

# テキスト分析の基本

～自由記述テキストの語のつながり等による分析手法概要 と 事例紹介～

フリーアンサーをなぜ分析するか？

フリーアンサーをどう分析するか？

舘野 昌一  
2017年3月

# 情報の種類

## 1. デモグラフィックス(人口統計学的属性)

年齢、収入、職業

## 2. ジオグラフィックス(地理的属性)

住居あるいは勤務地域

## 3. サイコグラフィックス(心理的属性)

ライフスタイル、行動、信念、価値観

# データの例（アンケート）

	A	B	C	D	E	F	G	H	I	J	K	L
1	SampleID	Q-001 使用	Q3_好きなところ(FA)	Q4_嫌いなところ(FA)	Q5_他に使用している化粧品の理由(F)	Q7_年代(V)	Q8_婚姻(V)	Q9_ファンデーションのタイプ	Q10_購入場所(V)			
2	1	1	じぶんにあってる	たかい	ない	40-49歳	既婚(子供あり)	クリームタイプ	ブランド専門のインターネット			
3	2	1	しっとり肌になじむ感じ、粉っぽく浮かない。	値段が高い	ない	25-29歳	未婚	パウダリータイプ	インターネット 以外の			
4	3	1	安い	ケースがダサい	ない	30-34歳	既婚(子供なし)	パウダリータイプ	ドラッグストア			
5	4	1	毛穴が消える。厚くならない。表情が明るくなる。	長持ちしない。	ない	25-29歳	未婚	パウダリータイプ	ドラッグストア			
6	5	1	あんまりぬらなくてもかばーしてくれところ。	てかりやすい。					デパート			
7	6	1	カバーメイクができる 紫外線カット 率が高い	色が少ない 香りがよい					化粧品専門店			
8	7	1	微粒子	特になし					インターネット 以外の			
9	8	1	使い勝手が良い 海綿でつきやすくべとつきも無く汗で取れる事も無い	添付されている説明書がわかりやすい					化粧品専門店			
10	9	1	保湿効果が高く、パウダーなのにしっとりした仕上がる。	カバー力に物足りない					デパート			
11	10	1	肌の付き具合、伸びが良い。肌がきれいに見える。	特になし					ドラッグストア			
12	11	1	値段はそんなに高くないのにカバー力がある	粉っぽい					ドラッグストア			
13	12	1	自分の肌に合ってる	特にない					ドラッグストア			
14	13	1	無香料でのびがいいところ	カバー力が少し悪い					インターネット 以外の			
15	14	1	しっとりして乾燥しない。ツヤがでる。透明感がある。化粧崩れしない。	以前目の周りなどに使用していたものが廃盤になったこと。					デパート			
16	15	1	ひんじつかどにかくよい。よくフィットする	手軽に使いづらい。					デパート			
17	16	1	薄く付くのに素肌っぽくカバー力もある。	特にない					化粧品専門店			
18	17	1	厚ぼったくならず自然に仕上がる点。	シミソバカスのカバー力がある					大型スーパー			
19	18	1	手軽に使える、UVカットも出来る	特になし					ブランド専門のインターネット			
20	19	1	年中使えるのが便利です	特に不便は感じていない					化粧品専門店			
21	20	1	もちがいい。	特にない。					その他			
22	21	1	安いのに、値段に劣らず、とてもカバー力があったりして、使いやすい	ケースがダサいところ					ドラッグストア			
23	22	1	つきが良い。色が好き	はがれやすい気がする					ドラッグストア			
24	23	1	カバー力があって原材料の成分が肌に優しい	夏場は崩れやすい					デパート			
25	24	1	ノリがいい	高い	ある	30-34歳	未婚	パウダリータイプ	デパート			
26	25	1	質感がよい					パウダリータイプ	ドラッグストア			
27	26	1	さらっとしているところ					パウダリータイプ	ドラッグストア	共あり)		
28	27	1	キラキラした物が入っているのでもって素肌っぽく仕上がる					パウダリータイプ	ドラッグストア	共あり)		
29	28	1	ナチュラルで、時間がたつにつれて自然になる					パウダリータイプ	デパート	共なし)		
30	29	1	のびが良い	なし	ない	25-29歳	未婚	パウダリータイプ	デパート			

