言語研究のデータ獲得３：コーパス

滝沢直宏

Ⅰ　はじめに

経験科学である言語研究の遂行にあたっては，主張を展開するにあたって強固な経験的基盤が不可欠である。その経験的基盤を成すのは，適切な方法で獲得された言語データである。その獲得には内省・直観，インタヴュー，コーパス利用など様々な方法がある。

2023 年度の立命館大学国際言語文化研究所連続講座では，言語の機械処理に加えて，言語研究のためのデータ獲得の方法論を一つの柱に据え，筆者は，他の方法と対比しつつ，特にコーパスの利用について述べた。本稿はその報告である。なお，コメンテーターには，立命館大学・言語教育情報研究科の David Coulson 教授に依頼した。

言語データの獲得について，Chomsky（1965: 18）は以下のように述べている（下線は引用者，以下，同様）。

> Clearly, the actual data of linguistic performance will provide much evidence for determining the correctness of hypotheses about underlying linguistic structure, along with introspective reports（by the native speaker, or the linguist who has learned the language）.

ここでは，内省に基づく報告（introspective reports）と共に，言語運用の実際の資料（actual data of linguistic performance）が，仮説の妥当性を決めるための証拠を提供すると述べられている。

また，梶田（1974: 169）は以下のように述べている。

> このように文というのは抽象物であるから，これを直接に認知することはできない。その認知は，言語に関係のあるさまざまな資料（data）から間接的に行うほかない。この目的に役立つ資料としては，(1) 言語運用の観察，とくに，読み書き聞き話すなど，意味に注意を払って行われる実際の場面での自然な（狭義の）言語運用の観察，(2) 言語についての内省とその報告（たとえば，与えられた文がどの程度自然に受け入れられるかという容認可能性（acceptability）に関する内省や二つ以上の文の間の同義関係に関する内省など），(3) 言語に関する各種の実験（たとえば，種々の文の記憶や理解の容易さの測定），その他がある。これら各種の資料は，いずれも，母国語使用者がその言語能力（つまり修得・記憶している文法）をなんらかの形で使用する活動であるという意味で，（広義の）言語運用の一種と見做して差し支えないものである。

梶田の（1）は Chomsky の the actual data of linguistic performance に相当し，（2）は Chomsky の introspective reports に相当する。（1）のみならず（2）も，「言語運用の一種」であるとしている点に注意する必要がある。加えて，Chomsky は言及していないが，「言語運用の一種」としての「（3）言語に関する各種の実験」も，資料になりうることが述べられている。

これらのデータ獲得方法は，どれも言語研究の遂行に有益・不可欠なものであるが，長所・短所があるので，その利用にあたっては，注意が必要である（滝沢（2017）も参照）。

Ⅱ　言語データ獲得の方法

1. 言語データの様々な獲得方法

前節で引用した梶田の（1）には，読書などによる例文の収集（語法，文法，表現に留意した読書・観察と手作業による例文の収集）が該当する。これには，研ぎ澄まされた言語観察眼が必要とされ，そのためには言語学的訓練が不可欠であるが，それ以外には特に留意するところはないと思われる。

また，梶田の（1）には，書籍・新聞・話し言葉などの資料（狭義の言語運用の結果物）を電子化して，コンピュータで処理できるようにしたコーパスの利用も含めて考えて良い。コーパスについては，その利用にあたって，多くの留意点がある。その点で，読書などによる例文の収集とは性格を異にする。

梶田の（2）には，内省･直観によるデータ獲得が該当する。例えば，よく知られているように，以下の例文（1）の「自分」の先行詞は，「太郎」と「次郎」のうち「太郎」に限定されるのに対して，例文（2）では「太郎」または「次郎」で曖昧である。日本語母語話者であれば，特別な訓練を経ていなくとも，明白に判断できる（しかし，実際に確認すると，大多数の被験者は，この通りに判断するものの，異なる判断をする母語話者が少数ながら存在することには，留意する必要がある）。

