

体とした調査環境でも利用されている。また、文字情報の電子化技術の進歩で、分析対象とすべき文字情報の大半は電子ファイル化されている。郵送や留置きで取得した自由回答文を電子的処理が可能な形としたデータも、テキスト型データと考えられる。

b. 自由回答の特質

調査における自由回答文の解析を考えるとき、以下の特質（利点、欠点）に留意した客観的な方法論が必要となる。

まず、利点としては、

- ① 設問選択肢等がない、あるいは影響を受けないので、予期しなかった回答や意見が得られるという期待がある。
- ② インターネット調査等では、回答量（書き込み量）が多くなるといわれている。
- ③ 登録者パネルを用いるインターネット調査等では、再回答率が高くなるといわれている。

ここで②③については、研究がまだ十分でなく慎重な対応が必要である。

一方、欠点としては、

- ① 考えたことがないことには答えにくく、いきなり設問を受けても答えることが難しい。
- ② 無記入が多くなる傾向があるとされている（回答が得られにくい、留置きや郵送であれば書き込みがない、面接であれば回答が得られない等）。
- ③ 他の選択肢型設問やその選択肢の影響を受ける（回答誘導の懸念）。
- ④ 適切なデータ解析法がない、特に事後のアフターコーディング処理の首尾一貫性確保の困難性。
- ⑤ 集計やデータ解析の手間がかかると考えられている。

10.2 自由回答の分析

10.2.1 自由回答とテキスト型データ

a. テキスト型データ

自由回答文・自由記述文として取得した文字情報はもとより、書籍や新聞記事等の文字情報を「テキスト型データ」(textual data) と総称する。情報処理技術の進歩により、日本語の高度な電子的処理が可能となったこと、特に言語情報処理や自然言語処理の諸研究が進んだことから、文章型データあるいはテキスト型データの取得方法や解析手法が多様化している。特に社会調査（意識調査、態度調査等）や市場調査等の分野では、定性情報の分析法の一環として、自由回答・自由記述設問(open-ended question, free response, free answer) の取得方法や取得後のデータ解析方法論が再認識されている。自由回答とは、選択肢型設問 (close-ended question) に対して、選択肢などを設げずに、回答者の自由な反応・意見を引き出す方法をいう。また、自由回答の取得方法は種々ある。たとえば、郵送自記式、留置き自記式、電話調査等の従来型調査の中で用いるだけでなく、近年はインターネット調査（電子メール調査、Web調査）等の電子機器を媒

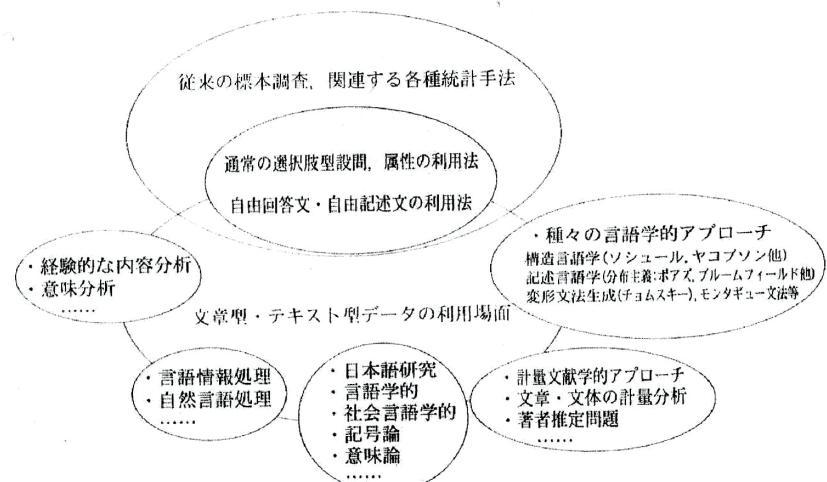


図 10.1 自由回答解析の位置づけ

- ⑥ 回答に均一性を欠くといわれている、あるいは再現性がない。
- ⑦ 設問文の意図がどう反映されたか、自由回答のみからは読み取りがたい。
- ⑧ 面接調査では調査員の不当な解釈、理解違いの意見が書き込まれる可能性がある。など、さまざまな事項が指摘されてきた。

