名，生命現象に関する専門用語については，比較的正確に翻訳されていた。しかし，完璧ではなく，しばしば不適切な語句が選択されることも見出された。

3–5　主語が省略された文の翻訳

最後に，CBDB における 2 つの項目，「入力」と「出力」について翻訳精度を調べた。本データベースにおける入力と出力は，環境からどのような刺激（入力）が与えられると，細胞はどのような行動（出力）を起こすのか，という関係性に基づいている。それぞれのデータは文で表現されるが，しばしば主語が省略される。これは，日本語の特徴として主語の省略が行われること，主語である細胞名は別項目で記載されているため，データベースを見れば主語が明確であることが挙げられる。したがって，英訳の際に「何に対して入力したのか？」と「何が出力されたのか？」という点が，この項目の文だけからでは判断が難しい。これは，CBDB に限らず，項目分けして表記するデータベースにおいてはよく見られる課題のひとつであると推測できる。

図 2 は，「入力」および「出力」それぞれにおける誤訳されたデータの割合を示している。GNMT は入力データの 29％と出力データの 15％に，DeepL では入力データの 14％と出力データの 16％に間違いが見られた。それぞれの文における誤訳のパターンを調べたところ（表 3），専門用語の間違いが最も多かった（GNMT 32 箇所，DeepL 22 箇所）。次によく見られたのが訳抜けで，主語を細胞ではなく別のものに設定した，文章の一部を訳さなかった，文の意味そのものが変更されたといったことを確認できた。さらに，動詞を名詞化する，学術的ではない

表２　誤訳された専門用語

細胞、組織、器官

GNMT およびDeepL	GNMT	DeepL
予定角膜内皮細胞	鱗の黒色素胞	メラノフォア
大型黒色素胞　（マクロメラノフォア）	背側尾部の皮膚の黒色素胞	大孔
神経冠（提）細胞	リンパ球様細胞	胚細胞 大孔
予定体腔嚢細胞	神経提（冠）細胞	胃の前駆細胞
神経冠 (提) 細胞	黒色素胞	内胚葉上皮
前側の割球 割球	原腸の細胞	外胚葉上皮
胚盤葉後部の上層の細胞群	砂嚢上皮	臍帯静脈の血管内皮細胞
予定中胚葉細胞	脊索	線維芽細胞様細胞
予定内胚葉細胞		体腔細胞、血球
予定原口背唇部直下の割球		後方部の脊索中胚葉

生物種

GNMT およびDeepL	GNMT	DeepL
イトマキヒトデ	カイコ	Volvox carteri の Gls/Reg 変異体
キイロタマホコリカビ	メダカ	Volvox carteri の Pcd-K1 および Sls-S1 変異体
マボヤ	メダカ（野生型）	Volvox carteri の S16 温度感受性変異体
群体ボヤ		イトマキヒトデ
		ショウジョウバエ
		スズガエル
		タイラギ（二枚貝）
		ホシムシ
		マウス
		マボヤ
		石灰カイメン

発生ステージ

GNMT およびDeepL	GNMT	DeepL
神経筋接合の形成時期		発生ステージ
11 日胚 初期胞胚期-後期原腸胚期	胎児 4 日胚から 7 日胚	原腸胚 幼生
後期原腸胚期	細胞集合期	オタマジャクシ 幼生
神経胚	胚（外鰓期）	受精直後　（背腹軸決定前）
原腸胚後期-ビピンナリア幼生 原腸胚後期	胚期	
卵割期		
原腸陥入期		
原始線条期、		
原条形成期		
原始線条期（胚内陥入時）		
胞胚		

現象名

GNMT およびDeepL	GNMT	DeepL
形態形成と基底層	食作用	生体防御
骨片形成	脊索形成	原始線条（原条）の退縮
細胞の接着力	細胞配列パターン	原腸の伸長
基底層	基質上での移動	原腸の誘導
包囲化	再生	食作用
多核状態の解消	細胞と力	脊索形成
体腔嚢形成	細胞の極性化	胚の再構築
大孔の運動	背地適応	肛門形成
培養基質上での伸展	サイズ依存性	
	軸形成	
	多核体形成	
	培養基質の形状	
	培養基質への接着	
	培養基質への反応	
	培養基質上での移動	
	培養基質上での動態	

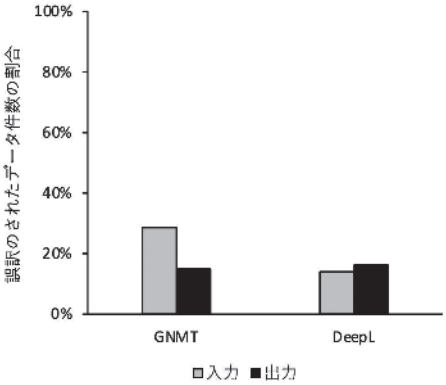

図 2　ニューラル機械翻訳における文の翻訳精度
細胞行動データベースの「入力」と「出力」の記述データを GNMT および DeepL で翻訳し，誤訳されたデータ件数の割合を示した。

表現，語句の省略，時間の表現が正確ではないなどの間違いも確認された（GNMT 9 箇所，DeepL 7 箇所）。

以上，文の翻訳でも文章の翻訳と同様に専門用語の間違いが多かった。訳抜けも散見され，文章の翻訳では見られなかった主語の取り違えも確認された。面白いことに，訳抜けする文は NMT システムによって異なっており，それぞれのシステムのアルゴリズムに依存した現象であることが示唆された。

4. 考察

本研究の目的は，学術研究データベースの自動英訳化に向けてニューラル機械翻訳（NMT）に必要な要件を明らかにすることであった。今回，多様な細胞行動がデータベース化されている細胞行動データベース（CBDB）内のデータを材料として，日本語から英語に NMT システムで翻訳された語句や文章の正当性を解析した。その結果，翻訳された CBDB の文章は英語を母語とする話者にとって意味のある文章として認識されることが示された。一方，原文からの翻訳の過程で，専門用語の誤訳や内容を誤解した訳文の作成などいくつかの問題があることも見出された。これらの結果から，現行の NMT によって研究用データベースを自動翻訳して発信するのは時期尚早であること，専門用語の翻訳精度の向上と文の誤訳を無くすことが特に重要であることが示唆された。以下，3 項目の視点から，機械翻訳ツールを取り込んだ英語化データベースの実現性について考察する。

