

A10.1-10.7

Corpora and applied linguistics

英語科3年 東野健太

目次

10.1 Introduction (紹介)

10.2 Lexicographic and lexical studies (辞書学/語彙研究)

10.3 Grammatical studies (文法研究)

10.4 Register variation and genre analysis (レジスター変異/ジャンル分析)

目次

10.5 Dialect distinction and language variety (方言の区別と言語変種)

10.6 Contrastive and translation studies (対照/翻訳研究)

10.7 Diachronic study and language change (通時的研究と言語変化)

10.1 Introduction

(紹介)

主題：言語研究でコーパスがどう使われるのか

- ◆ Leech (1997) 「ほとんどすべての言語学/言語学習の領域で」
- ◆ 実証性に強み
 - 話者の直感を蓄積、言語分析をより客観的に
- ◆ 10.2-10.8 コーパスが使われる主要な言語学の領域チェック
 - ◆ (10.9-10.14はコーパスを使い始めた他の領域)
 - ◆ (10.5は言語研究におけるコーパス利用の制限)

10.2 Lexicographic and lexical studies

(辞書学/語彙研究)

辞書学/語彙の研究においてコーパスは貴重

→1990s以降の辞書はコーパス準拠

初めに辞書学について。

辞書編纂者にとっての便利な点（3点）

①機械で読める

→生の典型的な語彙の用法をテキスト群から数秒で抽出できる

②コロケーションの頻度情報と定量化

e.g. COBUILD1995/Longman 1995

→教材を書く人や学習者に便利

③マークアップとアノテーションの恩恵

(1)テキスト情報 (e.g. ジャンル) /社会言語学的情報 (e.g. 性別)

→語彙の使用のより正確な記述へ

(2)POSタグや語彙の曖昧さ排除

→より実用的な多義語や同形異音異議語 (e.g. wind) のグルーピングを可能に

(3)モニターコーパス

→語彙の意味と用法の変遷を辿り辞書を最新にできる

(4)コーパスエビデンスが個々の編纂者の本能を補完/訂正

→辞書が正確に

以上のようなコーパスによって起きた変化をまとめたものが…

Hunston (2002) 5つの「強調」

- ◆ An emphasis on frequency: 頻度の強調
- ◆ An emphasis on collocation and phraseology: コロケーションと語法の強調
- ◆ An emphasis on variation: 変種の強調
- ◆ An emphasis on lexis in grammar: 文法における語彙の強調
- ◆ An emphasis on authenticity: 真正性の強調 *authentic: based on facts

借用語は以前辞書に入れるのかは編纂者の直観だったが…

Podhakecka and Piotrowski (2003)

“Russianisms” in BNC vs. OED/LDOCE

【予想通りだったもの】

360の借用語のうち半分がBNCに一度しか登場せず。
頻度が高いものはごくわずか

【予想外だったこと】

語源準拠でOEDからの選出した項目と頻度準拠でBNCから選出した項目 (item) の分布が同じだった #itemとは？分布とは？

彼らは借用語のコンテクストは考慮せず。それに対して、

Kurtböke and Potter (2000)

トルコ語コーパスの英語借用語/英語コーパスのイタリア語借用語調査
「借用語周辺のコロケーションのパターンも借用語の扱いに含めるべき」

「頻度数に基づく同化基準 (assimilation criteria) は従来考えられていたよりも信頼性が低く、メタファー等の代替基準も考慮に入れる必要がある」 #assimilation criteriaとは？

ここからは語彙研究について。まずは形式の話。

語彙研究では、

- ①コロケーション = 意味の韻律 (**semantic prosody**)
- ②コロケーションの意味 = 意味の選好 (**semantic preference**)

に焦点を当てる

Firth (1957): collocationの生みの親

Firth (1968): collocationは任意の語の慣習的な位置に関する表現

→量的/コロケーションへの統計手法は多くのコーパス言語学者が追随

→彼らはコロケーションは語のパターンの特徴的な共起だとする点で一致

Greenbaum (1974)

