

コーパス言語学から日本語教育へ

深田 淳
パデュー大学

要 旨

本稿の目的は、コーパス言語学の諸分野を、特に日本語研究、日本語教育との関連に焦点を当てながら紹介することである。まず、コーパスの定義とその用途、コーパス言語学の言語観、コロケーションの概念とその重要性、また辞書編纂との関わり、などのトピックをカバーし、その後でコーパスに基づく日本語研究および日本語教育現場での実践を具体例を通して紹介する。

なお本稿は 2009 年 8 月 15 日～16 日に開催された CAJLE 年次大会における講演に基づいているが、内容をさらに吟味し、新しい情報を盛り込んだ上で執筆したものである。

1. はじめに

日本語のコーパス整備が進み、ウェブを使った簡便なアクセスが一部可能になってきたためか、近年のコーパス言語学に対する興味の高まりには著しいものがある。まず 2006 年に『日本語教育』の 130 号で「コーパスと日本語教育—現状と課題—」という特集が組まれたことは記憶に新しい。また 2007 年 10 月には、国立国語研究所が刊行している『日本語科学』がその第 22 号において、『コーパス日本語学の射程』と題する特集を組んだ。さらに「現代日本語書き言葉均衡コーパス」構築が現在進行中であり、2011 年に公開予定であるということがある。日本語では初めての、大規模な均衡コーパスである。このプロジェクトチームには日本語教育班も編成されているので、公開後の活躍が期待される。また、視野を外国語教育まで広げると、2008 年になって Meunier と Granger 編集による *Phraseology in Foreign Language Learning and Teaching* が出版されたことは特筆に値す

る。Phraseology とは、ここではコーパスを使ったコロケーションの研究のことを指している。

本稿は、いくつかの具体的な研究事例を紹介しながらコーパス言語学の諸分野を概観し、最終的に日本語教育での取り組みや実践的な利用方法を考えることを目的とする。

2. コーパスとその用途

コーパス（英語では corpus、複数形は corpora）は、漢語で言えば「言語資料体」となり、大量の言語資料を集めたものをいう。それらの資料を計算機処理することを前提に電子化したものが「電子コーパス」と呼ばれていた時期もあったが、現在ではそれが当たり前となり、単にコーパスと呼ばれる。

コーパス言語学では、コーパスを利用して言語を研究するわけだが、その研究・利用の可能性は、通時的研究から文化論まで大変幅広い。以下は網羅的なリストではないことを断っておく。

● 通時的研究

例えば、各時代ごとに（例：100年周期）幅広いジャンルの書き言葉、話し言葉を集めたコーパスがあれば、通時的に言語変化を調べることが可能である。

● 文体論

文体論では、例えば、新聞では体言止めが多いか、またフォーマルな書き言葉では受動文が多用される傾向があるか、などが研究できよう。

● 外国語教育・習得研究

一般的な日本語コーパスから取り出すことのできる文字、語彙、文法情報は、日本語学や日本語教育の資料になり得る。また、学習者コーパスを分析することで、学習者の誤用パターンや習得パターンを発見したりすることが可能である。

- 対照研究、方言研究

多言語コーパスを使って、複数の言語を対照させて分析することもできる。同じ言語内のバリエーション（いわゆる方言）を対照的に調査することもできる。例えば、英語には、アメリカ英語、イギリス英語、オーストラリア英語、インド英語などのコーパスが存在している。

- 語彙調査、用字調査（言語政策）

漢字などの用字調査や語彙調査は、言語政策の基礎資料になる。大規模な調査は、既にコーパス無しではできなくなっている。

- 語法・文法研究

従来の語法や文法の研究に必要な例文を探すためにコーパスを使ったり、インターネットの検索エンジンを使ったりするのをコーパス言語学と呼ぶ必要はないだろう。ここでは、例えば、統計的な手法を使って、内省では出てこないような新たなパターンを発見したり、新しい言語事実を発掘したりといった研究が考えられる。

- 社会学、文化論

Kennedy (2008) は、イギリス英語において、いくつかの動詞とその目的語のコロケーションを調べた。例えば、**enjoy** の目的語になりやすい名詞をコーパスから抽出しているが、これを例えば、アメリカ英語、イギリス英語、オーストラリア英語、インド英語のコーパスで比較すれば、それぞれの社会で人が何を **enjoy** するかを比較することができ、社会、文化の比較が可能になるだろう。