4–1　NMT による誤訳

NMT は 2014 年ごろから主流になった機械翻訳技術で，文章の流暢性が格段に上がったこ

表3　入力と出力におけるニューラル機械翻訳の誤訳

		専門用語	訳抜け				そのほか			
		適切な用語が選択されなかった	語句が省略された	主語が別のものに設定された	内容が別の意味になった	ごまかした訳（精度が低い）	直訳	表現が学術的でなかった	時間表現	名詞化
入力	GNMT および DeepL	4	0	0	0	0	1	0	1	0
	GNMT	17	0	3	1	2	0	0	0	0
	Deep L	5	0	0	4	0	0	0	0	0
出力	GNMT および DeepL	7	0	0	0	0	0	1	0	0
	GNMT	4	0	1	1	1	1	1	0	0
	Deep L	6	1	1	1	0	0	0	0	1
合計	GNMT および DeepL	11	0	0	0	0	1	1	1	0
	GNMT	21	0	4	2	3	1	1	0	0
	Deep L	11	1	1	5	0	0	0	0	1

とが特徴のひとつである。一方で，従来の統計型機械翻訳と比較して，NMT は不適切な用語の選択や訳抜けが生じやすいことが指摘されている。現状を垣間見ても，多くの専門用語を含む特許翻訳の分野ではすでに特許庁による NMT サービスが活用されているが，誤訳をいかに少なく抑えるかということが今後の課題となっている［12］。同様に海外でも，別の NMT で英訳された公開特許データベースに専門用語の誤訳がしばしば含まれることが指摘されている［13］。

今回の研究では，自然科学系のデータベースである CBDB が GNMT と DeepL という 2 種類の NMT で英訳されているのだが，専門用語の誤訳（表 2）や，原文にある単語情報が抜け落ちてしまう訳抜けや原文にない単語情報が意図せずに訳出されてしまう湧き出しなどが確認された（表 3）。また，日本語の特徴に主語の省略があり，主語の省略された文は英訳された際に実際とは異なる主語が意訳によって生じ，データの意味そのものが変更されてしまうことも見出された。これらの結果は，現状の NMT で自然科学系の研究データベースを翻訳した場合，専門用語の誤訳や意味の取り違えが生じうる懸念を生じさせる。研究用データベースは，他の科学者がデータを利用して研究を発展させるため，その精度の高さはデータの信頼性と直結する。仮に，誤訳によって誤ったデータが公開されると，そのデータをもとに研究が進められてしまうなど極めて大きな悪影響が生じる。

4–2　母語話者による翻訳文への評価

前述したように NMT の技術的発展に従い訳文の流暢さは大きく向上したが，皮肉なことに，訳の自然さに起因して間違いに気が付きにくくなってしまう可能性が指摘されている［12］。本研究でも，翻訳文に対する英語を母語とした校閲者の評価は高く，日本語から英語への翻訳文はかなり自然に感じられるようになってきていると考えられる。しかしながら，校閲された内容は文法や明瞭さの向上が中心であり，誤訳に対する修正はほとんどなされなかった。もし現行の NMT で英語化されたデータベースがそのまま発信されてしまった場合，データの中に誤訳に基づく重大な誤りがあったとしても，使用者が気付くことなく誤りのあるデータを使用

してしまう危険性が生じる。

4–3　翻訳前後の処理の重要性

NMT の誤訳の問題に対し様々な研究が行われている。専門用語の間違いは，NMT が機械学習の過程で使用する辞書の用語数が制限されることが理由のひとつと考えられている。機械学習に使用する辞書の用語を増やすことは難しいため，専門用語集と連携させることが行われ，一定の効果をあげている［14］，［15］。生命科学の領域では，日英他医薬を収めているライフサイエンス辞書（LSD）［16］などと連携することで，NMT の弱点を補い専門用語の翻訳の質の向上が期待される。

一方，今後データを作成する場合は，原文となる日本語をできるだけ翻訳しやすい形に仕上げることも有効かもしれない。例えば，日本語の特徴のひとつに主語の省略がある。文章の場合は前後関係から主語の推測ができるが，文だけの場合は正確な翻訳が難しく，今回の CBDB の英語化においてもしばしば主語の取り違えが起こることが見出された。原文に主語を明記するなど翻訳しやすい文章にあらかじめ加工しておくことで，こういった問題は解消すると期待できる。また，BLEU スコア［17］や RIBES 法［18］などの翻訳の自動評価法により，発信前に英文のクオリティをスクリーニングすることも有効であると思われる。以上，NMT のさらなる発展・改良に加え，用語集との連携や翻訳前後の処理によって，信頼性の高い英訳データを発信していくことが可能になっていくと期待できる。

5. 結論

われわれは，本研究について幾つかの限界があることを認識している。1 つ目は，研究対象が CBDB のみであり（CBDB は多岐にわたる生命現象を集めており，少ないデータでも語句の網羅性は高いと判断されるが），語句や文章の使用ケースが限られていることである。2 つ目は，今回使用した GNMT と DeepL 以外の NMT が未試行な点である。GNMT と DeepL は機械翻訳の研究において比較検討されることが極めて多い NMT であり，評価も高く，データベースの自動翻訳においても現実的な選択肢となる可能性が極めて高いことから，今回はこの 2 つを評価検討した。3 つ目は，翻訳前後の処理によりどの程度翻訳精度が高くなるのかが未知数なことである。このような課題を積み残してはいるが，研究用データベースを国際的に発信する上で，NMT は強力な味方となりうる。NMT のデータベースへの実装においては，今回の研究からも見出されたように，NMT の弱点である専門用語の不適切な選択と文の誤訳を解決しなくてはならない。データベースの価値とも換言できるデータ品質の維持を担保すべく，機械翻訳技術の今後の発展に期待すると同時に，用語集との連携，翻訳し易い処理，英文の質チェックなどを加えることを新たに提案したい。

謝辞

本研究は，慶應義塾大学学事振興資金および慶應義塾大学自然科学研究教育センター研究プロジェクト費による助成を受けたものである。

引用文献

[1]　E. W. Sayers et al., “Database resources of the national center for biotechnology information,” Nucleic Acids Res., vol. 39, no. suppl_ 1, pp. D38–D51, 2010.

[2]　E. Motschall and Y. Falck-Ytter, “Searching the MEDLINE literature database through PubMed: a short guide,” Oncol. Res. Treat., vol. 28, no. 10, pp. 517–522, 2005.

[3]　G. M. Boratyn et al., “BLAST: a more efficient report with usability improvements,” Nucleic Acids Res., vol. 41, no. W1, pp. W29–W33, 2013.

