

# 人材開発部門の データ活用

「社員意識調査」と「多面評価」を柱として

半蔵門オフィス 代表 南雲道朋

## 第10回 テキストデータの分析の仕方

前回まで、選択式設問の回答データの分析について述べてきました。設問には、問題だということやその改善提案を自由に述べてもらったり、選択式設問で特定の回答をした理由を述べてもらうなど、文章形式で自由に回答させる設問を含めることも効果的です。今回は、そのような「自由記述設問」の設計および回答データ（テキストデータ）の分析方法について述べます。計量テキスト分析（テキストマイニング）手法を用いますが、ここでも、専門のツールに頼る前に、Excelを用いて分析を行うことを重視します。

### 自由記述設問の設計方法

#### (1) 要望の収集でなく現状の把握

自由記述設問は、「社員が何をどう感じているのか」について具体的な情報を得ることで、選択式回答データの定量的分析結果を裏づけるためのものとして有用であるだけでなく、具体的な要望を得ることによって今後の施策検討のヒントとしても有用となります。

そのことから、調査の設問としては選択式設問よりも自由記述設問をむしろ重視し、調査を「社員の要望を吸い上げる機会」として、積極的に活用しようという調査姿勢がとられることもあります。しかし、そのような調査姿勢は、調査が社員の御用聞きになってしまうリスクをはらみ、社員を巻き込んだ調査結果の効果的活用を、かえって妨げてしまう可能性があります。要望を聞けば聞くほど、社員には「要望すれば会社が対応してくれる」という期待が生まれ、会社任せの姿勢になりやすくなると同時に、期待したような変化がみられない場合には、かえって不満や失望が生じやすくなるためです。事務局が組織のあらゆる問題の解決を引き受けるようなスタンスをとってしまった結果、毎年調査を行い、それに基づいて対策を打っても社員の満足度が高まらず、逆に指摘が多くなるばかり、という悪循環に陥るケースさえあります。

よって、自由記述設問を設けるにあたっては、あくまでも「現状

を把握するため」という姿勢を保ち、経営陣、主管部門、現場の管理職、そして社員一人ひとりが、それぞれの立場で主体的に結果を受け止めるようにすることが重要です。そうすることで調査を通じた組織の活性化が期待できます。

#### (2) 建設的な回答を引き出す工夫

自由記述設問には、単なる不満の書き連ねを呼び起こすリスクもあります。それは回答者自身にとっても、不満を表明された組織や対象者にとっても、好ましい影響をもたらさず、調査の効果を引き下げることになってしまいます。よって、「いかに建設的な回答を引き出すか」ということが重要になります。

そのためには、まず「良い点」を聞いて、次に、より良くするためにどうしたらよいかという見地から、「要改善点」を聞くことがポイントです。社員意識調査の場合には、まず「当社の良い点や強みは何か」をたずね、次に「当社がより強くなるために何が必要か」という順番で、多面評価の場合は、「対象者の良い点や優れた

点は何か」、そして「対象者により活躍してもらうために何を期待するか」という順番で設問を設けます。課題解決施策や改善アクションの王道は「強みを活かし弱みを補う」ということであり、そのような見地から、そのまま今後へのヒントとして活用できるようにします。

また、たとえば「経営陣に伝えたいことを自由にお書きください」といった形で1つの設問にまとめてしまうことは、分析の技術的な観点からも避けたほうがいいのです。後述するテキストマイニング（計量テキスト分析）は、文章中の単語の出現数を数えることにより、その回答が「何について」語っているのか、ということを活かすことができます。しかし、それがポジティブな文脈で語られているのか、ネガティブな文脈で語られているのかを機械的に判別することは難しく、それがはっきりしない設問項目では解釈・活用ができません。そのため、ポジティブな（良い）側面を聞く設問と、ネガティブな（改善を期待する）側面を聞く設問とは、別にしておくほうがいいのです。

### （3）分析を容易にする定型化の手法

完全な自由記述回答のかわりに、「定型自由文形式」という、空欄を埋めてもらう設問形式があります。たとえば、「当社のあるべき姿は、（ ）で（ ）で（ ）

であるのに対し、当社の現状の姿は（ ）で（ ）で（ ）である」といった文章の空欄に単語を入れてもらうのです。あらかじめ回答が単語に分解されているため定量的な処理が容易になり、あるべき姿と現状とを対比させる問題分析も容易です。ただし、完全な自由記述回答から得られる「経営陣／対象者への手紙」としての生き生きとしたニュアンスが失われるデメリットはあります。