(1) 太郎は次郎に写真を自分の家で見せた。
(2) 太郎は次郎に写真を自分の家で見させた。

こうした判断の他に，柴谷（1978: 4）は「母国語話者の直感・内省的知識には次のようなものがある」とし，以下を挙げている。

(a) 与えられた文が日本語の文であるか，そうでないかを識別する直感
(b) 個々の語や，語のグループの文法範疇に関する直感
(c) 文の構造に関する直感
(d) 文の両義性・多義性に関する内省的知識
(e) 同義的な文に関する内省的知識

こうした内省・直観による判断は，明瞭な判断が得られる場合には，有力なデータの供給源

となる。しかし，常に明瞭な判断が得られるわけではない。例えば，以下の 3 つの文の容認可能性判断は，実際に多くの日本語母語話者に確認したところ，かなり判断が揺れることがわかっている。（ある文献によれば，(3) は容認不可，(4) はやや不自然，(5) は容認可能であるという。実際のところ，この直観を共有しない日本語母語話者は一定数，存在している。）

(3) 学生が犬と三人歩いた。
(4) 学生が犬とゆっくり三人歩いた。
(5) 学生が犬とゆっくり公園まで三人歩いた。

筆者の立場は，明白な判断が母語話者の（ほぼ）全てから得られる場合には，母語話者の内省・直観を積極的に利用すべきだが，判断に揺れがある場合には，慎重の上にも慎重に扱うべき，ということである。

他にも，Elicitation（聞き出し），非侵襲的な脳観測，反応時間の測定なども言語研究のための資料の獲得方法として利用可能であると思われるが，本稿では詳述しない。

2.「いささかの疑いも容れないほど明白に」判断できる場合とできない場合

前節で，明白な判断が母語話者の（ほぼ）全てから得られる場合には，内省・直観を積極的に利用すべきと述べたが，梶田 (1974: 170) は，「いささかの疑いも容れないほど明白に文法的（あるいは非文法的）な文（あるいは非文）」と判断できる場合には，その判断を文法研究に対する経験的事実の基盤を提供するものとして扱って差し支えないと述べている。となると，「いささかの疑いも容れないほど明白に」判断できない場合はどうするかという問題が生じてくる。

明白でない要因は多岐にわたるが，「個人間（intersubjective）の揺れ」がまず考えられる。ある被験者は「明白に容認可能と判断する」が，別の被験者は「明白に容認不可能と判断する」という場合である。これは言語の変異（方言差，性差，年齢差，職業の違いなど）に関わるが，特定個人の中では「いささかの疑いも容れないほど明白」な場合もあり，データの扱いは厄介である。

加えて，「個人内（intrasubjective）の揺れ」の問題もある。容認可能か否かを明白に判断できないことは言語研究者が日頃，遭遇する厄介な事態である。「今日は，このように判断するが，明日には，判断が変わっているかもしれない」と思うことも実際に多い。また，「数十回，読んでいたら「容認不可」と思った文も「やや不自然」程度に改善された。もう数十回，読んだら，「容認可」になる可能性もある」という経験をしている言語学者も多い。

実際，インプットの量による質の変化（量質変化）の問題は，言語研究においても重要に関わってくる。以下は，裁判所における裁判官と弁護士のやり取りである。弁護士の発言に注目しよう。

裁判官： 次の日程ですが，月曜の 3 時ではいかがでしょう。
弁護士： 差しつかえです。

この種のやり取りについて，ある弁護士は，自身のブログに「最初こういった言葉を聞いた

とき，「普通に言え～！」と思いましたが今では日常的にそういった言葉を使っている自分がいることに驚きです。」（https://yotsubalegal.com/blog/words-often-used-by-awyers/）と書いている。何度も聞いているうちに，全く不自然ではないように聞こえるようになり，また自分自身も全く普通に使うようになったという事例で，こうしたことはさほど珍しいことではない。

これについては，柴谷（1978: 14-15）の以下の点とも通ずるところがある。

> 文の文法性に関しても，文法家に接しているうちに，判断が徐々に文法家の仮定にそってくる傾向を示すこともある。すなわち，母国語話者を相手に資料を集めているうちに，資料提供者を文法家として訓練してしまうことがあって，文法家に都合のいい，人工的な直感・内省的知識をうえつけてしまうことがある。