[4]　M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, and M. Tanabe, “KEGG: integrating viruses and cellular organisms.,” Nucleic Acids Res., vol. 49, no. D1, pp. D545–D551, Jan. 2021.

[5]　A. Fukuda, Y. Kodama, J. Mashima, T. Fujisawa, and O. Ogasawara, “DDBJ update: streamlining submission and access of human data.,” Nucleic Acids Res., vol. 49, no. D1, pp. D71–D75, Jan. 2021.

[6]　“Cell behavior Database 細胞行動データベース .” 慶應義塾大学，http://cmdb.keio.ac.jp/cbdb/.（現在限定公開のため，使用を希望される方は著者までご連絡ください）

[7]　Y. Wu et al., “Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation,” 2016.

[8]　中澤敏明，“機械翻訳の新しいパラダイム：ニューラル機械翻訳の原理，” 情報管理，vol. 60，no. 5，pp. 299–306，2017.

[9]　“Google Translate. Available online: https://translate.google.co.jp (accessed on 6 Feb 2022).”.

[10]　“DeepL. Available online: https://www.deepl.com/translator (accessed on 6 Feb 2022).”.

[11]　岩爪道昭，団まりな，金子洋之，浅利今朝男，“細胞行動データベース 生物界の意味論と自然言語の意味論のマッピングを通して，” in 人工知能学会全国大会論文集 第 20 回（2006），2006，p. 81.

[12]　園尾聡，“ニューラル機械翻訳による特許機械翻訳システムの開発，” Japio year B., vol. 2019, pp. 296–301, 2019.

［13］ 田畑文也，“中国特許翻訳情報の検証，”情報プロフェッショナルシンポジウム予稿集，vol. 2021，pp. 7–12，2021.

［14］ M.-T. Luong, I. Sutskever, Q. V Le, O. Vinyals, and W. Zaremba, “Addressing the rare word problem in neural machine translation,” arXiv Prepr. arXiv1410.8206, 2014.

［15］ 山本ゆうじ，“シンプルな用語集形式 UTX――ニューラル機械翻訳時代の用語管理の意義，”情報の科学と技術，vol. 69，no. 9，pp. 415–420，2019.

［16］ 金子周司 et al.，“ライフサイエンス辞書から生命科学オントロジーへ，”情報知識学会誌，vol. 15，no. 4，pp. 1–10，2005.

［17］ K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “Bleu: a method for automatic evaluation of machine translation,” in Proceedings of the 40 th annual meeting of the Association for Computational Linguistics, 2002, pp. 311–318.

［18］ 平尾努，磯崎秀樹，須藤克仁，塚田元，永田昌明，“語順の相関に基づく機械翻訳の自動評価法，”自然言語処理，vol. 21，no. 3，pp. 421–444，2014.

補足資料 1　細胞行動データ（説明以外）　ID 1–9

ID	細胞・組織・器官名	生物名	発生ステージ	入力	出力	生命現象
1	ケラチノサイト	ヒョウガエル	成体	表皮域のメラノサイトに接する	メラノサイトの樹状突起を食べる	食作用
2	メラノフォア	メダカ	成体	メラノフォア刺激ホルモン（MSH）で処理される	メラノソームを拡散させる	体色変化
3	予定角膜内皮細胞	ニワトリ	4 日胚から 7 日胚	角膜基質上に進入して角膜基質を覆う	単層上皮の形態をそなえた角膜内皮をつくる	上皮構造形成
4	アメーバ細胞	細胞性粘菌 Dictyostelium discoideum	細胞集合期	カウンチン 2 タンパク質の濃度を感じる	集合を続ける（子実体形成を抑制する）	集団サイズの決定
5	筋肉細胞	ショウジョウバエ Drosophila melanogaster	神経筋接合の形成時期	神経が接近してくる	エンドサイトーシスにより細胞表面タンパクの組成を変える	エンドサイトーシス，神経筋接合
6	筋肉細胞	ショウジョウバエ Drosophila melanogaster	神経筋接合の形成時期	神経が接近してくる	細胞表面からミオポディアをのばして迎えに行く	細胞の変形，神経筋接合
7	大孔 Osculum	Sponge Ephydatia fluviatilis	Adult	筒状に伸びた状態で基部に中程度の電気刺激を受ける	刺激された部位での急激な収縮に続き，約 15 分間かけて，全体を縦横に収縮させる（図：b, c）。	大孔の運動
8	肺原基	マウス	11 日胚	浸透圧が高い外液中で培養される	より多く分枝する	分枝形成
9	変形細胞, sclerocyte, amoebocyte	石灰カイメン Calcareous sponge Cycon ciliatum	Adult	骨片を作ろうとする	数匹が集まり，septate junction で結合して閉じた袋を作る	細胞間結合 Cell-cell junction

補足資料 2　細胞行動データ（説明以外）　ID 10–16

ID	細胞・組織・器官名	生物名	発生ステージ	入力	出力	生命現象
10	変形細胞 sclerocyte, amoebocyte	石灰カイメン Calcareous sponge Cycon ciliatum	Adult	数匹が集まって septate junction で結合して，閉じた袋を作る	その袋の内部に向かって，骨片を包み込むタンパク質の鞘と，骨片の主成分である炭酸カルシウムを分泌する	骨片形成 spicule formation
11	液胞細胞 vacuolated cell （体腔細胞，血球）	マボヤ Halocynthia roretzi	成体	培養基質上に置かれる	シャクトリムシ様の運動方法により，平均速度 17.4 μm/min で動き回る	基質上での移動
12	鱗の黒色素胞	メダカ（野生型）	成体	メダカ（野生型）が黒い背景に 30 日間適応させられた後，白い背景に 30 日間適応させられる	細胞死（アポトーシス）を引き起こし，徐々に消滅する	細胞死（アポトーシス），背地適応
13	背側尾部の皮膚の黒色素胞	スズガエル	胚（外鰓期）–幼生，オタマジャクシ	受精後 6 日目には，ランダムに配列している	発生が進むにつれて，多数が直角方向に樹状突起を出しあって，規則的なパターン（網目構造）を形成する	紋様形成
14	顆粒細胞（血球）	カイコ	Larval stage	体腔内に注入された直径 25.7 μm のラテックスビーズに触れる	多くの細胞で，1 つのビーズを被い包む	生体防御，包囲化
15	卵母細胞	アフリカツメガエル	卵母細胞期	UV を照射された後，成熟，受精，発生する	軸構造を形成せずに，腹側化する	背側決定因子
16	多極性細胞 multipolar cell （体腔細胞，血球）	マボヤ Halocynthia roretzi	成体	培養基質上に置かれる	多数の糸状仮足を真っすぐに延ばして，接着する	培養基質への接着