使い勝手が良い 海綿でつきやすくべとつきも無く汗で取れる事も無い

保湿効果が高く、パウダーなのにしっとりした仕上がる。

肌の付き具合、伸びが良い。肌がきれいに見える。

値段はそんなに高くないのにカバー力がある

自分の肌に合ってる

無香料でのびがいいところ

しっとりして乾燥しない。ツヤがでる。透明感がある。化粧崩れしない。

人々の思いを集約できれば

使い勝手が良い 海綿でつきやすくべとつきも無く汗で取れる事も無い  
 保湿効果が高く、パウダーなのにしっとりした仕上がる。  
 肌の付き具合、伸びが良い。肌がきれいに見える。  
 値段はそんなに高くないのにカバー力がある  
 自分の肌に合ってる  
 無香料でのびがいいところ  
 しっとりして乾燥しない。ツヤがでる。透明感がある。化粧崩れしない。

人々の思いを集約できれば

# 手法について

分析手法の概要  
テキストの解析手法  
主な出力表示  
分析の流れ

# 分析手法の概要

## 本技術でできること

「コンピューターの解析と人の解釈により、大量のテキストを効率的に集約」  
「お客様の声」を経営に活かすお手伝いをします。

例えば・・・

### ◆お客様の声・社員の声

コールセンター業務、顧客満足度調査、フェアのアンケート、モラールサーベイなど

### ◆市場調査・社会調査（自由記述）

商品マーケティング、広告効果測定、ブランド戦略、Webアンケート

### ◆世論調査・投書・Webへの書き込み

民意・政策立案

等々



大量のテキストデータを持て余していませんか・・・

# 分析手法の概要

## テキストを定量化して解析

数値化が可能な情報は[定量情報]、これに対して人間が自由に記述するテキストは[定性情報]と言えます。

この定性情報を定量化する事で、その内容を解釈する事がテキスト解析の基本的な考え方となります。



定性的な  
テキストを・・・

「意味を数える」



数値解析可能な  
定量情報に。

# テキストの解析手法

## 語に分解し意味チャンク単位に整理

テキストを語に分解して品詞を判別(形態素解析)し、係り語・助詞・受け語のまとまり(意味チャンク)単位に整理します。

## 語の重み付けをカウントし可視化

語同士のつながりの多さや、語の出現頻度などを、コンピュータで解析・可視化。  
解釈支援の為の表示も併せて生成します。

## 原文テキスト例

マナーが悪い人が多いと思います

## 形態素解析

マナー・が・悪い・人・が・多い・と・思い・ます

## 意味チャンクを解析

マナーが → 悪い → 人が → 多いと → 思います

体言 用言 → 連用 → 連体 □ 意味チャンク

# テキストの解析手法

## 機械による解析と人の解釈を融合

言葉だけでは解釈しきれない場合があります。

例えば・・・

「美しい少女の母」という言葉で、美しいのは、少女と母の2通りの解釈が可能です。

従って、テキストをある程度までは機械で解析できますが、最終的には人の解釈を加える必要があり、本技術もその考え方に基づいています。

### 「美しい少女の母」

美しいのはどっち？

- ・少女？
- ・母？

見ればわかるが  
言葉だけでは不明



### テキスト解析

コンピューターによる解析



人の解釈







# 主な出力表示

## 解釈支援表示

### 意味チャンク

テキスト内の[係り語・助詞・受け語]のまとめり(意味チャンク)を、係り語又は受け語毎にその出現頻度順に確認可能。

フォルダー

意味チャンク

[縮める](#) | [伸ばす](#)  
☒ [肌 # \(574\)](#)  
☒ [カバー力 # \(272\)](#)  
☒ [安価 # \(101\)](#)  
☒ [自然 # \(215\)](#)  
☒ [仕上がる \(48\)](#)  
☒ [に \(31\)](#)

なる;自然\_に;;仕上がる;;;;15  
 ナチュラル\_に;;仕上がる;;;;93  
 こまかい;自然\_に;;仕上がる;;;;158  
 いる\_が;自然\_に;;仕上がる;;;;180  
 ナチュラル\_に;;仕上がる;;;;261

### 原文表

多語グラフに出現する語を色付けした原文を、任意の語を含むテキストに絞って参照できます。

原文表

MID	好きなところ(FA)	年代 (V)	肌#	自然 #	色#	値段 #	カバ ー力	自分	きれい#	手ご ろ#	感じ #
326	自分の素肌に近い色である。伸びがよくむらにならない。自然な仕上がりが好き。	40代	0	1	1	0	0	1	0	0	0
430	値段がそんなに高くない。仕上がりが自然。パフが気持ちいい。	19歳 以下	0	1	0	1	0	0	0	0	0
966	薄づきで自然な仕上がりになるところ	30代	0	1	0	0	0	0	0	0	0
1556	ナチュラル感があり、さらりとしている。値段も手ごろ。	30代	0	1	0	1	0	0	0	1	0
1580	色が肌に合っており、自然な感じに着く。化粧くずれしにくく、持ちがよい。	40代	1	1	1	0	0	0	0	0	1
1820	薄付きなのに、肌がきれいに見えるところ、自然な感じに仕上がる場所が好きです。	20代	1	1	0	0	0	0	1	0	1