しかし、従来型調査の実施環境の急激な変化（実査の困難性、回収率の低下等）が指摘されるなかで、従来とは異なる意味で、定性調査への関心と期待が高い。特にインターネットの普及により、インターネット調査による自由回答の電子的取得が容易となつたことから、自由回答・自由記述型設問方式を多用する傾向が顕著である。従来の調査法のノウハウとインターネット調査のような新たな調査法との接合面を保持し、また相互の比較検証を進めながら、これと言語関連研究（言語学、自然言語処理）の諸要素とを総合的に組み合わせた、調査における新たな自由回答取得法やそのデータ解析法を実証的に検討する時期にある。図10.1に、こうした自由回答・自由記述文に関する研究と、関連諸研究分野との相互の関連を模式的に表した。

10.2.2 テキスト型データの特徴と形態素解析

a. テキスト型データの特徴

言語類型学的には、言語を孤立語、膠着語、屈折語等と分類する（加賀野井, 1995, 1999）。日本語が欧米の言語と大きく異なることのひとつに、テキスト型データが「べた書き」（膠着言語）であって「分かち書き」されていないことがある。さらに、日本語は漢字、カタカナ、ひらがな、外来語（それにあてられた漢字や仮名等）が混在するという特徴もある。欧米語は単語という単位で仕切られた言語であることから、分析処理はこれを単位として扱うことができるという容易性がある。日本語はこれが困難であるだけでなく、複数の語が連結されて複合語を形成することが多い。

b. 自然言語処理、言語情報処理

テキスト型データに限らず、日本語情報処理解析をコンピュータを利用して行う研究分野がある。たとえば、自然言語処理や言語情報処理では、言語をコンピュータで解析処理する自然言語解析と、言語をコンピュータで生成する自然言語生成等がその研究の中心課題となっている。このほか、全文検索技術、人工知能的アプローチあるいは発想支援ツール開発等の分野がある。こうした分野の研究成果は、形態素解析ツール、構文解析ツール、全文検索システム、文章要約化システム等として具体化されている。

c. 形態素解析とその支援ツール

日本言語処理の重要な技術要素のひとつが「形態素解析」（morphological analysis）である。形態素（morpheme, morphology）とは、表記された文章中に含まれる「最小の有意義な意味ある単位、意味をもつ最小の単位」（池上, 1993）あるいは「単語や接辞など、文法上、最小の単位となる要素」（長尾ほか, 1998）をいう。したがって、絶対的な概念ではなく、あくまでもひとつの便宜的な約束事であり、しかも元来は日本語独自の考え方でもない（池上によると Bloomfield (1933) によって唱えられた概念であ

る）。いずれにしても、日本語データ解析処理においては、何らかの意味で文章をある単位に分けねばならない。このために形態素解析が必要であり、コンピュータ処理機能の向上のおかげで、いろいろな解析方式が提案されている。

d. テキスト型データの分かち書き処理

日本語の処理の一環として、まず文章をある要素単位に分解する「分かち書き」（segmentation）が必要となる。形態素解析の機能のひとつがいわゆる分かち書き処理であり、日本語のような膠着言語の場合、データ処理・解析を容易とするために必須の操作である。

通常は分かち書き処理で得た要素単位の候補を辞書（形態素辞書）と照合し、次にそれを解釈可能な候補に絞り込み、文字・文章の文法的な接続関係（word connection）を検証する。そのうえで、必要に応じてその分かち書き単位についての品詞の同定を行い、統いて辞書にはない語の処理を行う等の手当をする。このような一連の処理過程が「形態素解析」である。したがって、その処理はかなり発見的であり、実際にいろいろな解析方式が提案してきた。たとえば、最長一致法、字種区切り法、文節数最小法、接続規則法等がある（全文検索システム協議会編, 1997）。このように、日本語処理の初めの過程として「分かち書き処理」や「辞書照合」の操作が不可欠となる。