補足資料 3　細胞行動データ（説明以外）　ID 17–23

ID	細胞・組織・器官名	生物名	発生ステージ	入力	出力	生命現象
17	リンパ球様細胞 lymphoid cell（体腔細胞，血球）	マボヤ Halocynthia roretzi	成体	培養基上に置かれる	アメーバ運動で基質上を動き回る	培養基質上での移動
18	大孔 Osculum	Sponge Ephydatia fluviatilis	Adult	静水中で筒状に伸びていたものが，かなり強い水流下におかれる	20 分ほどかけて縮んで，平たくなる	大孔の運動
19	Phago-amoebocyte（Coelomocyte, Hemocyte）	Halocynthia roretzi. Ascidian	Adult	Placed on cultural substratum	Migrate randomly by amoeboid movement at an average speed of 33.0 μm/min	Migration on cultural substratum
20	融合細胞 fusogenic phagocyte（体腔細胞，血球）	マボヤ Halocynthia roretzi	成体	培養基上に置かれる	薄い円盤状に伸展して，基質に接着する	培養基質への接着，細胞移動
21	リンパ球様細胞 lymphoid cell（体腔細胞，血球）	マボヤ Halocynthia roretzi	成体	培養基質上に置かれる	柔らかい糸状仮足を延ばして，接着する	培養基質への接着
22	変形細胞, sclerocyte, amoebocyte	海綿動物 Sponge	Adult	数匹が集まって閉じた袋を作る	協同して，種に固有の，特定の形の骨片を形成する	骨片形成 spicule formation
23	胚細胞	イトマキヒトデ Asterina pectinifera	初期胞胚期 - 後期原腸胚期	個々に解離されて海水中に浮遊した状態で，他の細胞と出会う	互いに接着して凝集塊を作る	胚の再構築，細胞の凝集

補足資料 4　細胞行動データ（説明以外）　ID 24–31

ID	細胞・組織・器官名	生物名	発生ステージ	入力	出力	生命現象
24	外胚葉細胞	イトマキヒトデ Asterina pectinifera	初期胞胚期 -後期原腸胚期	バラバラに解離された後に，凝集塊に組み込まれる	凝集塊の外縁に移動し，そこで他の外胚葉細胞と出会って上皮を形成する	胚の再構築，細胞選別
25	内胚葉細胞	イトマキヒトデ Asterina pectinifera	初期胞胚期 -後期原腸胚期	バラバラに解離された後に，凝集塊に組み込まれる	他の内胚葉細胞と出会って，凝集塊の内部で，頂部を内側に向けた上皮を形成する	胚の再構築，細胞選別
26	大孔 Osculum	Sponge Ephydatia fluviatilis	Adult	流水中で縮んで平たくなっていた状態で流水が止む	30 分以内に伸びて筒状になる	大孔の運動
27	大孔 Osculum	Sponge Ephydatia fluviatilis	Adult	筒状に伸びた状態で側面に電気刺激を受ける	刺激された側に湾曲する	大孔の運動
28	胃の前駆細胞	イトマキヒトデ Asterina pectinifera	後期原腸胚期	バラバラに解離された後，再構築中の凝集塊に取り込まれる	内胚葉部分に定位する	胚の再構築，細胞選別
29	内胚葉上皮	イトマキヒトデ Asterina pectinifera	初期胞胚期 -後期原腸胚期	再構築胚内で他の内胚葉上皮と出会う	合体して，より大きな上皮を形成する	胚の再構築
30	胚細胞	イトマキヒトデ Asterina pectinifera	初期胞胚期 -後期原腸胚期	再構築胚内で外胚葉と内胚葉の上皮が出会う	外胚葉上皮が，内胚葉上皮の一点をこじ開けて割り込み，そこを中心として内胚葉上皮と連結する	胚の再構築，肛門形成
31	内胚葉上皮	イトマキヒトデ Asterina pectinifera	初期胞胚期 -後期原腸胚期	肛門が開いてしばらくする	肛門の反対極をふくらませて，食道を分化させる	胚の再構築

補足資料 5　細胞行動データ（説明以外）　ID 32–40

ID	細胞・組織・器官名	生物名	発生ステージ	入力	出力	生命現象
32	融合細胞 fusogenic phagocyte（体腔細胞，血球）	マボヤ Halocynthia roretzi	成体	培養基質上に置かれる	優雅なすべり運動により，平均速度 10.8 μm/min で動き回る	基質上での移動
33	食道上皮	イトマキヒトデ Asterina pectinifera	初期胞胚期-後期原腸胚期	食道上皮として完成すると間もなく	外胚葉上皮と連結して，口を開ける	胚の再構築，口形成
34	外胚葉上皮	イトマキヒトデ	原腸胚	周囲の海水がグリシン，アラビノース，尿素，食塩などの低分子化合物によって高調になる	septate junction の septa を解消する	上皮の開閉
35	食細胞 Phagocyte	イトマキヒトデ Asterina pectinifera	成体	体液中に浮いているときは	花びら状の仮足を複数枚，広げている	細胞の形
36	食細胞 Phagocyte	キヒトデ Asterias amurensis	成体	培養基質に接触する	葉状仮足を円盤状に広げて伸展する	培養基質上での伸展
37	食細胞 Phagocyte	キヒトデ Asterias amurensis	成体	培養基質上で他の食細胞に出会う	融合する	細胞融合
38	食細胞 Phagocyte	イトマキヒトデ Asterina pectinifera	成体	ある程度以上の密度で，20 時間ほど培養下におかれる	独特のレース状の多核体を作る	多核体形成
39	顆粒細胞（血球）	カイコ	幼虫	体腔内に注入された直径の 5.7 μm のラテックスビーズに触れる	葉状仮足を伸展させて，ビーズをとりこむ	生体防御，食作用
40	食細胞 Phagocyte	サンショウウニ Temnopleures toreumaticus	成体	ある程度以上の密度で，20 時間ほど培養下におかれる	独特のレース状の多核体を作る	多核体形成