容認可能性の判断を求めること自体は，重要なデータ獲得方法ではあるが，その判断から得られた結果の信憑性については，様々な問題があり，慎重に判断する必要がある。また，判断を求める行為自体がその判断を改変してしまう可能性もあるのである。

以上を踏まえると，「微妙に異なる全ての母語話者の頭の中にある文法の積集合（共通部分）」については，完全に一致した明白な判断がいつでも得られ，内省・直観が有効に発揮されると言えるだろう。例えば，日本語における「格助詞は名詞の右側に置く」という一般化に関わる内省・直観判断がこれに該当する。それ以外の揺れが些かでも見られる場合には，よほど慎重な姿勢が必要になってくる。

Ⅲ　コーパス

Ⅲでは，データ獲得方法の一つのしてのコーパスの利用について扱う。

1. コーパスとは何か

コーパスとは「電子化された大規模な言語の資料で，言語の記述や分析の便宜に供され（う）るもの」を指す。（う）を読み込むか否かで，言語研究用に編まれた狭義のコーパスと，それ以外の目的に電子化された広義のコーパス（新聞のデータベースなど）に分かれる。

英語に関していうと，現代英語の代表的コーパスとして，The British National Corpus（BNC）XML Edition：1 億語（書籍 1 冊が 10 万語として 1,000 冊相当）がある。例として，The air was scented now with chocolate, coffee and freshly baked bread.（G0Y.xml）という文を見てみると，このコーパスの中では，この文を構成する各語に，lemma（辞書形）と POS（品詞情報）が以下のように付与されている。

<s n="1281"><w c5="AT0" hw="the" pos="ART">The </w><w c5="NN1" hw="air" pos="SUBST">air </w><w c5="VBD" hw="be" pos="VERB">was </w><w c5="VVN" hw="scent" pos="VERB">scented </w><w c5="AV0" hw="now" pos="ADV">now </w><w c5="PRP" hw="with" pos="PREP">with </w><w c5="NN1" hw="chocolate" pos="SUBST">chocolate</w><c c5="PUN">, </c><w c5="NN1" hw="coffee" pos="SUBST">coffee </w><w c5="CJC" hw="and" pos="CONJ">and </w><w c5="AV0" hw="freshly" pos="ADV">freshly </w><w c5="AJ0-VVN" hw="baked" pos="ADJ">baked </w><w c5="NN1" hw="bread" pos="SUBST">bread</w><c c5="PUN">.</c></s>

品詞情報が与えられているということは，品詞に基づく検索が可能であることを意味する。但し，大規模なコーパスの場合，lemma 情報および品詞情報の付与はそのためのソフトに頼って機械的に行わざるを得ない。ということは，特に英語のようにゼロ派生が頻繁に見られる言語の場合には，付与された情報が誤っていることもまま見られることになる（滝沢（2016）参照）。コーパスを利用するにあたっては，その点も踏まえて利用する必要がある。筆者は，極力，誤りが含まれている可能性があるという理由で，可能な限り，lemma 情報や品詞情報には依拠しないことにしている。

なお，世界初の電子化されたコーパスは，1964 年に完成した The Brown University Standard Corpus of Present-Day American English（100 万語）である。100 万語ということは，普通の厚さの書籍でおよそ 10 冊程度に相当する。わずか 10 冊程度の規模では得られる情報は少なく，言語学的に興味深い例の抽出はほとんど不可能である。今では，iWeb: The Intelligent Web-based Corpus というコーパスがあり，140 億語という規模である。書籍にして 14 万冊相当であるから，コロケーションなどの慣習的な言語現象や周辺的な言語現象を捉える際に有効である。