- コーパス分析、検索ツールの開発

以上で述べたようなコーパスの利用には、必要な情報を抽出、整理したり、新しい発見、発掘をしたりするためのコーパス検索・分析ツールの存在が前提となる。したがって、こういったツールを開発するというのも極めて重要な研究課題である。このようなツールなしでは、いくら巨大なコーパスを構築したところで、まさに宝の持ち腐れである。

3. コーパスの種類

コーパスの種類を内容別に見ると、次のような分類ができる。括弧内は具体例である。

- 新聞コーパス（毎日、朝日、読売、中日など）
- 書籍コーパス（CASTEL/J、青空文庫など）
- シナリオコーパス（CASTEL/J、『寅さん』など）
- 話し言葉コーパス（名大会話コーパスなど）
- 学習者コーパス（KY コーパスなど）
- 多言語コーパス（中日対訳コーパスなど）

学習者コーパスは、学習者の言語産出が格納されたものである。多言語コーパスは、対訳コーパスなど複数の言語で同じ内容の文章が収録されたものである。

構造的な分類としては、均衡コーパス、非均衡コーパスがある。均衡コーパスとは、幅広いジャンルの母集団を規定し、そこから体系的にサンプリングを行なって構築されたコーパスのことである。例えば、現在構築作業中の現代日本語書き言葉均衡コーパスが均衡コーパスの例である。英語では、Brown Corpus、LOB Corpus などいくつもある。

もう一つの分類は、付加情報の種類による分類である。例えば LOB Corpus は、品詞タグ付きコーパスである。他には London-Lund Corpus⁽¹⁾のようになりにかなり詳細にわたる音声表記を含むもの、The Penn Treebank⁽²⁾のように統語情報を含むものもある。

4. 現代日本語書き言葉均衡コーパス

国立国語研究所のコーパス整備計画 KOTONOHA⁽³⁾の一環で、2011年公開予定の略名 BCCWJ で知られる現代日本語書き言葉均衡コーパスは、日本初の大規模均衡コーパスである。規模は、約一億語で、幅広いジャンル、レジスターのサンプルを格納する予定である。

内部構造としては、三部に分かれており、出版（生産実態）サブコーパス（3,500万語）、図書館（流通実態）サブコーパス（3,000万語）、特定目的サブコーパス（3,500万語）から構成される。

幅広いジャンルやレジスターから体系的にサンプリングして大量に集められた言語資料を一カ所に格納した均衡コーパスは、母集団の特性を反映する。これを統計的的代表性（statistical representativeness）という。英語においては、初期から均衡コーパスが整備されていた（Brown Corpus、LOB Corpus など）。

ウェブという巨大なコーパスで検索エンジンを使って用例検索が手軽にできるようになった現在、均衡コーパスが必要なのかという声を耳にすることがあるので、この点について少し述べておきたい。前川（2007）によると、ウェブの問題点は、以下の三点である。

- (1) 検索の再現性がない
- (2) ジャンル・著者情報がない
- (3) 整列条件が非公開

第一点は、実証的研究には欠かせないものである。第二点の著者については、日本語であっても非ネイティブが書いている可能性があるし、自動機械翻訳による文章の可能性さえある。

以上に加えて、

- (4) 執筆時期が必ずしもわからない

という問題も加えておこう。インターネットが普及してまだそれほどの年数が経っていないので意識されないが、年数を経れば時代による語法の差異が出てくるのは必定である。

5. コーパス言語学の言語観

コーパス言語学は、実在する言語資料を対象とするところから、経験主義的、帰納主義的に言語を観る。コーパスに依拠した言語研究に対する Chomsky の批判は有名である。Leech (1991: 204) によると、氏は

Some sentences won't occur because they are obvious, others because they are false, still others because they are impolite.

のような批判をしており、さらに、Kennedy (1998) によると、同氏は「コーパスの中で I live in New York. が I live in Dayton, Ohio. より高頻度で出現したとしても、その事実は言語理論、言語記述に何の関係もない。」

(p.23 筆者訳) とも述べたとある。「可能な文」を規定するのを目的とする限り、いくら数多くの文を集めても意味はない、というところだろう。

しかしながら、言語の創造的側面を強調する Chomsky の立場を批判する者も少なくない。外国語教育に携わる者なら、「教室で教えた規則には則っているが、ネイティブなら絶対にこうは言わない」というような文に毎日のように遭遇していることだろう。Leech (1990) は、通常の内省によるデータ収集だと一人またはごく少数の話者についてのデータしか得られないので、例えば社会言語学的変異の研究では使えないが、コーパスなら多数の話者や著者のデータが一度に得られることを指摘した。さらに氏は、seem と appear、almost と nearly という類義語の対に関して、直感的には同意語だが、コーパスで調べると明らかな使い分けが見られることを明らかにして、コーパスデータの有用性を強調した。