形態素解析を行うツールは多数開発されている。特に、言語処理関連の研究機関やコンピュータ・ソフトウェアの主要企業では、ほとんどといってよいほどこの種の分析ツールの開発を手がけている。たとえば、Juman（京都大学）、紫苑（奈良先端科学技術大学院大学）、QJP（リコー）、BreakFast（富士通）、SuperMorpho-J（オムロン）、Happiness/AiBASE（平和情報センター）等、多数ある。ちなみに、形態素解析を行うとどのような情報が得られるかを簡単な例で示す。ここでは、形態素解析ソフト QJP とテキスト型データ解析システム WordMiner の分かち書き機能（Happiness/AiBASE）を用いた。

形態素解析と分かち書きの例 ある Web 調査（大隅ほか (1999)）で取得した自由回答データであるが、用いる設問は統計数理研究所による「日本人の国民性調査」にある設問に合わせて用意した以下の 2 問である*。

（質問 1） あなたにとって、一番大切と思うものは何ですか。1つだけあげてください（どんなことでもかまいません）。

（質問 2） では、この他に大切なものとして、何がありますか。いくつでもあげてください。

この 2 問への自由回答を併合して 1 文として形態素解析を行うと、表 10.1 の結果を得る。なお、いずれの例も原文は省略した。また、例 1 は分かち書きと品詞の同定結果まるで、出力をそのままに、例 2 は分かち書き結果のみを記号で区切って示した。

この例からも、分かち書き処理が、必ずしも適切に行われるわけではない。例 1 では

* 日本人の国民性調査は面接員による面接法調査であり、設問文の表現も若干異なる。ここにあげる Web 調査で用いた設問は 2 つに分かれるが、国民性調査では第 8 回調査までは 1 問のみ、第 9 回調査(1993 年)からは 2 問を用いている。

表 10.1 分かち書き処理の例

[例 1] QJP を用いた場合

【形態素分割】

目標に向かって一生懸命努力する|こと。家族、や友達それとお金かなお金、にゆとりが|ない。と心が狭く|なる。もの

【単語リスト】

*[1](0)目標	(目標)	(41)名詞
.[2](4)に	(に)	(51)ニ=格助
*[3](6)向かって	〈向かう〉	(15.22)動:五ワ用 b
.[4](12)て	(て)	(55)テ=接助
*[5](14)一生	(一生)	(48)時副詞名詞
*[6](18)懸命努力	(懸命努力)	(89)サ変名詞
.[7](26)する	〈する〉	(60.4)する=体
*[8](30)こと	(こと)	(43)形式名詞
.[9](35)	()	(92)読点
*[10](36)家族	(家族)	(41)名詞
.[11](40)や	(や)	(51)ヤ=格助
*[12](42)友達	(友達)	(41)名詞
*[13](47)それ	(それ)	(44)指示名詞
.[14](51)と	(と)	(51)ト=格助
*[15](53)お金かなお(お金かなお)	(0)	
*[16](64)金	(金)	(41)名詞
.[17](66)に	(に)	(51)ニ=格助
*[18](68)ゆとり	〈ゆとり〉	(0)
.[19](74)が	(が)	(51)ガ=格助
*[20](76)ない	〈ない〉	(21.3)形:ク終
.[21](80)と	(と)	(55)ト=接助
*[22](82)心	(心)	(41)名詞
.[23](84)が	(が)	(51)ガ=格助
*[24](86)狭く	〈狭い〉	(21.22)形:ク用 b
*[25](90)なる	〈なる〉	(15.3)動:五ラ終
.[26](94)もの	(もの)	(54)モノ=終助

[例 2] Happiness/AiBASE を用いた場合

例 2.1

自分が自分らしく生きて行く|事。|その|自分らしさ|を見つける|のに苦労|して|います。|ある|程度|の|収入|も|必要|になつて|くる|し|仕事|に|充実感|も|ほしい|し。|でも|仕事|ばかり|に|夢中|に|も|なつて|いられ|ない|し。|死ぬ時に|ああ生きて|いて|良かった|と|思えたら|それで|いい|と|思つて|います|が|そう|思える|様|に|なる|には|どう|した|らい|いかが|悩み|の|種|です。|家族|、|友達|、|人|と|の|つながり|を|大事|に|思つて|います。