補足資料 6　細胞行動データ（説明以外）　ID 41–49

ID	細胞・組織・器官名	生物名	発生ステージ	入力	出力	生命現象
41	食細胞 Phagocyte	ニホンクモヒトデ Ophioplocus japonicus	成体	ある程度以上の密度で，20 時間ほど培養下におかれると	独特のレース状の多核体を作る	多核体形成
42	外胚葉上皮	イトマキヒトデ Asterina pectinifera	初期胞胚期-後期原腸胚期	外胚葉上皮が完成し，胞胚腔を広げながら 15 時間ほどたつ	胚の後部（＝肛門の場所）となるべき外胚葉が陥入を開始する	胚の再構築，体軸の決定
43	食細胞 Phagocyte	マナマコ Stichopus japonicus	成体	ある程度以上の密度で，長時間，培養下におかれる	分厚く，良くまとまった多核体を作る	多核体形成
44	食細胞 Phagocyte	ホシムシ Sipuncula cumanense	成体	ある程度以上の密度で，培養基質上に長時間おかれる	食細胞同士が側方で接着して，繊維芽細胞様の編み目パターンを作る	培養基質上での伸展
45	ジャイアント細胞 giant cell（体腔細胞，血球）	マボヤ Halocynthia roretzi	成体	培養基質上に置かれる	激しいすべり運動により，平均速度 33.0 μm/min で動き回る	基質上での移動
46	食細胞 Phagocyte	タイラギ（二枚貝） Atrina pectinata	成体	ある程度以上の密度で，培養基質上に長時間おかれる	食細胞同士が側方で接着して，繊維芽細胞様の編み目パターンを作る	培養基質上での伸展
47	臍帯静脈の内皮細胞	ヒト	胎児	バラバラの状態で，マトリゲル（基底層）上に播種される	3 時間以内に，互いに連結した網目状のパターンを作る	細胞配列パターン，基底層
48	臍帯静脈の内皮細胞	ヒト	胎児	ラミニン受容体に対する抗体の存在下でマトリゲル上に播種される	網目状パターンを作らなくなる	形態形成と基底層
49	臍帯静脈の内皮細胞	ヒト	胎児	柔らかい基質にコートされたマトリゲル（基底層）の上に播種される	基質をシワシワに変形させる	細胞の張力

補足資料 7　細胞行動データ（説明以外）　ID 50–59

ID	細胞・組織・器官名	生物名	発生ステージ	入力	出力	生命現象
50	臍帯静脈の内皮細胞	ヒト	胎児	基底層にシワができる	シワに沿って仮足を伸ばす	基質の形状，細胞移動
51	臍帯静脈の内皮細胞	ヒト	胎児	マトリゲル（基底層）でコートされた堅い基質上に置かれる	離れた細胞との間のマトリゲルに細いシワを作り出す	細胞の張力
52	ジャイアント細胞 giant cell （体腔細胞，血球）	マボヤ Halocynthia roretzi	成体	培養基質上で，基質に強く接着している他種の細胞と出会う	方向変換するか，細胞体を２つに割って，跨いで通る	回避反応
53	全能性血球 Totipotent blood cell	群体ボヤ Botrylloides leachi	成体	親群体から切り離された血管の小片内に取り残される	個虫を形成する	再生 Vascular budding
54	歯肉由来の繊維芽細胞	ヒト	成体	溝のあるチタンコートされた基質上におかれる	溝に沿って長く伸びる	培養基質の形状と細胞の形
55	食細胞 Phagocyte	ユムシ Urechis unicinctus	成体	ある程度以上の密度で，培養基質上に長時間おかれる	食細胞同士が側方で接着して，繊維芽細胞様の編み目パターンを作る	培養基質上での伸展
56	食細胞 Phagocyte	アカウニ Pseudocentrotus depressus	成体	ある程度以上の密度で，20時間ほど培養下におかれる	独特のレース状の多核体を作る	多核体形成
57	歯肉由来の繊維芽細胞	ヒト	成体	溝のあるチタンコートされた基質上におかれる	平らなチタン基質におかれた細胞よりも多くの細胞外マトリックスを分泌する	培養基質の形状と細胞外マトリックスの分泌
58	心臓由来の繊維芽細胞	ニワトリ	胚期	シリコンラバーの基質の上におかれる	基質にシワを作り出す	細胞の張力
59	心臓の小片	ニワトリ	胚期	シリコンラバーの基質の上におかれる	基質にシワを作り出す	細胞集団，細胞の張力

補足資料 8　細胞行動データ（説明以外）　ID 60–68

ID	細胞・組織・器官名	生物名	発生ステージ	入力	出力	生命現象
60	PTK-1 細胞（株細胞），腎臓	Rat kangaroo Potorous tridactylis	成体	シリコンラバーの基質の上におかれる	基質にシワを作り出す	細胞の張力
61	毛細管内皮細胞	ヒト	成体	基質への伸展状態に応じて	自殺の率を変える	細胞の伸展，アポトーシス
62	毛細管内皮細胞	ヒト	成体	基質への伸展の面積に応じて	自殺の率を変える	細胞の伸展，アポトーシス
63	臍帯静脈の血管内皮細胞	ヒト	胎児	培養下で水流を受けると	基質にしがみつく	細胞と力
64	大動脈由来の血管内皮細胞	ニワトリ	成体	血流にさらされる	基質にしがみつく	細胞の接着力
65	アメーバ様細胞 phago-amoebocyte（体腔細胞，血球）	マボヤ	成体	培養基質上でヒツジの赤血球と出会う	食べる	食作用
66	血管内皮細胞	ウシ	成体	さらされている血流の強さに応じ	ストレスファイバーを発達させる	接着の強さ，細胞の力学的性質
67	融合細胞 fusogenic phagocyte（体腔細胞，血球）	マボヤ	成体	培養基質上でヒツジの赤血球と出会う	食べる	食作用
68	リンフォイド細胞 lymphoid cell （体腔細胞，血球）	マボヤ	成体	培養基質上で固定されたヒツジの赤血球と出会う	複数まとめて抱え込む	抱え込み作用