「言語の二つの語彙項目の頻繁な共起」

→統計的に重要なコロケーションのみを示すと想定

collocability, collocable: 共起可能性のある語に対して

collocation, collocate : 実際に頻繁に共起する語に対して

ただ、「どのくらい」頻度が高ければいいのかわからないので…

Hoey (1991)

文脈上ランダムの確立よりも大きい確率で語彙項目が共起すればコロケーション

#MI (相互情報量) , t/z score (t値/z値) とは

Hoey (1991) にすべての言語学者が賛同したわけではない

Herbst (1996)

コロケーションへの統計的手法に反対

「Houseに対する限定詞the/thisや動詞sellが来ることは不思議でもないし興味深くもない」

限定詞 (determiner) が名詞の共起表現の上位に来てしまうのは避けられない。

↓

中心語 (node) と特定の語の文法クラス (grammatical class) のコロケーションはコリゲーション (**colligation**) と解釈される

↓

コロケーションリストに対する手法を選択的にしなければならない

共起表現の調査はコンピューターの時代にコーパス言語学が進歩し、適切なソフトウェアによって比較的些細な作業に



計算の容易さのみならず言語学者・辞書編纂者を直観頼りなコーパス研究から解放させた



Greenbaumの頃（70年代）はコーパス資源がなく、大量のネイティブスピーカーの直感頼り。コロケーションについて意見が一致しないことも珍しくなかった（Partington, 1998）。



直観頼りの問題点が浮き彫りに。

統計手法による解決にはコーパスが不可欠に。

ここまででは語彙の形式について。ここからは語彙の意味について。

中心語 (node) と共に語 (collocate) の相互作用によって現れるコロケーションの意味を意味の韻律 (**semantic prosody**) と呼ぶ

- ①書き手・話し手の態度や評価を表現することが主な機能。
- ②ネガティブなものが大半だが時にポジティブのものもある。

e.g. Sinclair (1987, 1991)

happen/set in が不快な状況を表す名詞と使われる

③皮肉や不誠実、ユーモアの際など意味の韻律 (semantic prosody) が破られることも (Louw, 1993)

④個々の単語では意味を超えて作用することがある

e.g. personal price

personal も price も中立的表現

personal priceだと認可望ましくないものを指示する
(BNCとBank of Englishにて証明)

意味の韻律（semantic prosody）はコロケーションと同じくらい話者の意識的な内省に近づくことができない

コーパスサイズの肥大化とツールの開発で意味の韻律（semantic prosody）は言語学者により取り組まれるようになってきている

- ①ある項目とその典型的な共起語（collocates）の相互作用による結果
- ②典型的な共起語（collocates）のコンテクスト以外では（その語から影響を受ける）感情的な意味は示さない
- ③典型的でない共起語（collocates）と使われるときにも感情的な意味を帯びる（②の逆）

任意の中心語（node）に関して、共起語（collocates）に共通する意味的特徴から浮かび上がる意味は意味の選好（semantic preference）とよばれる。

「ある意味的特徴を共有する、頻繁に現れる共起語群（collocates）」と定義される。

e.g. largeは質と大きさを示す意味群の項目と典型的に共起する（number, scaleなど）

意味の選好 (semantic preference) と意味の韻律 (semantic prosody) は別個の独立した意味である

Sinclar (1996, 1998) and Stubbs (2001)

意味の韻律 (semantic prosody) は語彙単位 (lexical units) 間の関係についての更なる抽象化の段階である

collocation: 中心語 (node) と個々の単語の関係

colligation: 中心語 (node) と文法項目の関係

semantic preference: 共起語群 (collocates)

semantic prosody: 典型的な共起語群を伴う任意の中心語 (node) の感情的意味

Partington (2004)

意味の選好 (semantic preference) と意味の韻律 (semantic prosody)
は異なる作動領域を持つ

↓

semantic preference:

中心語 (node) を特定の意味群からのほかの項目に関連付ける。共起語群 (collocates) の特徴と見ることもできる。

semantic prosody:

より広いテキストの範囲に影響する。中心語 (node) の特徴とみることもできる。

相互作用:

意味の韻律は中心語の選好的な選択を制限する一般的な環境を決めるのに対して、意味の選好は韻律を作るのに強力に貢献している

意味の韻律 (semantic prosody) が内包的な意味 (connotative meaning) なのは異なる意見がある #内包的な意味とは

Partington (1998), Stubbs (2001), and Hunston (2002)

意味の韻律 (semantic prosody) は内包的

Louw (2000)

意味の韻律 (semantic prosody) は、背後での力が内包のスキーマ的アスペクトよりも強力に連語的であるから、単に連語的であるというだけではない。

#スキーマ的アспект (schematic aspects of connotation) とは

著者ら

内包は連語的にも非連語的にもなれるが、意味の韻律は連語的にしかなれない

語彙研究には形態素分析も含む

執筆当時は入手可能な形態素のアノテーション付きコーパスはなく、University College Londonが形態素のアノテーションをPOSタグ・構文解析済みのICE-GBに統合しようとしている。

形態論者、統語論者、辞書編纂者のみならず、学習者にも便利
e.g. -ic/-icalで終わる形容詞の意味・分布の相違 (Gries, 2003)

10.3 Grammatical studies

(文法研究)

文法も辞書学/語彙と並んでコーパスデータを頻繁に用いる領域

質的な言語や言語変種の統語的特徴を量化するだけでなく、文法理論から仮説を試すのに便利

文法の参考書にも影響を与えている

Huston(2002)

「コーパスのことを着たことがない人でもコーパス準拠の調査製品を使っている」

A comprehensive grammar of the English languageから
Longman Grammar of Spoken and Written Englishへ

コーパス準拠の文法には多様な点で唯一無二のもの

話し言葉と書き言葉の違い、レジスターの変異（register variation）を説明するなど

↓

話し言葉に対する文法で取り扱う範囲が特に大切

↓

従来の文法研究は書き言葉に集中していたが、伝統的な文法では話し言葉を記述するには不十分

ノッティンガム学派

話し言葉の文法に談話の観点からアプローチする
書き言葉文法には無い/軽視されていた、多くの話し言葉文法の
特徴を発見

McCarthy (1995)

過去進行形で現れるsayやtellといった伝達動詞のような、頻出する伝達現象が見落とされていたことを指摘。

伝達動詞の過去形は発話の内容に対して真正性を与えていたのに
対して、過去進行形は発話 자체に焦点を当てていた。

話し言葉の文法は書き言葉とは根本的に異なり、話し言葉のデータをもとにモデル化する必要性を唱える

ノッティンガム学派と逆の考え方をしたのが…

Biber et al (1999); Huddleston and Pullum (2000)

話し言葉と書き言葉の同一性に注目し、同じフレームワークで説明しようとした

バーミンガム学派 (Sinclair, Hunston, Francis and Manning)

文法説明に語彙に注目 (pattern grammar)。

語彙と文法の区別を拒否 (Sinclair et al, 1990)。

パターンと意味の関連性に重要性を置き、Collins COBUILD Grammar Patterns seriesで体現した。

↓

Francis et al (1998)

Bank of Englishの10000の名詞/形容詞から200のパターンを提示、意味と結び付けた

↓

言葉を孤立させずにフレーズの一部として扱うボキャブラリービルディングへの資源となる

10.4 Register variation and genre analysis

(レジスター変異とジャンル分析)

レジスター (register) とジャンル (genre)

コーパス設計は外的基準 (or状況パラメーター) に依存
→使用の社会的・コミュニケーション的文脈からレジスターを定義する。

Biberはレジスターとジャンルに互換性を持たせている

本書では厳密な定義はせず、排他的にならないようにする

コーパス（特に均衡サンプルコーパス）は幅広いレジスターとジャンルをカバーしているため、レジスター変異やジャンル分析の研究に適している

Oh(2000)