例 2.2

0060 家族。|一人|では|生きて|いけ|ない|し。|何|か|あれば|損得|勘定|なく。|愛情|で|働く|仲間|だから。|長い|年数つき|あつて|いる|親友。|平均的|な|生活|を|送れる|お金。|周囲|に|対する|思いやり|や。|感謝|の|心|。

「一生|懸命努力|する」となつて|いる(ここは誤記)。また、例 2.1 では「ああ生きて|いい|て|良かった|…」で「ああ|生きて」とならず、例 2.2 では「長い|年数つき|あって」では「長い|年数|つきあって」とはなつて|いない。このように、テキスト型データの解析では、まず分かち書き処理を必要とし、しかもどのような方式を用いても処理結果にノイズの混入は避けられない。また、用いる形態素解析ツールに固有のクセもある。しかし、コンピュータの処理機能、特に処理速度の向上で、パーソナルコンピュータでもこの種の処理が十分に可能となつて|いる。

一般に、自然言語処理や言語情報処理では、日本語解析の一環として、まず表記内容の構造を形態素解析、構文解析により確認し、続いて意味的なアプローチから意味解析、意味理解といった操作が行われる。また、最近はシソーラス情報の電子化も進み、用いた単語の重要度や単語間の意味的な連鎖(語彙的連鎖)の分析に用いて、文章の要約化や特徴抽出を行う方法等が提案されている。シソーラスとして、分類語彙表(国立国語研究所)、日本語語彙体系(岩波書店)、シソーラス活用辞典(言語工学研究所)等が電子媒体(CD-ROM)として公開されている。

いずれにしてもコンピュータ処理の支援は不可欠で、しかも通常は相当量の計算処理時間要する。また、自然言語処理的な観点に立つと、その要素技術は言語学的というよりも、さわめて工学的な考え方や研究が多い。一方、調査における自由回答文処理は、設問の内容や主題をかなり絞り込んで、得られる記述内容や記法・表記が乱れることが一般的であり、現状の情報技術では、単語や語、意味内容を確実に同定することはあまり期待できない。また、コンピュータのみによる処理形態が、実際に人が行う言語処理行動(回答行動)を反映させた解析法を提供するかは、現時点の研究成果だけで十分に説得力があるとはいえない。しかも、言語学的観点からは、日本語は流動的な変化や変容が日常的であり、その意味では言語学研究そのものが発展過程にある。

このほか、日本語の曖昧性(本当に曖昧かどうかの議論もある)、デノテーション(語の明示的な意味、表向きの意味)とコノテーション(語の言外の意味、含意)、「テニヲハ」の考慮、カテゴリー論や記号論との関係、最近話題の認知科学的なアプローチからのメタファー(比喩)の重要性等、「日本語の構造的な特徴」をめぐる諸研究や議論が多くのある。また、単純に日常的な生活環境の中でも、ワードプロセッサーの登場による表記法の変化やインターネット利用下における電子メール語(E-mail 語)、チャット語、さらには携帯電話用語(ケータイ語)の登場と、日本語の様相は多様である。

10.2.3 テキスト型データの解析ツール

a. 調査データ分析向きの多次元データ解析ツール

テキスト型データの解析ツールはさまざまな形で提供されているが、ここでは 2 つに大別して考える。ひとつは自然言語処理、人口知能等の研究に依拠する場合である。この種の多くのソフトが形態素解析に始まる一連の言語情報処理的な視点から開発されていて、全文検索システム、文章要約化システム、構文解析ツールあるいはテキスト・マ

イニング・ツールとして提供してきた。しかし、こうしたソフトウェアは、調査で取得する自由回答文のデータ解析には必ずしも適してはいない。

もうひとつは調査データ解析向きの統計システムである。調査環境下において取得した自由回答の分析では、この種のデータに特有の性質まで考えたデータ重視・実践型機能を実装した記述的・探索的解析を設計指針とすることを要件とする。