補足資料 9　細胞行動データ（説明以外）　ID 69–76

ID	細胞・組織・器官名	生物名	発生ステージ	入力	出力	生命現象
69	黒色素胞 Melanophore	アフリカツメガエル	幼生　オタマジャクシ	MSH 存在下で培養されて基質上に伸展し，色素顆粒を拡散させることが出来るようになる	増殖する	培養基質上での動態
70	液胞細胞 vacuolated cell （体腔細胞，血球）	マボヤ	成体	培養基質上でヒツジの赤血球と出会う	食べる	食作用
71	大型黒色素胞（マクロメラノフォア）	プラティフィッシュ	成体	小型色素胞しかいない皮膚の領域に移植される	近傍の小型色素胞を消滅させていく	紋様形成，細胞間相互作用
72	神経提（冠）細胞	ウズラ	神経胚	ファイブロネクチンをコートした基質上で，密度の高い状態で培養される	一部のものがメラノサイトに分化する	細胞分化，密度効果
73	神経提（冠）細胞	ウズラ	神経胚	集団でラミニンをコートしたカバーガラス上で培養される	一部がメラニサイトに分化する	細胞分化
74	黒色素胞	ウシガエル	幼生	培養基質上におかれる	太い樹状突起（仮足）を放射状に出して伸展する	培養基質上での動態
75	ジャイアント細胞　giant cell（体腔細胞，血球）	マボヤ Halocynthia roretzi	成体	培養基質に触れる	四方八方に激しく仮足を出し入れした後，急激に葉状仮足を半円形に広げて，基質上に伸展する	培養基質への反応
76	融合細胞 fusogenic phagocyte（体腔細胞，血球）	マボヤ Halocynthia roretzi	成体	培養基質上で他種の細胞と接触する	方向変換して通り過ぎる	回避反応

補足資料 10　細胞行動データ（説明以外）　ID 77–84

ID	細胞・組織・器官名	生物名	発生ステージ	入力	出力	生命現象
77	融合細胞 fusogenic phagocyte（体腔細胞，血球）	マボヤ Halocynthia roretzi	成体	多種類の体腔細胞からなる巨大な細胞集塊として培養基質上に置かれる	同種の細胞同士で巨大な融合体を形成する	細胞融合
78	Type 1 細胞（血球）	ナメクジ	成体	個体が外傷を受ける	傷口表面で伸展する	傷の治癒
79	線維芽細胞様細胞　fibroblastic cell （体腔細胞，血球）	マボヤ Halocynthia roretzi	成体	培養基上に置かれる	細胞外マトリックスを分泌しながら，ゆっくりと伸展する	培養基質への反応
80	神経冠（提）細胞	ウズラ	神経胚	集団で，ファイブロネクチンをコートしたカバーガラスの上で培養される	広がって，フラットな形態をとる	培養基頂上での動態
81	クラミドモナス	クラミドモナス	なし	2 本の鞭毛のうち，片方を切断される	切断された部位に新しく鞭毛を伸長させると同時に残った方の鞭毛を縮める	鞭毛の再生
82	クラミドモナス	クラミドモナス	なし	片方の鞭毛を切断され，その部位に新しく鞭毛が伸長し始め，残った鞭毛が縮んで 2 本の長さが揃う	2 本の鞭毛を同時に伸長させ始める	鞭毛の再生
83	Saccharomyces cerevisiae	Saccharomyces cerevisiae	なし	α 因子を受容する	極性を持つようになる	細胞の極性化
84	アメーバ細胞	キイロタマホコリカビ	なし	環状 AMP をマイクロピペットからしみ出すように与えられる	環状 AMP 源の方向に移動する	走化性

補足資料 11　細胞行動データ（説明以外）　ID 85–93

ID	細胞・組織・器官名	生物名	発生ステージ	入力	出力	生命現象
85	胚性癌腫細胞 P 19 株（P 19 embryonic carcinoma cell: P 19 EC-cell）	ネズミ	成体	レチノイン酸存在下で	神経細胞接着因子（NCAMs）を発現する	分化誘導
87	黒色素胞	ウシガエル	幼生	培養下で表皮細胞と接触する	太い樹状突起を放射状に出した形状（真皮性黒色素胞型の形態）から，細い突起をもつ双極性の形状（表皮性黒色素胞型の形態）に変化する。	細胞間相互作用
88	腹側の皮膚	アフリカツメガエル	幼生，オタマジャクシ	小片として培養下におかれ，細胞が培養基質に遊出する	一部の細胞が黒色素胞に分化する	細胞分化
89	予定体腔嚢細胞	イトマキヒトデ	原腸胚後期–ビピンナリア幼生	原腸先端が膨潤し終わる	“腸体腔型”で体腔嚢を形成する	体腔嚢形成
90	間充織細胞	イトマキヒトデ Asterina pectinifera	原腸胚後期	異物であるプラスティック皿上で多核状態に変化した後，振動を続ける	単核状態に戻る	多核状態の解消
91	原腸の細胞	Lytechinus pictus	Late gastrula	原腸陥入の第 2 段階に入る	互いに位置を変える	原腸の伸長
92	間充織細胞	イトマキヒトデ Asterina pectinifera	原腸胚後期	異物であるプラスティック皿に触れる	互いに融合することにより多核体を形成する	多核体形成
93	神経冠（提）細胞	ウズラ	神経胚	集団でラミニンをコートしたカバーガラスの上に培養される	突起を伸ばしたり，紡錘状の形態をとり，凝集塊をつくる	培養基頂上での動態

補足資料 12　細胞行動データ（説明以外）　ID 94–100

ID	細胞・組織・器官名	生物名	発生ステージ	入力	出力	生命現象
94	前側の割球	Volvox carteri	16 細胞期	他の割球から単離される	ゴニディア（無性生殖細胞）を 1 個分化させる	細胞分化，調節性，無性生殖
95	割球	Volvox carteri の Gls/Reg 変異体	64 細胞期	熱処理（42.5℃，40 分に続き 45 ℃，20 分）を加えられる	全ての細胞が体細胞に分化する	細胞分化，サイズ依存性，無性生殖
96	割球	Volvox carteri の Gls/Reg 変異体	16 細胞期	熱処理（42.5℃，40 分に続き 45 ℃，20 分）を加えられる，または，この時点での細胞の直径が 11 – 14 ミクロンの間である	その後 1 回の分裂の後に分裂を停止し，全ての細胞がゴニディア（無性生殖細胞）に分化する	細胞分化，サイズ依存性，無性生殖
97	割球	Volvox carteri	卵割期	32–128 細胞期に熱処理により分裂を停止させられる	ゴニディア（無性生殖細胞）を過剰に持った個体に発生する	細胞分化，サイズ依存性，無性生殖
98	割球	Volvox carteri の S 16 温度感受性変異体	卵割期	卵割終了時に割球の直径が 8 μm 以上ある	ゴニディア（無性生殖細胞）に分化する	細胞分化，サイズ依存性，無性生殖
99	割球	Volvox carteri の Pcd-K 1 および Sls-S 1 変異体	卵割期	卵割終了時に直径 9 μm 以上の大きさがあると	ゴニディア（無性生殖細胞）に分化する	細胞分化，サイズ依存性，無性生殖
100	砂嚢上皮	ニワトリ	6 日胚	前胃間充織と組み合わせて培養される	前胃上皮へ分化する	誘導，上皮間充織相互作用