Switchboard corpusとBrown corpusを用いて、

- ①話し言葉ではactuallyがin factよりも3.7倍頻度が高い
- ②話し言葉と書き言葉双方で、actuallyは非発話末で用いられることが好まれること

を発見

MF/MD法

最も強力なレジスターとジャンルの変異への手法はBiber (1988)によって開発された**MF/MD法 (multi-feature/multi-dimensional analytical framework)**



LOBとLondon-Lund corporaから481のテキストを機能的に関係した67の言語特徴に基づいて、話し言葉/書き言葉のイギリス英語における21ジャンルの完全な分析を示す。

MF/MD法は因子分析に基づいている

因子分析は社会/行動科学において多変量間の相互関係を簡潔な形で要約するために用いられる。

Biberは因子分析を言語特徴の頻度数と合わせて、テキスト内で高頻度で共起する特徴群を特定した。

67の言語特徴を7つの因子 (dimension/factor) に落とし込むことができた。

MF/MD法の非技術的な説明

すべての言語特徴の相関マトリックスに基づいて因子の構造が確定し、各言語特徴の因子の重み (factor loading/weight) が計算される。因子の重みは任意の因子が個々の言語特徴にどれくらい一般化できるかを示している。

重みは正負どちらも取り、相関の方向を示す。ある因子に対してより大きな重みの絶対値を取るほど、その言語特徴は因子の代表性が大きいことになる。

MF/MD法の非技術的な説明（つづき）

Biberは重要で顕著な重みのみがそれぞれの因子の一部として解釈されるべきだと考え、0.30未満の絶対値を取る重みをもつすべての言語特徴を除外

多くの言語特徴がほとんどの因子に重みをもっていた為、因子の値の計算に含む言語特徴を決める基準を0.35まで引き上げた。

ある言語特徴が複数の因子に対して0.35以上の重みをもっていた場合は最も高い重みをもっている因子の値にのみ含まれた。

最終的な 7 つの因子

- ①情報的 vs. 自己関与が強い生産 (informational vs. involved production)
- ②口語的 vs. 非口語的関係 (narrative vs. non-narrative concerns)
- ③明示的 vs. 状況依存参照 (explicit vs. situation-dependent reference)
- ④説得の明白な表現 (overt expression of persuasion)
- ⑤抽象的情報 vs. 非抽象的情報 (abstract vs. non-abstract information)
- ⑥オンライン情報の精巧さ (online informational elaboration)
- ⑦学問的ヘッジ (academic hedging, Biber も詳しく説明せず)

様々な話し言葉/書き言葉英語のジャンルの近似性や違いを説明できるようになった。

e.g. 対面での会話と公的文書では①③⑤で大きく異なり、④では似ている

応用事例

MF/MD法はもともとは英語の書き言葉と話し言葉のレジスターを比較するために開発されたものであったが、

- ①特定のレジスター・ジャンルや著者のスタイルの共時的な分析
- ②レジスターの進化を記述し、初期近代英語の文学・非文学のジャンルの違いを調査する通時的な研究
- ③非西洋語のレジスター研究や対照分析
- ④コーパスデザイン問題
- ⑤レジスター/ジャンルやテキストタイプの定義問題

に応用されている。

単語連鎖 (lexical bundles) もレジスター変種の信頼できる指標であると分かつてきた

Biber and Conrad (1999)

会話と学問的散文での単語連鎖の構造的な種類が顕著に異なる

Biber (2003)

4語の単語連鎖に関する15の主要な種類の分布を、①会話②教室指導③教科書
④学問的散文の4レジスターで調査

話し言葉のレジスターで単語連鎖が顕著に頻繁に

会話では単語連鎖の90%近くが平叙文か疑問詞節内

学問的散文の場合は単語連鎖は節ではなく句から成っていた

②が会話と学問的散文の双方を扱うため、もっとも単語連鎖が多くなっていた

10.5 Dialect distinction and language variety

(方言の区別と言語変種)