2. コーパスの用途

コーパスの用途は利用者次第であるが，筆者は，一つには「知らず知らずのうちに使っている慣習的側面」を捉える目的でコーパスを利用している。例えば，副詞 prohibitively は高価を表す expensive や high fees の high との顕著な共起傾向を示しているが，こうしたことを調査する際にも，大規模コーパスは有益である。前述の通り，狭義のコーパスには品詞情報が付与されているのが普通だから，具体的な語を指定せずに，例えば「副詞＋形容詞」の連鎖を網羅的に抽出し，統計の力を借りて，両者の結び付きが強い連鎖を抽出することもできる。visually handicapped, deathly pale, ideologically unsound, racially discriminatory, blissfully unaware などがその例である。広義のコーパスには，lemma 情報や品詞情報が付与されていないのが普通であるが，品詞タグ付けソフトを活用することで，lemma や品詞に関する情報付与を行うことが可能であり，一度，その情報が付与されたら，狭義のコーパスと同様の利用が可能となる。

筆者がコーパスを利用している今一つの目的は，「周辺的言語現象を研究すること」である（田野村（2004）も参照）。英語には，ごく初歩の段階で習う基本的ルールから逸脱した構文がある。例えば，現代英語の普通の散文に，Formal training does not necessarily a good teacher make. のように語順が SOV になっている文がある（SOV 構文。詳細は滝沢（2017））。また，Columns

became blogs became tweets. のように，接続詞を伴うことなく単一の節の中に時制をもった動詞が複数ある文がある（ABC 構文。詳細は滝沢（2017））。こうした例は，周辺的で極めて低頻度ではあるが，英語の中に存在することは疑いを容れない。そうである以上，これらを構文として認め，その特徴を捉えることは，英語の記述をする上で不可欠な作業であり，なぜこのような例外が起こるのかを考えることは言語理論的にも重要である。こうした構文は，極めて低頻度なので，読書で遭遇することは滅多になく，また母語話者の判断も揺れるので，大規模コーパスの利用が考えうる唯一の研究方法である。このように大規模コーパスから一定数の例が抽出されれば，そこから一般性を探り出すことが可能となる。これまでの方法では研究できなかった現象を研究対象にすることができるのである。なお，II.2 節で「インプットの量による質の変化（量質変化）」に言及したが，多くの文を見ていると，SOV 構文や ABC 構文がごく自然な英文に見えてくることは，筆者自身，経験している。

上で例示した慣習的・周辺的事例については，言語知識を脳内にもっている母語話者であっても，自らの言語知識を明確に意識化できるとは限らないと言える。ここから以下のことが導き出される。

・内省・直観のみに依拠することは，時に危険でありうる。
・内省･直観を使って判断する行為は，母語話者が自らの無意識の言語知識を（時に無理矢理に）意識化させる行為と言える。可能な限り，無理なく内省・直観を使うように心掛けることが大切である。
・大脳の中に内在化された言語知識の無意識の発動の結果物（実例）の利用も考えるべきである。コーパスが正にその資料となる。

このような方法を取ることで，母語話者も気付いていない「言語事実」や「傾向」の発見が促される。

3. コーパス利用の留意点と限界（否定的証拠の欠如）

コーパス利用の留意点としては，以下の 2 点を指摘すべきだろう。

・コーパスは，様々な脳内の「文法」の産出物の寄せ集めなので，どのような属性の書き手・話し手によるものなのかを，出典の吟味によって調査する必要が生じることもある。
・コーパスは，どんなに巨大でも可能な言語表現の氷山の一角に過ぎない。したがって，コーパスだけで言語研究を完結させることは原理的に不可能である。

また，コーパスからは，直接否定的情報（direct negative evidence）は基本的には得られない点も重要である。しかし，間接否定証拠（indirect negative evidence）は得られる可能性はある。例えば，以下のような状況があったとする。

・「私は寿司を食べた。」が大規模コーパスに 1 億件あった。

・「は私を寿司食べた。」が大規模コーパスでも 0 件だった。

ここから,「日本語では, 助詞を名詞句の右側に置ける」と断定することは可能である。しかし,「日本語では，助詞を名詞句の左側に置けない」とは断定できない。なぜなら，どんなに巨大でも氷山の一角だからである。しかし,「日本語では，助詞を名詞句の左側に置けないようだ」と推測することは可能である。その確認には，内省・直観による判断（母語の場合）・インタヴュー（内省のきかない言語・方言の場合）という方法をとる必要がある。