補足資料 13　細胞行動データ（説明以外）　ID 102–110

ID	細胞・組織・器官名	生物名	発生ステージ	入力	出力	生命現象
102	後方部の脊索中胚葉	有尾両生類	原腸陥入期	別の胚の腹部に移植される	尾部を誘導する	尾部誘導 spinocaudal induction
103	胚盤葉後部の上層の細胞群	ニワトリ	原始線条期，原条形成期	胞胚期に到達する	初原条部に集まる	原始線条（primitive streak）の形成，原条形成
104	ヘンゼン結節が存在していた領域下の中胚葉細胞群	ニワトリ	原始線条期，原条形成期	原始線条（原条）が退縮する _x 000 D_	脊索になる	脊索形成
105	脊索	ニワトリ	原始線条期，原条形成期	ヘンゼン結節が退縮する	退縮した領域を追って長くなる	原始線条（原条）の退縮と脊索形成
106	予定中胚葉細胞	ニワトリ	原始線条期（胚内陥入時），原条形成期	原始線条（原条）から胚内に入るときに	ボトル形に変形する	細胞の変形
107	予定内胚葉細胞	ニワトリ	原始線条期，原条形成期	原始線条（原条）およびヘンゼン結節を通って胚盤葉下層に入る	内胚葉を形成する	内胚葉の形成
108	受精卵	有尾両生類	受精直後（背腹軸決定前）	精子侵入点が最上部になるように 90°回転させられる	将来灰色半月と 180°ずれた部位に原口を形成する	軸形成
109	受精卵	有尾両生類	受精直後（背腹軸決定期）	精子侵入点が最上部になるように 90°回転させられる	将来 2 つの原口を形成する	軸形成
110	予定原口背唇部直下の割球	両生類	胞胚	別の胞胚の植物半球に移植される	移植された場所の細胞群にもうひとつの原腸を形成させる	原腸の誘導

補足資料 14　細胞行動データ（説明）　ID 1–7

ID	説明
1	動物の皮膚には，メラノサイトと呼ばれる黒色の色素細胞が分布している。真皮域のメラノサイトは，自身の細胞内でのメラノソームの凝集と拡散により，体色変化を引き起こす。一方，表皮域でのメラノサイトはその樹状突起を周りのケラチノサイトに食作用させることにより皮膚を着色する。日焼けによる皮膚の黒化も同様のメカニズムで生じる。図は，ヒョウガエルの脊側の皮膚において，メラノソームが，樹状突起を伸ばしたメラノサイトからケラチノサイトに移動していることを示している。この状態は表皮性メラニン単位と呼ばれる。
2	メダカの鱗にはメラノフォアと呼ばれる黒色の色素細胞が分布している。メラノフォアは，自身の細胞内の色素顆粒（メラノソーム）を暗所では凝集させ，明所では拡散させることにより，メダカの体色変化を引き起こす。このメラノソームの拡散は，メラノフォア刺激ホルモン（MSH）により制御されている。図はメダカの鱗から遊離したメラノフォアの MSH への反応を示している。暗所（図の上部の黒線）で凝集しているメラノソームに MSH を添加すると，メラノソームは拡散し（図中↑），MSH を除去すると凝集する（図中↓）。
3	角膜内皮は，角膜の裏側（外界に面していない側）を覆う単層上皮である。ニワトリ 3 日胚ごろに，角膜，水晶体，眼胞からなる眼球の原型ができたときには，角膜の裏側は，細胞で覆われておらず，基質がむきだしになっている。4 日胚ごろになると（図 a），周辺から繊維芽細胞の形態をもった予定内皮細胞が進入してきて（図では右上と左下から），5 日胚でほぼ全面に散在するに至る。6 日胚（図 b）では細胞間の隙間がほぼなくなり，7 日胚（図 c）では上皮の形態を備えるようになる。
4	細胞性粘菌で，countin 2 と呼ばれる遺伝子を破壊すると，通常より小さな細胞集合体を作った段階で，子実体形成が始まる。countin 2 がコードするタンパク質は細胞外に分泌されると推測され，細胞集合体が子実体への分化を始めるときの最小サイズを決めているらしい。countin 2 破壊株では，この制限がはずれて小さい細胞集合体の段階で分化が始まる。40％のアミノ酸同一性を持つ countin が既に報告されており，こちらは細胞集合体の最大サイズを決めているらしい。図 A：同じ数の細胞から出発してできてきた細胞集合体の数の比較。図 B：同じ倍率でとった細胞性粘菌の集合体（スケール：0.5 mm）。A × 2：野生型，countin 2 ^ (-)：countin 2 破壊株，countin^ (-)：countin 破壊株，countin^ (-) and countin 2 ^ (-)：countin と countin 2 の破壊株。
5	神経筋接合が形成される時期になると，それまで筋肉細胞表面に存在していた commissureless 遺伝子産物（COMM）が細胞内のエンドゾームに移動する（図 a，b）。COMM を持たないミュータントでは，神経筋接合ができないが，野生型の COMM を再導入するとレスキューされる。このとき COMM の細胞内部分を短くして細胞内にとりこまれない形にして導入したのではレスキューできない。これらの結果から，筋肉表面には元々神経の結合を阻害する膜タンパクが分布しているが，これらが COMM の移動に伴って細胞内に移動し，神経が結合できるようになるという仮説がたてられた。
7	淡水カイメンの Ephydatia fluviatilis は，静水中で飼育すると大孔を筒状に伸ばし（図 a），静かに水を吐き出している。この状態で基部に電気刺激（図 b：矢印）を加えると，刺激された部位で急激に収縮し（図 b），続いて，約 15 分かけて，全体を縦横に収縮させる（図 c）。なお，大孔部分の中膠には平滑筋細胞が存在することが知られている。ただし，からだのどこを探しても，神経細胞は見つかっていない。