言語変種 (language variety) とは

言語変種 (language variety) には、

- ①標準語 (standard language)
- ②方言 (dialect)
- ③社会方言 (sociolect)
- ④個人言語 (idiolect)
- ⑤専門語 (jargon)

などがある

言語変種 (language variety) とは

- ①本書では言語変種も方言も地理的に定義する
- ②国家変種は言語変種とみなす
- ③地域変種は方言とみなす
- ④他の変種（社会言語、個人言語、専門語）は言語変異（language variation）とみなす
- ⑤標準語は本書の定義としては方言とみなさないが、便宜的な判断であり、標準語を方言と認めないということではない
- ⑥言語変種にはピジンやクレオールのような言語変種も含める
- ⑦方言や言語変種における変異（variation）は発音や綴り、語の選択にみられるが、文法的な相違は珍しい
- ⑧方言は量的な違いがある

英語の方言や変種では中心的な文法的な構造は相対的に統一されている

Biber et al (1999)

Have gotten (アメリカ英語、特に会話内)

助動詞のあとにdoを続ける (イギリス英語)

Biber (1987)

名詞化や受け身、it分裂構造がアメリカ英語で顕著

時間や場所の副詞相当句、従位接続詞の消去はイギリス英語で顕著

Case study 2 in section C

helpの後に原形不定詞が続くことを好むアメリカ英語

Hundt (1998)

WWCとFLOBで過去形の規則動詞/不規則動詞の頻度や割合を調査
ニュージーランド英語とイギリス英語で大きな差 (56.4規則/68.7規則)
相当する動詞をBrown Corpusで調査、96.7%が規則的
3変種ともに違いがあり、特にNZ英語はBr英語に近い
*3コーパスの時期の差に留意

Tagnin and Teixeria (2003)

料理レシピの比較可能なコーパスで4つの言語変種（イギリス英語、アメリカ英語、ブラジルポルトガル語、欧州ポルトガル語）を調査。
様々なレベルでポルトガル変種間の違いの方が英語よりも顕著だった。

10.6 Contrastive and translation studies

(对照/翻訳研究)

同一言語の方言や変種の話から異なる言語間の話へ

多言語コーパス (multilingual corpora) を用いた主要な研究は、

①対照研究 (contrastive studies)

②翻訳研究 (translation studies)

のふたつ

翻訳研究について

コーパス準拠の翻訳研究は(1)理論的と(2)実用的の二種類が存在

- (1)ある言語での表現がほかの言語でどう伝わるのかを探求し、翻訳の対象言語と翻訳された言語間での言語特徴や頻度を比較することで翻訳プロセスを研究するのにコーパスを用いる
- (2)翻訳者の訓練の仕事台になったり、機械翻訳 (machine translation, MT) や機械支援翻訳 (computer-assisted translation) システムといったアプリケーションの開発の基礎に

ある言語での表現がほかの言語でどう伝わるのかを調べるためにパラレルコーパスは適している

Xiao and McEnery (2002)

英中パラレルコーパスを用いて英語での時制と相の意味が中国語でどう表現されるか調査。両言語とも進行相を持っているが、進行形は意味の範囲が異なっていた。

英語では93.5%が規範的に状況の進行する性質を指示している一方、一般的な進行形の定義に収まらないような特別な用法が数多く存在した（可能性のある慣習的/反復的状況を示す、予期された未来の出来事を指示する、特別な感情的効果を加えるためのイディオム的用法）。

一方中国語ではzaiで表現される進行相は動的状況での進行的性質のみに対応する。英語の進行形で示される状況の58%のみが、中国語で明白に或いは暗に示された進行/継続相を取った。

Xiao and McEnry (2002) (つづき)

状況相 (= 状況がnatural final endpointを持つか) と観点相 (e.g. 完了相 vs. 未完了相) 間での相互作用も翻訳者の観点相の選択に影響を及ぼす。

natural final endpointがある状況(約65%) や進行性と想いれない状況では、観点相の移動を経て、中国語の翻訳では完了相で現れる可能性がより高い。