# 主な出力表示

## 属性による傾向比較

### 比率差グラフ

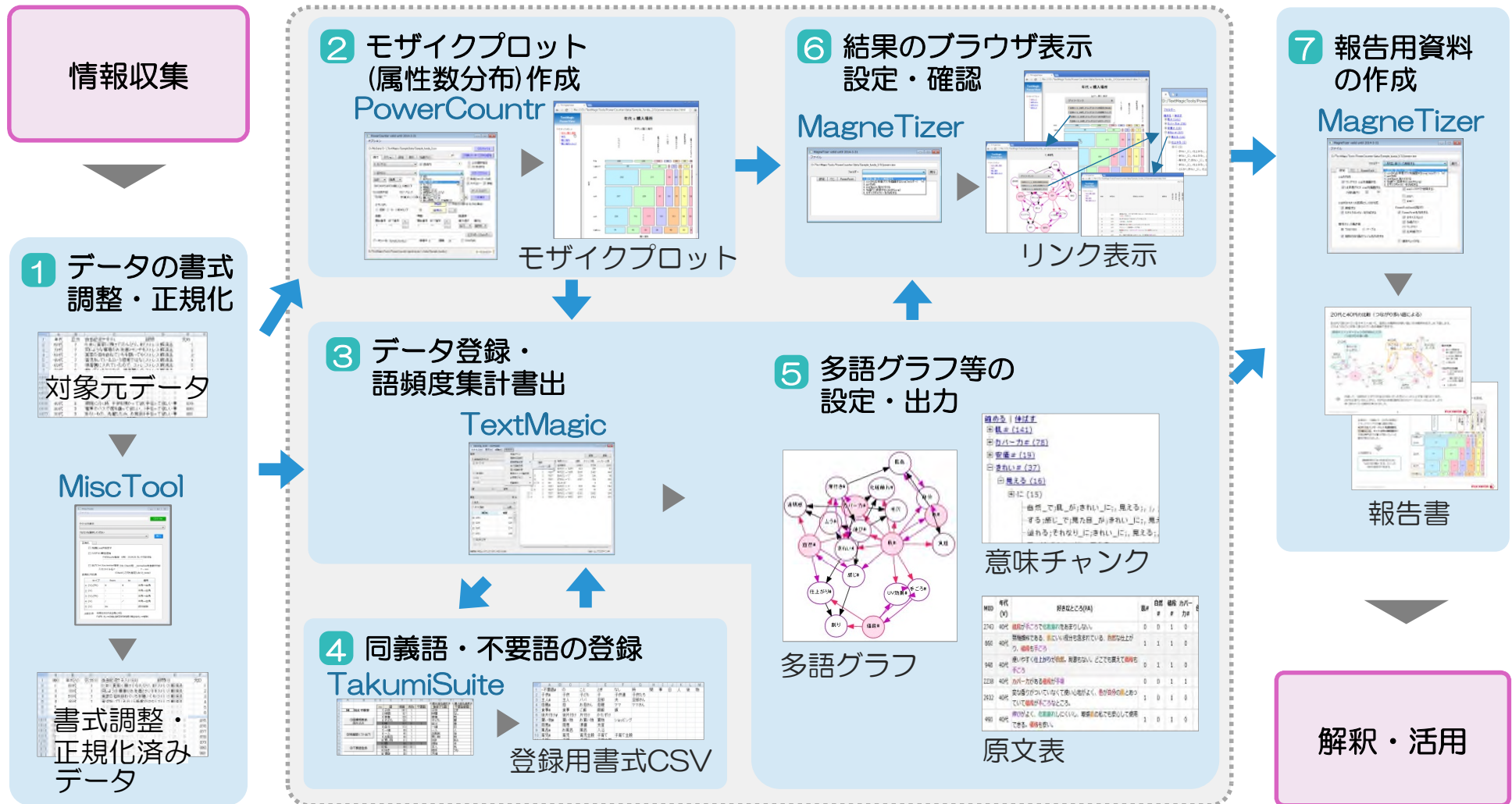
属性による違いを、語の出現比率の全体に対する差で表示します。青色の領域が多いほど、その語が他よりも語られる率が高く、赤色領域が多いほど率が低いことを示します。