しかし、自由回答文のデータ処理機能を備え、しかも選択肢型設問・属性情報を相互補完的に併用する調査データ解析に適したソフトウェアはほとんど例がない。稀少な例がフランスで開発された SPAD.T (Système Portable pour l'Analyse des Données, Donnée Textuelles) であり、また “Sphinx Survey: Plus2 & Lexica” (Sphinx Development, 1998) であるが、これには当然日本語処理機能（分かち書き機能、日本語辞書機能等）は含まれない。

したがって、日本語テキスト型データの解析（特に多次元データ解析）を行うには、分かち書き機能ほかの日本語処理機能を付加した独自の統計システムの開発が必要とされる。国内ではこうした試みは少ないが、ひとつの例として WordMiner がある。これは SPAD.T を基本エンジンとし、これに分かち書き処理等の日本語解析に必要な機能を追加した統計システムであり、日仏の研究者とソフト会社との産学協同開発の成果のひとつである。

たとえば、こうした解析システムが行う主な処理機能には、

- ① 入力データ（自由回答文、他の選択肢型設問・属性情報等）の初動探査（データ検証、集計作業、リコード等）
- ② 自由回答文の分かち書き処理による解析用データの生成
- ③ 分かち書きデータによる出現単語の頻度集計、異なり単語頻度集計等
- ④ 分かち書きデータの探査観察と再編集（単語辞書作成、単語の削除と置換等）
- ⑤ 多次元データ解析（対応分析・数量化法類、クラスター化処理ほか）による回答サンプルと抽出単語の関係、選択肢型設問・属性情報と抽出単語の関係等の解析
- ⑥ 回答サンプル、抽出単語それぞれのクラスター化情報の分析

等が含まれる。特に、統計的データ解析に耐えるデータセットを用意するためには、單なる分かち書きだけでは不十分で、細かい処理を必要とする（このあたりが自然言語処理とは異なる）。たとえば、状況に応じて以下のような手当を必要とする。

- ① もとの自由回答文（原文）の確認と事前編集：誤記、日本語変換のミス等、インターネット調査であれば回答の入力中断エラー等の点検や編集が必要である。
- ② 簡単な編集：句読点や特殊記号の削除、必要に応じて助詞の削除等の基礎編集作業。
- ③ 確認済みの原文の分かち書き処理：分かち書き処理は自動化されているが、分かち書きで得た結果の再吟味が必要となる。分かち書きのミス（辞書照合の不一致等）、辞書にない言葉の扱い、人の判断を必要とする判定、類語や同義語のくくり、多義語の扱い等である。つまり自由回答の記述内容が曖昧であることから生じるや

むをえない手当である。

- ④ 分かち書きとキーワードのファイル以外に、分かち書き・キーワードの回答者別利用頻度数等を必要に応じて解析対象として利用する。
- ⑤ これら解析対象ファイルについての統計解析の処理。たとえば、総単語数、異なり単語数（率）等の統計値の算出や比較を行う。
- ⑥ 解析結果の視覚化、グラフィカル表現等。

b. テキストマイニングと関連ソフトウェア

一方、形態素解析ツール、言語情報処理や人工知能研究等の援用を受けて、多数のテキストマイニング関連ソフトウェアが登場している。「テキストマイニング」とは、テキスト型データをはじめ、文書・文章型データのような定性型情報の解析ソフトと、全文検索ツールや文書データベース、データウェアハウス等の要素技術を統合化したソフトウェア体系をいう。いわゆる「膨大な文書・テキスト情報の中から有用な情報を掘り出