補足資料 15　細胞行動データ（説明）　説明　ID　8―16

ID	説明
8	器官培養の条件下で肺原基の分枝形成（branching morphogenesis）を起こさせるときに，培養液に 100 mM 程度のショ糖を加えておくと，　分枝形成が促進された。ショ糖のかわりにラクトースや食塩でも同様の現象がみられ，浸透圧によるものらしい。外液の浸透圧を上げる代わりに，肺原基の気管支部分に細いナイロン糸を挿入し，内外の液の流れを確保してやっても同様に分枝形成の促進が見られることから，原基内部に水がたまるのを防ぐことが重要らしい。図 A：0 mM ショ糖を加えた培養液中での肺原基。培養 3 日目。図 B，C，D，E はそれぞれ 20 mM，30 mM，50 mM，100 mM ショ糖を加えた培養液中での肺原基。ショ糖濃度が上がるに従ってより分枝している。
9	海綿動物の細胞は septum のない簡単な結合で結ばれるのがふつうである。それに対し，中膠に存在する変形細胞は，骨片を形成する際には数匹が集まり（図 A ＊），水溶性分子の通過を許さない septate junction で結合して閉じた袋を作る（図 B）。 Bergquist PR. 1978. の仕事。＊：Brusca RC, Brusca GJ. 1990. Invertebrates. Sinauer Asociates. p. 192. Figure 8 A から借用。
12	白い背景に移す前のメダカ成体の鱗には，多数の黒色素胞が存在している（図 A）。このメダカを白い背景に移して 6 日後の鱗では，黒色素胞が徐々にアポトーシスにより消滅している（図 B）。そしてアポトーシスを引き起こした黒色素胞では，断片化した黒い小胞が観察される（図 B）。さらに 10 日後の鱗では，黒色素胞の消滅はかなり進行している。
13	スズガエルにおいては，発生に伴い，胚の背側尾部における黒色素胞の配列が変化する。受精後 6 日目では，背側尾部の皮膚における黒色素胞の配列はランダムであるが（A），発生が進むにつれて，多数の黒色素胞が直角方向に樹状突起を出しあって，規則的なパターンを形成するようになる（B：受精後 9 日目，C：受精後 16 日目）。
14	昆虫は，自然界において様々な病原体の体腔内への侵入を受けた場合や，実験的に粒子を体腔内に注入された場合に，顆粒細胞やプラズマ細胞によってその異物を処理する。異物処理方法として，食作用，ノジュール形成が知られているが，包囲化も処理方法の 1 つである。包囲化は，異物のサイズが大きい場合に観察される。図は，カイコ幼虫に直径 25.7 μm のラテックスビーズを注入したときに，多数の顆粒細胞が 1 つのビーズを包囲化する様子を示している（図 1 は，顆粒細胞がラテックスビーズを包み始めた状態。図 2 は，ラテックスビーズを完全に包み込んだ状態）。
15	UV 照射を受けたアフリカツメガエルの卵母細胞は，成熟，受精後，発生すると，軸構造形成が阻止された腹側化胚を生じる（図）。成熟卵に対する UV 照射実験でも同様の結果が得られるが，卵母細胞の場合，成熟卵で判明している微小管の機能阻害とは関係なく，軸構造の決定因子そのものが破壊されたことによると考えられている。
16	マボヤの体腔細胞の一種である multipolar cell は，海水中でプラスティックシャーレやガラス上に置かれると，図のように多数の糸状仮足をピンと延ばして，基質に接着する。この細胞は，基質上を移動しない。この細胞は，少なくとも fusogenic phagocyte（図中の f 1）と phago-amoebocyte（図中の p）との接触を回避しない。図 a，b は位相差顕微鏡像，図 c は走査型電顕像。

補足資料 16　細胞行動データ（説明）　説明　ID 22―60

ID	説明
22	カイメン動物の中膠に存在する変形細胞は，必要に応じて数匹が集まり，協同で作った袋の中に，種に固有な，特定の形の骨片を作る。
23	イトマキヒトデの初期胞胚期から後期原腸胚の胚（図 A）を，1.2 M グリシン溶液中でピペッティングすると，胚を個々の細胞の状態にまでバラバラにすることができる（図 B）。こうして解離された細胞を海水中に懸濁すると（図 B），細胞の種類の違いに関係なく，直ちに集合を開始し，様々な大きさの凝集塊を作る（図 C）。この状態の凝集塊は，互いに出会って合体し，さらに大きな凝集塊に成長する（図 D）。
34	胞胚期以降のヒトデ胚の上皮細胞（内胚葉と外胚葉）は，互いに septate junction で結合しており（図 A：矢印，梯子状構造），海水中の巨大分子（例えば IgG：免疫グロブリンなど）を体内に通すことはない（図 B，B’）。しかし，海水にグリシン，アラビノース，尿素，食塩などの低分子化合物を 0.3 M の濃度で加えて高調にすると，胚は体内に巨大分子（ここでは蛍光ラベルした IgG）を取り込む（図 C，C’）。この時，外胚葉の septate junction を電子顕微鏡で見ると，隔壁が消えてなくなっている（図 D）。図 A，図 D は電子顕微鏡写真。図 B，図 C は，上が光学顕微鏡写真，下が同じ胚を IgG の抗体で染色した蛍光抗体写真。
35	ヒトデの体腔液中に浮遊している食細胞は，体液から取り出されたばかりの状態では，下図のように単独で存在しており，複数の葉状仮足を花びらのような特徴的な形に広げて（矢印），浮遊に適した形を取る。
47	ヒトの臍帯静脈の内皮細胞は，バラバラに解離された状態で，マトリゲル（基底層）でコートされた培養皿の上に播種されると，その直後はほぼ均一に分布するが，3 時間以内に互いに連結して，網目状のパターン（下図）を作り出す。なお，ラミニン受容体に対する抗体の存在下では，網目状パターンを作らない。
53	ヨーロッパ産の群体ボヤ Botrylloides leachi（図 A）から全ての個虫を取り除き，被嚢とその内部に走る血管だけを基質上（ここではスライドグラス）に残してやる（図 B：6 つの小片が取り残されている）。すると 2 日後には，小片の血管内部に血球の丸い凝集体（＝芽体）が 1 個から数個出現する（図 C）。この芽体は盛んに分裂を繰り返して，中空の球（図 D）を経て，72 時間以内に 1 個の個虫を生じる（図 E）。なお，複数の芽体が出現した小片では，1 つの芽体だけが個虫に発達し，残りは吸収されてしまう。また，切り離された小片は，100 - 200 個以上の血球を含むものであれば，大きさは問わない。
60	PTK-1 細胞をトリプシンでバラバラにした後シリコンラバーの基質上においてやると，その上に円盤状に伸展する。伸展初期の細胞と基質との接着面を見ると，基質に不規則なシワが発生している（下図）。このことから，細胞は仮足を伸ばすさいに，細胞中心部で力を出していることが分かる。