母語話者、特に仮説を立てる立場にある言語学者自身による、かなり恣意的な内省判断に基づく研究方法は近年反省されつつあり、Experimental Syntax という新しいパラダイムが芽生え、発展しつつある (Schütze 1996, Cowart 1997)。これは実験的な手法を取り入れて、多数の母語話者からデータを取り、統計的な分析方法を使用して統語現象の解明を目指すアプローチである。

内省判断と現実との隔たりを示す例として、使役形の「沸かせる」を挙げよう。Shibatani (1973) を受けて、宮川 (1989) は、「沸かせる」という形式が「沸かす」という他動詞の存在にブロックされるため存在しない、と述べている。ところがコーパスを検索すると「沸かせる」の使用例が現代の新聞に多数見られるのである (深田 2007)。さらにコロケーション情報をコンピュータで抽出してみると「沸かせる」が「会場を沸かせる」などのように比喩的な意味で用いられていることがわかり、「沸かす」との棲み分けが明らかになった。

結論的に言うと、いくら大規模なコーパスを構築しても、「可能な文」に迫ることはできないという批判は正論であるが、それだからといって内省判断一点張りでは現実の言語使用を反映しない机上の空論になりかねない。結局、帰納的、演繹的、両方のアプローチから多面的に言語を見ていくのが最善だと思われる。

6. 日本のコーパス言語学

日本では、近年まで、コーパス整備の立ち遅れということが叫ばれていたが、意外にもコーパス言語学の歴史は古い。宮島 (2007) によると、国立国語研究所は 1948 年設立後、最新の機器や統計的手法を用いて語彙調査や方言調査などの社会言語学的実態調査に取り組んだと言う。その成果は、1951 年『現代語の助詞・助動詞』、1952 年『語彙調査—現代新聞用語の一例—』などに残されている。また 1956 年には、計量国語学会が創立され、その機関誌『計量国語学』は 50 年続いており、これは世界に類をみないと言う。1964 年の『分類語彙表』は、国立国語研究所が日本語教育における基本語彙を選定した時に資料として使われたのをはじめ、現代においても頻繁に参照される日本語研究者必携の資料である。また、国語施策のためにも用字調査 (特に漢字) は重要である。これにもコーパス

の利用は必要不可欠になっていると言っても過言ではないだろう。この分野の最近の研究に小椋・相澤（2007）がある。

7. コロケーション

文字と語のレベルを超えた、語と語の結びつきのレベル、いわゆるコロケーション（連語）を次に見てみよう。語とその使用文脈の不可分性については、Firth（1957: 11）の *You shall know a word by the company it keeps.* がよく引き合いに出されるが、Fries や Harris などいわゆる構造主義者の考え方の中に見られた。そしてコロケーションの重要性は、Harris の指導生 Chomsky によって否定され、周辺的な現象に追いやられることになる。ところが、コロケーションはそこで死に絶えたわけではなく、認知言語学では *constructions*、コーパス言語学では *phraseology* として生き続け、*Construction Grammar*（Goldberg 1995, 2003）、*Radical Construction Grammar*（Croft 2001）においては、中心的な役割を果たすこととなる。構造主義というとオーディオリンガル法、そしてオーディオリンガル法というとパターンプラクティスが想起され、批判的に見られることが多いが、「雨が」に対しては「止まる」ではなく「やむ」や「あがる」、「お茶」は「作る」のではなく「入れる」などの組み合わせを教えたり、練習したりするのは十分意味のあることだろう。コロケーションを覚えることは、ネイティブらしい自然な日本語の習得につながるからである。

滝沢（2007）はいくつか興味深いコロケーションの例を挙げている。例えば、「法外に」という副詞は「高い」「高価な」と共起しやすい。「甘い」に続く名詞としては、「球」「香り」「汁」「言葉」「考え」「マスク」などが高頻度で出現するという。「はなはだ」には、「遺憾」「疑問」「厳しい」「残念」「心もとない」「迷惑」などネガティブな意味を持つ語が続く傾向がある。「あながち」は、「とは言えない」「とは言いきれない」と不連続的に結びついている。さらにその間に名詞が入る場合は、

「間違い」「誇張」「的外れ」「暴言」など否定的な響きを持つものが現れるという。

8. コロケーションの重要性

コロケーションの研究者として著名な Sinclair は、コロケーションやイディオムなど語の塊の重要性を Principle of Idiom という形で言い表した (Sinclair 1991)。

Principle of Idiom:

A language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments.