表 10.2 単語編集の例（削除と置換）

[削除の例]
、。・：；？！^_一ちゃん、…（）「」十、—&+ 100% のが にをとでなことものはやかです特になしとくにありません別になし
[置換の例]
1 鶏肉=チキン、鶏、若鳥、鳥、鳥肉、鳥肉類 2 漬物=おつけもの、お漬物、漬け物 3 お好み焼き=お好み焼、好み焼 4 ごはん=ごはんもの、ご飯、ご飯類、ごはんもの 5 スパゲッティ=スパゲッティ、スパゲッティー、スパゲティ、スパゲティー、 スパゲティイ、スパッギ 6 丂物=どんぶり物、丂モノ、丂もの、丂料理 7 刺身=おさしみ、お刺しみ、お刺身、お刺身、さしみ 8 何でも=なんでも 9 焼き肉=焼き肉系、焼肉、焼き肉 10 惣菜=総菜、おそうざい 11 天ぷら=てんぷら、テンプラ、天麩羅 12 とんかつ=トンカツ=カツ、カツレツ、トンカツ、とんかつ、ヒレカツ、豚カツ 13 餃子=ギョウザ、ぎょうざ、ぎょーざ、ギョーザ、焼餃子 14 メキシカン=Mexican、メキシコ 15 エスニック料理=エスニック、エスニック料理、エスニック系 16 イタリア=Italian、イタリア、イタリアン、イタリア系、イタリヤ 17 寿司=ごり寿司、バラ寿司、おすし、お寿司、握り寿司、寿司、生寿司、鮓、お鮓 18 安い=安く、安くして、安ければ、安価、安上がり 19 定食=定職 20 家族団欒=家族そろって、家族みんな、家族揃って、家族団らん 21 規則正しい=規則正しく 22 日本食=和食和風=純日本食、純和食、純和風、日本食、和食、和食系、和風 23 唐揚げ=からあげ、から揚げ、とりから揚げ 24 蕎麦=そば=お蕎麦、きそば、そば、ソバ、もり蕎麦、蕎麦、日本そば、ざるそば

す、宝を探り当てる」ことを標語に登場した。最近は、企業の顧客データベース化、文書電子データ化、あるいはその共有化が進んでいる。また、インターネット調査やコールセンターにおける収集データ（顧客・消費者の意見、生の声）の分析、マーケティングリサーチにおけるブランドイメージ分析等、多様な利用場面があるとされる。商用化された例としてDE-FACTO（デファクト）（電通リサーチ・富士通）やそのもととなったHIPS(Hybrid Idea Processing System, 富士通)がある。このほか、AIDE(Augmented Informative Discussion Environment, ATR), KnowledgeOcean (NTTデータ), VextMiner (コマツソフト), ConceptBase/Clustering&Summarizer (ジャストシステム), SurveyAnalyzer(NEC), SynfoWARE/TextMining Server(富士通), TrueTeller(トゥルーテラー) (野村総合研究所)など、多数のソフトが登場している。その多くはかなり高価であり、また広告会社とソフト会社の共同開発等も多い。ただ、その機能や分析手法の内容については暗箱化されており、利用者からは、解析処理の内容が不透明で、結果の信頼性や妥当性について今後の改善や情報開示を求める声が多い。

c. 自由回答の簡単な解析例

ある Web 調査において用いた次の 2 つの設問について取得した自由回答文を併せて分析した(大隅, 2000)。ここで、有効回答者数は 1105 名(男性が 659 名、女性が 446 名)、解析に用いた抽出単語数は 92(語) であった。また、ここでは表示が煩雑となる。