あるいは、あらかじめテーマを選択したうえで、自由記述回答を記入させる形式もあります。すなわち、顧客対応／商品開発／オペレーション／職場環境／コミュニケーション／働き方／評価と処遇／人材育成など、テーマの分類を設定しておき、テーマを選択してから回答を記入してもらうことにするのです。この方式であれば、単純な集計だけでも、何について語られているのかという傾向を把握できます。ただし、すべてのテーマに対して回答を求めるよう誘導することになってしまい、回答者の負担が大きくなったりと、指摘があったことに対して事務局として何らかの対応をする責任を負うリスクが高まったりするデメリットがあります。

以上から、良い点と要改善点のそれぞれについて、思いついたことをそのまま文章で書いていただき、あとからその内容を分析する

ほうが、社員の考えの把握と振り返りにつながりやすいというのが本稿の立場です。

### （4）回答理由を聞くことの是非

選択式設問ごとに、選択した理由を記入できる自由記述欄を設けておくことで、回答傾向の背景を設問ごとに直接調べることができます。ただし、回答理由の記入を必須にすると回答の負担が極端に大きくなってしまい、必須にしないといわざわざ記入する人が少なくなり、適切なサンプルにならないという難点があります。

そこで、否定的な回答の場合に限って回答理由の記載を求める方法もありますが、そうすると今度は、否定的な回答をしないように誘導することになってしまい、それにも難点があります。

以上から、回答理由を聞くのであれば、特に重点的に分析したい一部の限られた設問にとどめるべきでしょう。

## 自由記述回答の分析方法

### （1）まずは全回答に目を通す

さて、得られた自由記述回答を分析しましょう。ここでまず重要なことは、分析の前に、回答テキストの文章をそのままの形ですべて読むことです。文章での回答を求めた以上、いきなりそれを単語に分解したりせず、まず文章そのものに目を通すことは、回答者に

対する礼儀であり、そのような姿勢を保ってはじめて、調査の主催者と回答者との間に信頼関係が成立するといえます（ただし、回答を依頼する際に「回答は、自由記述回答も含めてすべて統計的に処理し、個別の回答がそのまま情報として用いられることはありません」と明示した場合にはその限りではありません）。

そして実際、自由記述回答の生の文章は、組織メンバーの想いが込められた貴重な情報であるといえます。そこから伝わってくる人や組織の状況のイメージを頭に入れておくことは、選択式設問から得られた定量データの解釈にあたって生きてきます。

それでも、回答が数千件を超えると、全回答に目を通すことは大変な負担となります。結果として全体に目を通すことができないと、全体傾向の理解が少数意見に引きずられたり、自由記述回答の中から有益な意見を取り出すにあたって客観性が保てず、恣意的に都合のよい意見を取り出すことになったりする問題が出てきます。また、生の回答テキストは玉石混交で内容重複が多いことも事実であり、特に経営陣が時間を費やして読むものとしては最善のものではありません。

そこで、テキストマイニング（定量テキスト分析）手法を用いて、情報を圧縮する工夫が必要に

図表1 簡単なテキストマイニング

		全体	働き方	評価	育成	
これらの単語が含まれる回答にフラグを立てて集計		フレックス 時短 在宅 コアタイム	評価 考課	育成 研修 能力 スキル		
回答率：		44%	4%	5%	3%	
回答数：		358	31	41	22	
性別	部門	会社をより良くするために何が必要か				
男	企画	異動の希望を出せるような仕組みが必要。	1	0	0	0
男	営業	業務効率化のための施策を考えてほしい。	1	0	0	0
男	営業	フレックスにしてください。	1	1	0	0
男	企画	目標以外の部分の過程も評価してもらいたい。	1	0	1	0
女	営業		0	0	0	0

なってきます。回答の全体像を定量的・客観的に把握したうえで、個別具体的な内容については回答の本文を読んで精査するという流れをつくります。

## (2) テキストマイニングの基本

テキストマイニングでは、単語の出現頻度に基づいてテキストデータを分析します。一つひとつの文章（パラグラフ）に、あらかじめ取り出しておいた頻出語・重要語が含まれているか（イチ）／含まれていないか（ゼロ）のフラグを立てる、図表1のような形式の表を準備することが、テキストマイニングの基本であり土台です。これによって、自由記述回答の内容を定量的に把握することができます。