以上から，コーパスだけに依拠した研究は危険あるいは無理であると言える。しかし，母語話者が誰もおらず古い記録しか残っていない言語については，どうするのかという問題がある。そうした記録には否定証拠は普通，存在しないだろう。

4. コーパスの有効活用のために必要なこと

コーパスを利用して面白い発見をするには，日頃の読書などで問題意識を高めておく必要がある。また，コーパスを使うには，コンピュータに対して，明示的な指示を与える必要がある。そのためには，当該言語に関する明示的知識そして記述文法・言語理論がコーパス検索の前提となる。

ルイ・パスツールの "Dans le domaine de la science, le hasard ne favorise que les esprits qui ont été préparés."「科学の領域において，偶然が微笑みかけるのは準備された心に対してだけである。」という言葉は言語学にもそのまま当てはまるのであり，心を言語研究に準備しておくことが必要であり，それなしにコーパスの技術的なことを習得しても，コーパスを言語研究に活用することは不可能である。

Ⅳ　コーパスの利用方法：一体的利用と分離した利用

昨今，コーパスと処理ツールを一体化した利用が一般的である。Web サイトでコーパスを使う場合がその代表例である。その場合，コーパス自体はそのサイトを通して利用することになるので，いわばコーパスをブラックボックスとして使うことになる。Web サイトのサーバー上にどのようなテキストがあるのか，どのような付加情報があるのかが全く見えない状態でコーパスを研究利用することは，研究目的にもよるが，危険でありうる。

また，一体的な利用の短所としては，以下を指摘できるだろう。

短所 1：作成者が意図している処理・検索しかできない。求める機能があっても，その機能を実装していない限り，その処理・検索はできない。利用者の要望は無限であるので，どんな優秀・多機能な Web サイトであっても，それに依拠してばかりいては，できることが限定される。

短所 2：Web 上のツールの場合，セキュリティ上の理由などのため，敢えて機能を制限していることもある。

したがって，可能な限り，コーパスをブラックボックス化せず，利用するコーパスがどのように構成されているかに注意を払いつつ，汎用的な処理ツールを使って，コーパスを処理すべきだと考える。そのためには，コーパスを自分で自由に扱うことのできる記憶媒体に載せておく必要がある。

Ⅴ　コーパス利用の事例

Ⅴ節では，コーパス利用の実際例をいくつか示す。

1. 既存の観察の再検討 1：zigzag, flip-flop 型

コーパスの利用可能領域は多岐にわたるが，既存の観察の再検討をする際にも利用できる。英語には，zigzag や flip-flop のように同じ子音が繰り返され，異なる母音が生じるタイプの語は多数ある。このタイプの語には，第 1 母音は i で，第 2 母音は a または o が現れることは，Jespersen（1965: §10.3）や Zandvoort and van Ek（1975：§829）などが指摘している。Jespersen や Zandvoort といったいわゆる伝統文法家の時代は，電子化されたコーパスは存在しない時代であった。この観察が今もって妥当性をもつのか否かは，コーパスによって再検討する必要がある。そこで，正規表現で以下のような指定をする。

正規表現：\b([a-z]+)i([a-z]+)-?\1[ao]\2\b

注：XiYX[ao]Y（X と Y は 1 文字以上の子音字。Y と X の間にハイフンがあっても良い）

このパターンにマッチする語には，以下のような語がある。hip-hop, flip-flop, zigzag, mishmash, crisscross, wishy-washy, chitchat, ping-pong, chit-chat, singsong, tiptop, flimflam, riffraff, tic-tac, knickknack, tittle-tattle, riff-raff, ding-dong, pitter-patter, clip-clop, ticky-tacky, dilly-dally, click-clack（新聞における頻度順）。しかし，言語学・英語学の観点で真に重要なのは，従来の観察では不可とされているパターンを正規表現で指定し，そのパターンにマッチする語が存在しないことの確認である。そのためには，以下のような指定をする。