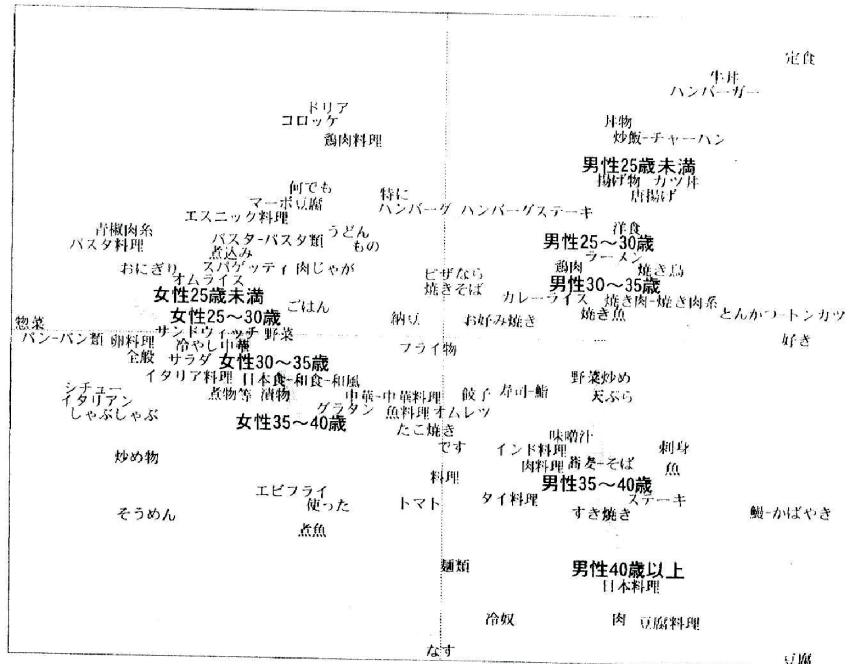


図 10.2 同時布置図：構成要素と性年齢区分の関係

卷之三

とを避けるため自由回答が簡潔な例を選んだが、一般には自由回答文の長さはまちまちであり、しかも長文となることが多い。

[用いる設問]

(質問1) あなたが好きで、よく食べる料理は何ですか？

(質問2) ほかに、あなたが好きでよく食べる料理は何ですか？

単語の編集 まず、分かち書きで抽出した単語群を観察して、これを編集する。たとえば、不要単語の削除、同義語の置換、誤記の訂正置換を表10.2のように行う。ここにはその一部を示したが、削除は句読点、特殊記号、カッコ類などを、また置換は表にあるように、等号の右側の単語群を左の1語で置換する等の処理を行う。

単語と属性（性・年齢区分）の関係 対応分析・数量化法由類を用いて性・年齢区分と単語との関係（同時有置図）を分析し、その結果の一部を図10.2に示す。この図から、回答者の記述した「好きな料理」と性・年齢区分との関係が観察される。たとえば、男性と女性の好みの違い、加齢に伴う食の変化などが見える。このほか、類似の解析手法で、回答と単語、選択肢型設問と単語の関係も探査的に分析できる。

用語検索による回答パターンの観察 探索的にデータの特徴を知る操作のひとつがコンコーダンス（用語検索）あるいはKWICリスト（key-word-in-context list）である。コンコーダンスは、コンテンツアナリシスの基本操作で、指定した単語を基準に原文を検索・ソートし出力する。ある特定の語が、自由回答文の中でどう用いられたかのパターンを知るための機能である。「麺類」をキーとして出力したコンコーダンスの例を表10.3に示した。

10.2.4 適用場面の拡大可能性と留意点

a. 適用場面の拡大

自由回答の処理機能の向上に伴い、他の定性情報分析の方法と合わせて適用範囲が広がる傾向にある。たとえば、消費者行動調査分析（商品モニターによる意見・評価データ）、製品ユーザーの意見聴取、製品に添付の意見はがきの自由回答、グループインタビュー等はもとより、一般書籍の文章解析、小説・文芸作品・新聞・雑誌記事等の分析、論述形式試験問題の評価分析等、さまざまである。さらに、情報技術やインターネットの普及に合わせて、インターネット調査、コールセンターやコンタクトセンターで収集のデータ、テキストデータベースの普及に伴うテキストコーパスの利用、さらには電子カルテの分析と適用分野は広範囲に拡大している。

b. 自由回答文の解析上の留意事項

既存の言語処理技法（形態素解析、特に分かち書き処理）と統計解析の諸要素技術を適切に組み合わせることで、個々の方法論だけでは解決できなかった調査分野のテキスト型データ解析手法としての新たな方向性と適用可能性がみえてくる。