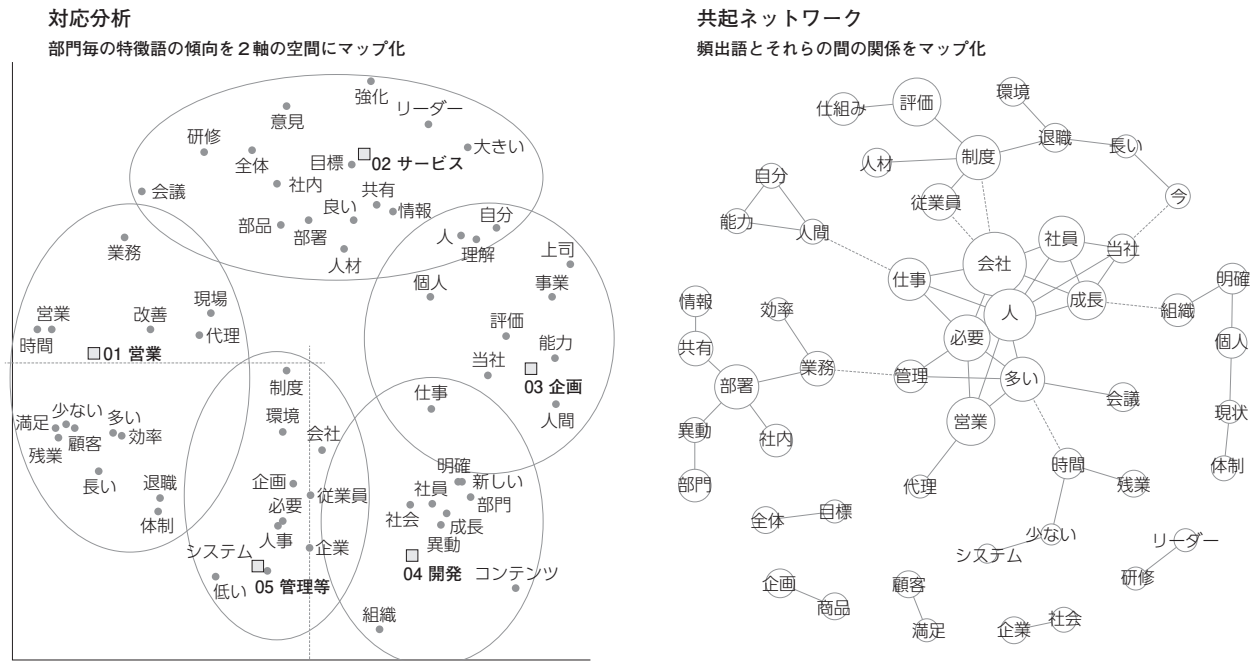
図表1の例であれば、「評価」または「考課」という単語が含まれている回答が全体で41件あり、回答者の5%が言及していることがわかります。いったんこのように数表化することで、どのような

分野への言及が多いかという傾向を1枚のグラフで把握することができます。さらに、男女別や年齢層別、部門別など、属性別の回答傾向もわかります。ここまではExcelで十分に対応できますが、これも立派なテキストマイニングツールといえます。

自由記述回答をこのような表に整理しておくことで、テキスト本文に目を通す際にも、単に順番に読んでいくのではなく、気になるコメントがあれば同じ分野に言及する他のコメントに範囲を広げて目を通す等、効率的な読み方ができます。あるいは、「実名を伴う告発や誹謗中傷」など、そのままのフィードバックに適さない回答の有無をチェックするのも使えます。たとえば、「パワハラ」、「○○さん」といった単語で検索することで、個人名をあげての告発が含まれていないかどうかをチェックできます。



図表2 本格的なテキストマイニング



### (3) 本格的なツールの活用

本格的なテキストマイニングツールも、基本的な原理は変わりません。文章全体を単語に分解し、それぞれの単語の出現数をカウントするとともに、ある単語とある単語が同じ文章（パラグラフ）中に出現する度合いを分析することで、文章の意味内容を定量的に分析します。

日本語テキストの本格的なテキストマイニングツールとして、「KH Coder」という、樋口耕一氏が開発した定番のフリーソフトウェアがあり、誰でも特別な専門知識なしに活用することができます。さまざまな分析機能がありますが、特に「対応分析」や「共起ネットワーク」は効果的で、頻出単語が

登場する文脈や頻出単語どうしの関係をマップ化し、回答の全体傾向を図示できます。図表2は、「会社をより良くするための社員の考え」をマップ化したものです。これらを全体地図として、あとはそれに沿って具体的な回答例を拾い上げていく手順をとることで、自由記述回答の扱いを客観化でき、恣意的に一部の回答にのみ焦点を当ててしまうことを避けることができます。

### (4) 分析結果のまとめ方

自由記述回答の分析結果は、テキストマイニングツールを用いて全体像をわかりやすく示したうえで、その全体像に沿った形で実際の回答テキストをテーマ別に分類し、並び替えて示すことが最善の

方法です。用いるテキストマイニングツールは、簡単なツールでも本格的なツールでも、どちらでもかまいません。

回答テキストを分類する手段として、本格的なテキストマイニングツールの「クラスター分析」を使って、似た回答どうしを自動的にグルーピングする方法もあります。ただし、適切に分類するためには、不要語の削除や同義語の統合といった分析用テキストの下準備に、意外と手間がかかることには注意が必要です。

なお、前回紹介した因子分析と同じく、KJ法のように、まとめたグループに名前やコンセプトを与えていく概念思考スキルが、ここでも有用です。