正規表現：\b([a-z]+)[ao]([a-z]+)-?\1i\2\b

注：X[ao]YXiY（X と Y は 1 文字以上の子音字。Y と X の間にハイフンがあっても良い）

この正規表現を用いて，可能な限り大規模なコーパスを検索することによって，従来の観察の妥当性を調査することができる。実際に大規模コーパス（iWeb など）を調査してみると，依然として，旧来の観察が基本的に維持できることが分かる。

2. 既存の観察の再検討 2：wend と one's way 構文

動詞 wend は現代英語ではさほど頻度の高くない動詞である。この動詞は，以下の文に見るように，one's way 構文で使われることが多い。

（6） As in years past, a bill that would allow couples to divorce is slowly wending its way through the Chilean Congress.（*The Daily Yomiuri,* 2004/1/4, p. 13）

実際，Jackendoff（1997: 173）は，以下のように指摘し，wend は worm と同様，one's way 構文以外では使われないと主張している。

> As with the resultative, there exist fixed expressions built out of the way-construction. For instance, the form *wend* occurs only in the expression *wend one's way PP;* the form *worm* appears as a verb only in the expression *worm one's way PP*.

コーパスが使えなかった時代であれば，このような母語話者の観察はそのまま受け入れるのが一般的であったかもしれない。しかし，今は，母語話者の観察だからといって鵜呑みにせず検証することが重要である。そのためには，以下のような正規表現を用いる。

正規表現：\bwend(s|ed|ing)?␣(?!(my|our|your|his|her|its|their)␣way\b)

注 1: ␣は，半角のスペース 1 つを表している。
注 2: wend(s|ed|ing)? は wend およびその屈折形を意味している。
注 3:（?! ...）は，「所有限定詞 + way が来ない」ことを示している。

これを用いて，コーパスを検索すると，one's way 構文ではない以下のような文に数多く出会う。以下の引用はごく一部である。

（7） After graduating from the medical school of Michigan State University in 1976, Stutes wended west and spent four years at Kaiser Permanente Medical Center in Sacramento.（*The New York Times* 紙）
（8） Water is a symbol of purification and nature in "Nine Songs," which wends through the river of life on its own theatrically imaginative terms.（*The New York Times* 紙）

このような場合の対処法は一様ではなく，少なくとも以下は行うべきである。

分布に偏りがないか出典を詳しく調べる。
複数の母語話者に見せて判断してもらう（インタヴュー）。

これを行って，分布に偏りがなく，且つ複数の母語話者が容認可能と判断するということになれば，少なくとも wend が one's way 構文以外で使えないということにはならない。仮に Jackendoff 教授の判断を仰ぐことができれば，一層良い。その際，もし Jackendoff 教授が（7）（8）などの文を「容認不可」というなら，「個人間の揺れ」の問題ということであり，母語話者の中には wend を one's way 構文以外でも使える者と使えない者がいるということになる。

一方，Jackendoff 教授が（7）（8）などの文を「容認可能」というなら，「個人内の揺れ」で，1997 年当時の判断と変わった可能性がある。あるいは，当初から，自分の意識の意識化がうまくできなった可能性がある（意識と実際の乖離）。

以上のような方法で，確認することが，十全な記述研究には必要になる。

3. 正規表現の効用：日本語のオノマトペの検索を例に

先行する V.1 節と V.2 節では，正規表現の効果的な使い方を見てきた。最後に，日本語のオノマトペの検索を例に正規表現の効用を見る。日本語には「ぽたぽた」「ごちゃごちゃ」型のような繰り返しを含むオノマトペがある。これらを網羅的に抽出するには，以下のような正規表現の利用が考えられる。