情報技術の普及に連動して、自由回答・自由記述等のテキスト型データの新たなデータ解析手法の研究が期待されている。多数のテキストマイニング手法の登場がそれを示

唆しているが、これらの技法は調査現場の実態からあまりにも乖離している。ここに、長い歴史に裏づけされた従来型調査の科学的方法論の援用を受けた新たな自由回答取得方法の研究が必要とされる大きな理由がある。しかし、自由回答取得方法やそのデータ解析技法の方法論の研究は進んでいるとはいえない。現実の現象解析にあたって最も重要なことは、問題とする対象を説明するために最も適したデータ取得法はどうあるべきか、またその取得データの解析法はどうあるべきかを、データ科学の視点から実証的に検証することである。

〔大隅 昇〕

● 10.2 の参考文献

- 1) 林知己夫 (2000), 「これから国民性研究へ—人間研究の立場と地域研究・国際研究から計量的文明論の構築へ」, 統計数理, 第48巻1号, 33-66.
- 2) 池上嘉彦 (1993), 「意味論—意味構造の分析と記述—」, 大修館書店.
- 3) 加賀野井秀一 (1999), 「日本語の復権」, 講談社現代新書1459, 講談社.
- 4) 加賀野井秀一 (1995), 「20世紀言語学入門」, 講談社現代新書1248, 講談社.
- 5) 小池清治他編集 (1997), 「日本語キーワード事典」, 朝倉書店.
- 6) L. Lebart, et al. (1998), *Exploring Textual Data*, Kluwer Academic Publishers, Dordrecht.
- 7) T. McEnery and A. Wilson (1997), *Corpus Linguistics*, Edinburgh University Press, United Kingdom, Edinburgh.
- 8) 西本一志ほか(1998), 「マルチエージェントによるグループ思考支援」, 電子情報通信学会論文集, D I Vol. J81 D I No.5, 478-487.
- 9) 長尾貞編 (1996), 「自然言語処理」, 「岩波講座・ソフトウェア科学」, 第15巻, 岩波書店.
- 10) M.P. Oakes (1998), *Statistics for Corpus Linguistics*, Edinburgh University Press, United Kingdom, Edinburgh.
- 11) 大隅昇, L. Lebart, A. Morineau, K. M. Warwick, 馬場康維 (1994), 「記述的多変量解析法」, 日科技速出版社.
- 12) 大隅昇 (2000a), 「調査環境の変化に対応した新たな調査法の研究」(報告書), 文部省科学研究費特定領域研究, ミクロ統計データ, 公募研究 (研究課題番号: 09206117).
- 13) 大隅昇 (2000b), 「調査における自由回答データの解析—Info Minerによる探索的テキスト型データ解析」, 統計数理, 第48巻2号, 339-376.
- 14) 大隅昇 (2000c), 「定性情報のマイニング—自由回答データの解析」, ESTRELA, 第74巻, 5月号, 14-26.
- 15) 大隅昇 (2001), 「電子調査, その周辺の話題—電子的データ取得法の現状と問題点」, 統計数理, 第49巻1号, 199-211.
- 16) N. Obsuni (2000), "From Data Analysis to Data Science", *Data Analysis, Classification, and Related Methods*, 329-334, Springer-Verlag, Heidelberg.
- 17) Sphinx Development (1998), Sphinx Survey: Plus2 & Lexica Editions, Software for Surveys, Statistics and Text Analysis, Reference Manual Version 2 for Windows, SCOLARI, Sage Publication Software, London.
- 18) 高橋和子(2000), 「自由回答のコーディング支援—格フレームによるSSM職業コーディング自動化システム」, 理論と方法, 第15巻1号, 149-164.
- 19) 「理もれた文書から“宝”を探す」, 日経コンピュータ, 2001年8月13日号, 40-46.
- 20) 渡部勇, 三木和男, 新田清, 杉山公造 (1995), 「ハイブリッド発想支援システム: HIPS」, 計測自動制御学会第17回システム工学部会研究会資料.

- 21) R.P. Weber (1990), *Basic Content Analysis*, Sage Publications, USA, Newbury Park.
- 22) 全文検索システム協議会編 (1999), 全文検索システムとは何か?, 第1部, 1-63.
- 23) 第10次日本人の国民性調査委員会編 (1999), 「国民性の研究第10次全国調査。1998年全
国調査」, 統計数理研究所研究リポート 83.