すなわち、自然言語の話者がパターン化されたフレーズや構文を大いに活用していることが大規模コーパスの分析から明らかになってきたということである。Erman & Warren (2000) によると、ネイティブの流暢な文章の約半分は Idiom Principle に則っているという。さらに話し言葉コーパスと書き言葉コーパスの比較から、話し言葉の方が Idiom Principle への依存度が高いとも述べている。

第一および第二言語習得や外国語教育でも、コロケーションの重要性は多くの学者が認めるところである。これらの分野でコロケーションという概念は、holophrases、prefabricated routines and patterns、formulaic speech、memorized sentences and lexicalized stems、lexical phrases、formulas、chunks、constructions などの用語で繰り返し登場している。

外国語教育においては、Michael Lewis が提唱した Lexical Approach がまさにこのコロケーションなどのフレーズを基本としたものである (Lewis 1993, 1997)。文法が基本にあつて、それが作り出す句構造の中のスロットに語を当てはめていく、というような Chomsky 的な考えを否定し、単語およびフレーズを lexis と呼び、それを基本単位とする言語観であり、シラバスも lexis が単位となる。

9. 辞書の編纂

コロケーション研究の成果を有効に世に送り出す一方法として、辞書や文法書への記載がある。Sinclair の COBUILD Project などに見られるように、コーパス構築と辞書編纂は密接に結びついてきた。このプロジェクトから現在、以下の辞書が出版されている。

- Collins COBUILD Advanced Learners
- Collins COBUILD Concise Learner's Dictionary
- Collins COBUILD Student's Dictionary
- Collins COBUILD Active English Dictionary

コロケーション情報を取り入れている英語の辞書は、これ以外にも以下のよういくつかある。

- Longman Dictionary of Contemporary English
- Oxford Advanced Learner's Dictionary of Current English
- Longman Essential Activator Dictionary

特に Longman Essential Activator Dictionary は、学習者コーパスから得たデータに基づき、学習者がよく犯す誤りに注意を喚起するように設計されている。

英和辞典でコーパスに依拠したものには、井上永幸・赤野一郎（編）『ウィズダム英和辞典』（2003）があり、最近ではデュアルディクショナリーと称して、紙の辞書とウェブで検索できる辞書を組み合わせた新形態で提供されている。用例コーパス機能もあり、コーパスはますます身近なものになるだろう。

10. 辞書の評価

日本語の辞書に関しては、『日常日本語バイリンガル辞典』（牧野・中田・大曾 1999）に関して指摘がある（深田・大曾・滝沢他 2002, 深田 2007）。同辞典には、動詞「追う」の動作主は有生名詞であるとの記載があるが、少なくとも受動形においては動作主が無生名詞であることの方が

圧倒的に多い。1991年の毎日新聞一年分を調べたところ、「追う」の受動文 434 例中、418 例（96%）が無生名詞の動作主を伴っていた。例としては、「仕事に追われる」「残業に追われる」「対応に追われる」「サラ金返済に追われる」などが見られた。ここでも、内省のみに基づく記述の問題点が浮き彫りになった。

11. コーパスに基づく日本語研究

日本語でも近年、科学研究費によるプロジェクト「日本語学習辞書編纂に向けた電子化コーパス利用によるコロケーション研究」（平成 13 年度～15 年度、研究代表者：大曾美恵子）をはじめ、コーパスを利用したコロケーションなどの研究が少しずつ始まっている。同プロジェクトは、プロジェクト期間終了後、名大会話コーパスを完成させ、一般公開している。これは、100 時間分の雑談を書き起こして収録したものである。このコーパスはパデュー大学茶漉サイト（深田 2007）において一般に公開されている⁽⁴⁾。

最近の日本語のコロケーション研究として、名大会話コーパスを分析した深田・大曾（2007）があるが、ここでは話し言葉における程度表現についてのみ紹介する。

調査した語は「すごい」「すごく」「大変」「とても」「非常に」「超」である。最多だろうと予想していた「とても」「とっても」は 66 例と少なく、「すごい」「すごく」が最多であった（表 1）。