対照的にnatural final endpointがない状況では通常zaiであらわされる進行形や-zheであらわされる継続相で翻訳される

#natural final endpointとは

ただし、この手の研究ではパラレルコーパスの翻訳の方向が大切

前頁の研究では中国語のアスペクトマーカーがどう英語に訳されるのか研究するのには適していない。中英パラレルコーパス（L1=中国語, L2=英語）が必要。

目標言語の翻訳の版が一つしかないコーパスのもう一つの問題点 翻訳がある個人の内観を示してしまうこと

Malmkjær (1998) 「なるべく多くの翻訳の版が必要」

実際にビルトするのは難しいし範囲が狭まる。文学の翻訳物であれば複数の版を持っていることがあるが作品が現代のものでなかつたり、版同士が何十年も時間に隔たりがある。

翻訳された言語の特徴は比較可能なL1テキストと比べることで特定され、翻訳プロセスに新しい光を投げかけ、翻訳の規範を特定する一助になる

Laviosa (1998)

L1/L2 の英語の話し言葉散文を調査

翻訳されたL2には四つの語彙使用のパターンがあった

- ①機能語 (function words) よりも語彙項目 (lexical words) の頻度が低い
- ②低頻度語よりも高頻度語の割合が高い
- ③最高頻度単語のより多い反復
- ④最も頻繁に使われる単語の多様性の低さ

そのほかの研究

翻訳された言語は語彙レベルを超えて名詞化・単純化・説明 (explication = 結束構造の高まり) ・削減 (sanitization, 内包的意味の減少) によって特徴づけられる

翻訳物と比較可能なL1テキストの研究（つづき）

Case Study 6

中国語翻訳物におけるアスペクトマーカーの頻度は比較可能なL1中国語データと比較して著しく低かった

こうした研究は翻訳の規範やFrawley (1984)が言うところの翻訳の“third code”を明らかにするのみならず、翻訳者や訓練翻訳者に注意喚起を促せる

翻訳された言語はtranslationeseとも呼べる言語の版を示す

↓

翻訳元の言語の影響はL2データを目標L1言語と知覚的に異なるものにしてしまう

一方向のパラレルコーパスは交差言語学的対照には不向き

↓

両方向のパラレルコーパスを用いることで軽減はされる

↓

調和された両方向のパラレルコーパスが翻訳と対照研究の架け橋に

↓

データの選択には同一のサンプリングフレームが必要

翻訳された言語は目標L1言語とは異なるとはいえ、パラレルコーパスは対象研究の信頼できる基礎とされてきた

- | | |
|-----------------|---|
| James (1980) | 「対照にとって最良の入手可能な基礎」 |
| Santos (1996) | 「対象研究にとって唯一の信頼できる理論」 |
| Mauranen (2002) | 「翻訳言語は使用される自然言語の一部であり、相応に扱われるべき。なぜなら言語は翻訳以外にも様々な方法で相互に影響を与えているからだ。」 |

著者らは疑問視

翻訳の際のソース言語への影響はほかの影響とは違う。通常言語接触は穏やかで翻訳言語におけるソース言語の影響に比べて体系だっていない。ゆえに翻訳言語はせいぜい代表性のない特殊な目標言語の変種だ。この特殊な変種を目標L1言語と混同して対照研究の唯一の基礎とするならば、結果はミスリードになり、第二言語指導・習得に長期的な影響を及ぼす恐れもある。

↓

パラレルコーパスでの発見はそれぞれの單一言語コーパスでのさらなる調査を促す点では、交差言語学対照の出発点として有用

実用的なコーパスの研究

便利で効果的な参照ツールや翻訳者や訓練者の作業台になる
単一言語のコーパスでも有用

Bowker (1998)

コーパス補助の翻訳は主題の領域の理解や正しい用語の選択、イディオム表現の点で慣習的な資源を使ったものよりも高い品質

Bernardini (1997)