正規表現：([ぁ-ん]{2,})\1

注 : ひらがな（[ぁ-ん]）2 文字以上）{2,}）も繰り返し（\1）を指定

この正規表現で日本語のコーパスを検索すると，数多くのオノマトペにマッチする一方で，オノマトペではない「すがすがしい」の「すがすが」，「いまいましい」の「いまいま」にもマッチする。これらは，実際に検索するまでもなく，容易に予測できる非該当例であるが，他にも，この正規表現にマッチしていながら，オノマトペではない文字列は多数ある。例えば，「いよいよ，おのおの，しばしば，ちょいちょい，ひとりひとり，まちまち，ゆめゆめ，わざわざ」などである。ここまでは何とか予想の範囲に含まれるとしても，以下はなかなか思い付かないのではないか。

あぶないあぶない
きたないきたない
しようしようと思っていたが
静かじゃないじゃないか
義務付けするものではないのではないか
これじゃ弁護じゃないじゃないか

正規表現を見て，どんな非該当例が混入するか（同時にどんな該当例を漏らしてしまいそうか）を，コーパス検索の前に熟考することは，言語分析能力を高めるために極めて良い訓練になる。コーパスは言語研究遂行上の情報源であるが，同時にコーパス検索に重要な役割を果たす正規表現に精通することによって，言語分析能力を高められるという重要な副産物もあるわけである（滝沢（2016）も参照）。

Ⅵ　まとめ

コーパスは，内省・直観，インタヴューと並んで言語研究を遂行する上で，重要である。本

稿では，以下のことを述べた。

(A)「いささかの疑いも容れないほど明白に」判断できる場合：積極的に内省・直観あるいはインタヴューに依拠する。コーパスを使うまでもない。
(B) 少しでも怪しいものは，コーパスで調査すべき。それを基に，内省・直観あるいはインタヴューによる調査を行うことも有益。
(C) 複数のデータ獲得方法を，長所・短所に気をつけながら使う。

「証拠」の扱いという点では，刑事訴訟とのアナロジーが感じられる。刑事訴訟における証拠には，自白，目撃証言，物証，各種の鑑定，アリバイなどがあると思われるが，全ての証拠が同じ方向性を向いていれば簡単だが，そうでない場合は，どのように扱うのか，こうした問題も，言語研究における証拠獲得の方法論を考える際に参考になると考えている。

コーパス利用にあたっては，以下の点を述べた。

(D) コーパスには lemma 情報や品詞情報が付与されていることがあるが，機械的に付与された情報なので，全幅の信頼を置くわけにはいかない。
(E) コーパスを利用するには，コンピュータに対して明示的な指示を与える必要がある。そのためには，前提として相当程度の言語学的訓練が不可欠である。
(F) コーパスはどんなに巨大でも，氷山の一角であり，コーパスにおける非存在は，その文が容認不可であることを意味しない。しかし，間接否定証拠は得られる可能性がある。
(J) 正規表現に精通することで既存の説の再検討が可能であると同時に，正規表現に何がマッチするか，何がマッチしないかを熟考することが言語分析能力自体の向上に有益である。

参考文献

Chomsky, Noam.（1965）. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Jackendoff, Ray（1997）. *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

Jespersen, Otto.（1965）. *A Modern English Grammar on Historical Principles. Vol. VI*. London : G. Allen & Unwin.

梶田優（1974）.「変形文法」in 太田朗・梶田優 .『文法論 II』（英語学大系第 4 巻）, pp. 163-667. 大修館書店 .

柴谷方良（1978）.『日本語の分析：生成文法の方法』大修館書店 .

滝沢直宏（2016）.「コーパスからの情報抽出と抽出データの意味づけに関わる諸問題」『英語コーパス研究』23: 45-60.

_____（2017）.『ことばの実際 2　コーパスと英文法』（シリーズ英文法を解き明かす―現代英語の文法と語法 10）内田聖二・八木克正・安井泉（編）. 研究社 .

田野村忠温（2004）.「周辺性・例外性と言語資料の性格 ― その相関の考察 ―」『日本語文法』4,2: 24-37.

Zandvoort, Reinard Willem and Jan Ate van Ek.（1975）. *A Handbook of English Grammar*. 7th Edition. London: Longmans, Green and Co., Tokyo: Maruzen.