表 1

語形	出現数
「すごい+形容詞/形容動詞語幹	354 例
「すごい+名詞一般」	214 例
「すごく」	383 例

また「大変」は 409 例中、形容動詞がほとんどで、程度表現は 2 例のみであった。「非常に」は、30 例で文体および話の内容との関係が明らかに見られた。10 例は丁寧体であった。以下は抜粋である。

TK3 : 非常に凝縮された文章ですからね、天声人語は。

KR1 : 非常に怖かったんですね。

「超」は、154 例で、「超+形容詞／形容動词语幹」が 84 例、「超+一般名詞」19 例であった。いくつか例を挙げておく。

IA1 : 私、超混むと思うけど。

WA1 : 上の人は超怖いね。

SE2 : うちのお母さん、うざいって言うと超怒るんだけど。

TK1 : 超一方的なんだね。

KM3 : ねー、超初歩的な会話になるね。

SE2 : *** (うん) 超田舎の方行って。

SE2 : 超味が出てるよ。

12. 日本語教育現場での利用

現場での利用でまず最初に思いつくのが、教授用資料、学習用資料にコーパス研究で得られる知見を反映させていくということである。コロケーションなどの情報も十分に盛り込んだ、使いやすい学習者用の辞書の編纂、教育文法書、教科書やその他の教材の充実が望まれる。さらにもっと身近なレベルでは、新出語の導入の際、コロケーションも同時に教えるなどの配慮ができるだろう。例えば、「匂い」を導入するなら、「嗅ぐ」「する」などとセットにしたいところだ。また「する」を考えると、匂いだけでなく、音や痛みなどの感覚にも使える（「音がする」「腹痛がする」）。

教授法に関して言えば、Bernardini (2004) が学習者用に準備されたコーパス検索ツール (コンコーダンサー) を用いた発見的学習 (discovery learning)、データ駆動型学習 (data-driven learning, DDL) を推奨している。

その根拠には、コンコーダンサーを用いて検索した結果をもとに新出語を覚えると、新しい状況においてもそれを応用できる力がつくという Cobb (1997) の研究結果があるという。日本語教育においては、データ駆動型学習を試みた実践報告に田辺他 (2009) がある。この試みでは日英パラレルコーパスを用いて、キーワード検索をした結果出てくる用例を対訳とともに表示することで、学習者の帰納的、発見的学習を促進しようとした。例えば「実施」という語を学習させる際、「『実施』の前にどんな修飾語が現れるか」「『実施』の後にどんな動詞が現れるか」などの課題を与え、コーパスツールで検索させ、発見に導いた。

一方、広谷 (2007) は日本語例文コーパスと学習者の作文コーパスとを組み合わせ、検索ツールで発見型学習を促進する試みをしている。まず日本語例文コーパスとは、教科書に出てくる例文を集めてコーパス化したものである。一方、学習者の作文コーパスとは、学習者がブログの課題で書く文章を蓄積したものである。この両方のコーパスを自在に検索できるコンコーダンサーツールを用意することで、学習者は自分の文章中の誤りや改善点に気づくことができるわけである。

以上のように日本語教育における実践例はまだ小規模かつ少数であるが、今後の発展が期待できる分野である。

13. コーパスの身近な利用法

前節ではコーパスを日本語コースの中で利用する実践例を見たが、コーパスは学習者、教師が普段からもっと手軽に利用できるものでもある。著作権などの問題があり、すべてのコーパスが一般公開されているわけではないが、いくつか利用可能なものがあるので紹介しておきたい。

最も身近なものは、おそらくインターネット上の検索エンジンによる語や語句の検索であろう。「お茶を濁す」などフレーズレベルの検索は、引用符で囲むなどしないと「お茶」と「濁す」が離れて現れている文章まで

ヒットしてしまうので注意が必要である。検索エンジンごとに色々な検索オプションがあるので、それをよく知った上で使えばよい。ただし第4節で述べたように、ウェブ検索には問題点もあるので、それを理解した上で利用することが必要である。

前述の茶漉では青空文庫コーパスと名大会話コーパスの検索ができるが、品詞や活用形などを指定しての複雑な検索も可能である。また茶漉には用例検索以外に、頻度などの統計およびコロケーションを抽出する機能もある。このサイトは一般公開されている。

さらに上述の『現代日本語書き言葉均衡コーパス』のサイトには、検索デモンストレーションのページがある。コーパスが（有料）公開された後のことはわからないが、とりあえず2009年現在においては簡単な検索を行うことができる。