伝統的な翻訳指導はLCC (large corpora condordancing) で補い、訓練者が未熟なアマチュアと翻訳者を区別する技能を身に着けることを提案

单一言語コーパスよりも比較可能なコーパスの方が有用

Zanettin (1998)

小さな比較可能なコーパスも、生徒のソース言語の理解や目標言語での翻訳をより流暢に生産する力を伸ばすよう設計された「翻訳者訓練ワークショップ」を考案するのに有用

Friedbichler nad Friedbichler (1997)

領域が特殊なコーパスでは特殊比較可能コーパスが有用。母語と全く異なる言葉を処理する必要があるため。翻訳の問題のために参照可能な比較可能なコーパスがあると翻訳者は母語への翻訳の際に生産性が上がり、ミスが減る。母語から外国語の場合は必要性は急激に増す。

单一言語コーパスよりも比較可能なコーパスの方が有用（つづき）

Gavioli and Zanetti (1997)

肝炎に関するテキストの特殊コーパス翻訳のが、翻訳仮説に確証を与える、領域が特殊な翻訳に関連する問題への解決策を提示する

单一言語コーパスや比較可能なコーパスは翻訳に使えるが、可能性のある仮説を作ることは難しいうえ、一致を確かめるのには時間がかかり、誤りを生みやすく、生産性が下がる

パラレルコーパスの有用性

パラレルコーパスはその点、特定の表現と同値なものに対して、より大きな確実性を提供し、適したツールと共に用いながら他言語で一致する文章についての如何なる表現の存在も見つけることができる

パラレルコーパスは翻訳者や訓練者が用語や語法で正確性を上げる

一つのソーステキストに多数の翻訳があるパラレルコーパスの特殊な使い方

目標言語に直接的な同等物がない言語特徴への体系的な翻訳ストラテジー

Buyse (1997)

フランス語en/yのスペイン語の翻訳。格パラレルコーパスの音声、韻律、形態素、意味、推論的特徴の定量的分析とL1目標言語の比較可能なコーパスでの定量的分析と合わせて解決策を提案

翻訳者育成に関するもう一つの問題が翻訳の評価

パラレルコーパスとソース・目標言語の比較可能なコーパスで構成された評価コーパスが、翻訳指導者が生徒の翻訳を評価し、より客観的なフィードバックを与えるのに役立つ

パラレルコーパスは機械翻訳システム開発の唯一無二の資源

従来のlinguistic rule-based approachに対する異議

Example-based MT (EBMT)、統計的MT等のパラレルコーパス
準拠のものを使用

機械支援翻訳ツール（Computer-assisted translation tool）に
も用いられる

10.7 Diachronic study and language change

(通時的研究と言語変化)

通時的研究はコーパスデータによってのみ扱える分野

この種の研究ではHelsinkiコーパス（ヘルシンキ大）が最も有名

分析は“English in transition: Change Computer Age: Change through variation”プロジェクトにて行われた

- ① Early English on the Computer Age: Exploration through the Helsinki Corpus (Rissanen, Kytö and Palander-Collin 1993)
- ② English in Transition: Corpus-based Studies in Linguistic Variation and Genre Styles (Rissanen, Kytö and Heikkonen 1997)
- ③ Grammaticalization at Work: Studies of Long-term Developments in English (Rissanen, Kytö and Heikkonen 1997)

1000年をカバーするサンプルのみならず、ジャンルや社会言語学的情報もエンコードされている。

Peitsara (1993)

受け身の構文で媒介者を導入する前置詞をヘルシンキコーサスの初期近代/近代英語で比較し、時代の初めではofとbyは同等に頻出していたが、時代の終わりにはbyが優勢になり、15世紀には3倍に、16世紀には8倍になった。特に公的文書・通信のジャンルで顕著であった。

Nevalainen (2000)

初期近代英語では女性の作者が動詞の接尾辞-thを-sで置き換え/主語の位置にyouを使う動きを見せたのに対して、男性は二重否定を単純否定に置き換える動きを見せた