14. むすびにかえて

コーパス言語学を内包する計算機言語学の一つの特徴は、コンピュータ技術との関係である。技術が大きく進歩することで、今まで不可能だったことが可能となり、また思いもよらなかったことが現実となり、そのためにまったく新しい研究・実践の可能性が広がるということがあり得る。日本語教育におけるコーパス利用は、そういう可能性にも常に目を配りながら発展させていくべき分野であろう。

注

1. <http://khnt.hit.uib.no/icame/manuals/londlund/index.htm>
2. <http://www.cis.upenn.edu/~treebank/>
3. <http://www.ninjal.ac.jp/kotonoha/>
4. <http://tell.fl.purdue.edu/chakoshi>

参考文献

- 井上永幸・赤野一郎（編）（2003）『ウィズダム英和辞典』三省堂
- 小椋秀樹・相澤正夫（2007）「現代雑誌 70 誌における漢字の使用実態と常用漢字表—国語施策へのコーパス活用に向けた基礎研究—」『日本語科学』22, 125-146 国立国語研究所
- 滝沢直宏（2007）「日本語研究・日本語教育とコーパス」
<http://www.lang.nagoya-u.ac.jp/nichigen/0-kyouiku/seminar/2007xian/takizawa.pdf>
- 田辺和子他（2009）「日英コーパスを利用した日本語 DDL 教材開発の試み」『2009 年度日本語教育学会春季大会予稿集』151-156.
- 広谷真紀（2007）「CMC コーパス利用の効果検証」In *Proceedings from The Fourth International Conference On Computer Assisted Systems For Teaching and Learning/Japanese (CASTEL/J)*, 177-180.
- 深田淳（2007）「日本語用例・コロケーション情報抽出システム『茶漉』」『日本語科学』22, 161-172 国立国語研究所
- 深田淳・大曾美恵子（2007）「茶漉で見る日常会話」In *Proceedings from The Fourth International Conference On Computer Assisted Systems For Teaching and Learning/Japanese (CASTEL/J)*, 125-128.
- 深田淳・大曾美恵子・滝沢直宏他（2002）「日本語コーパスからコロケーション情報を抽出するソフトウェアシステム」In *Proceedings from The Third International Conference On Computer Assisted Systems For Teaching and Learning/Japanese (CASTEL/J)*, 37-40.
- 前川喜久雄（2007）「コーパス日本語学の可能性—大規模均衡コーパスがもたらすもの—」『日本語科学』22, 13-28 国立国語研究所
- 牧野成一・中田清一・大曾美恵子（1999）『日常日本語バイリンガル辞典』講談社インターナショナル
- 宮川繁（1989）「使役形と語彙部門」久野暉・柴谷方良編『日本語学の新展開』187-211 くろしお出版

宮島達夫 (2007) 「語彙調査からコーパスへ」 『日本語科学』 22, 29-46

国立国語研究所

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In J. Sinclair (Ed.), *How to Use Corpora in Language Teaching*, 15-36. Amsterdam: John Benjamins.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System* 25(3), 301-315.

Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks: Sage Publications.

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Erman, B. & Warren B. (2000). The idiom principle and the open choice principle. *Text*, 20, 29-62.

Firth, J.R. (1957). *A Synopsis of Linguistic Theory. 1930-1955*. Oxford: Basil Blackwell.

Goldberg, A.E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, A.E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Science*, 7, 219-224.

Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London: Longman.

Kennedy, G. (2008). Phraseology and language pedagogy: Semantic preference associated with English verbs in the British National Corpus. In F. Meunier & S. Granger (Eds.), *Phraseology in Foreign Language Learning and Teaching*. 21-41. Amsterdam: John Benjamins.

Leech, G. (1990). The value of a corpus in linguistic research: A reappraisal, 筧壽雄教授還暦記念論集編集委員会編『ことばの饗宴—筧壽雄教授還暦記念論集』 115-26. くろしお出版.

Leech, G. (1991). The state of the art in corpus linguistics, In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, 8-29. London: Longman.

- Lewis, M. (1993). *The Lexical Approach: The State of ELT and the Way Forward*. Hove, England: Language Teaching Publications.
- Lewis, M. (1997). *Implementing the Lexical Approach: Putting Theory into Practice*. Hove, England: Language Teaching Publications.
- Meunier, F. & Granger, S. (2008). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Shibatani, M. (1973). Semantics of Japanese Causativization, *Foundation of Language* 9, 327-373.
- Schütze, C.T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.