比率差グラフ



# 分析の流れ

TextMagicを中心とした富士ゼロックス独自の自然言語分析は概ね以下に示す流れで実施します。



## 表示例

分析対象設定と対象数傾向（モザイクロット）

語の出現頻度による傾向比較（比率差グラフ）

語のつながりによる意味概要（多語グラフ・ウニグラフ）

解釈支援表示（意味チャンク・原文表）



# 分析対象設定と対象数傾向

テキストの分類分けが設定されていれば  
(例えば発言者の年代など)・・・

2つの分類(属性)の掛け合わせに該当する  
テキスト数を、矩形の大きさで表示

## ➡ 分類数の傾向把握が可能

【右の例では】化粧品の購入場所として・・・  
若者はドラッグストアなどで、年配になるほど  
通信販売の利用が増える傾向などが確認できる。

それぞれの矩形もしくは  
縦軸ラベル集計単位で

所属するテキストの  
分析等の表示が可能

## モザイクプロット



# 語の出現頻度による傾向比較

## 比率差グラフ

分類別のキーワードの出現比率と全体における比率との差により、多い場合(+)は青色で、少ない場合(-)は赤色でその度合いを表示

語	全体集合	20代	30代	40代
1 肌#	9.75%	0.17%	-0.22%	-1.03%
2 カバー力#	5.08%	0.73%	0.35%	-0.13%
3 自然#	4.88%	-0.85%	-0.32%	0.13%
4 色#	4.10%	0.68%	0.20%	-0.80%
5 値段#	3.89%	-1.29%	0.21%	0.94%
6 きれい#	3.24%	1.20%	-0.32%	-0.59%
7 伸び#	3.23%	0.19%	-0.20%	0.19%
8 薄付き#	2.48%	0.66%	0.44%	-0.60%
9 感じ#	1.99%	-0.69%	-0.09%	0.14%
10 自分	1.95%	0.03%	0.40%	-0.19%
11 手ごろ#	1.90%	-0.88%	0.20%	0.39%
12 心地#	1.82%	0.44%	0.43%	-0.64%
13 毛穴	1.64%	0.55%	0.10%	-0.58%
14 仕上がり#	1.51%	0.27%	0.03%	-0.21%
15 透明感	1.09%	0.48%	-0.32%	-0.03%
16 化粧崩れ#	0.98%	-0.09%	-0.52%	0.50%
17 リキッドタイプ#	0.94%	0.36%	-0.23%	-0.00%
18 ムラ#	0.79%	-0.38%	-0.03%	0.33%
19 UV効果#	0.79%	-0.11%	-0.13%	0.27%
20 ファンデーション#	0.73%	0.02%	-0.06%	-0.08%

■ 全体に比べて多い      ■ 全体に比べて少ない

➡ 分類別の多く語られている語の傾向確認が可能

【左の例では】  
使用中ファンデーションの好きなところとして・・・

20代では・・・  
色・きれい・毛穴・透明感といった仕上がりの見た目に関する語が、

40代では・・・  
値段が手頃なこと、及び化粧崩れ、ムラ、UV効果等の機能的なことにに関する語が、

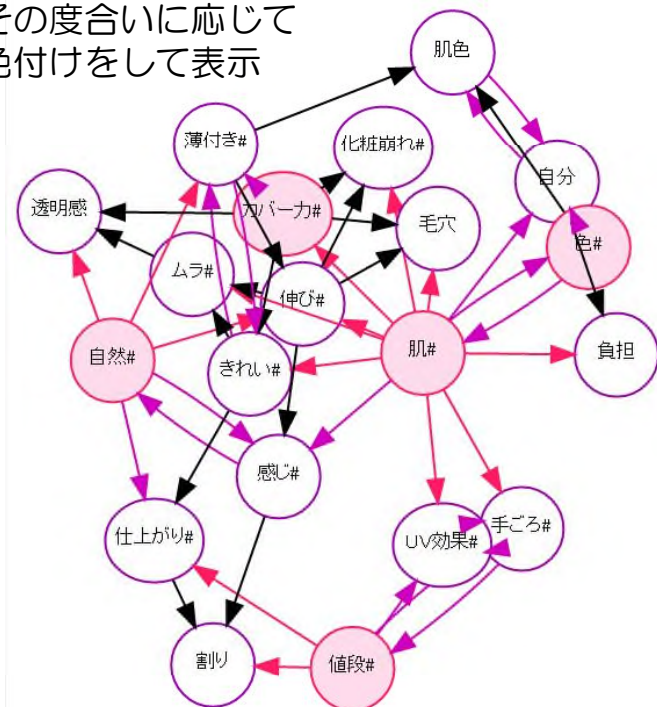
比較的多く語られていることが、概観できます。

# 語のつながりによる意味概要

「使用中ファンデーションの良い所は？」に対し、女性3,000名に自由記述アンケート実施した内の、40代(870名)の回答結果

## 多語グラフ

つながりの多い語とそのつながり方を、  
その度合いに応じて  
色付けをして表示

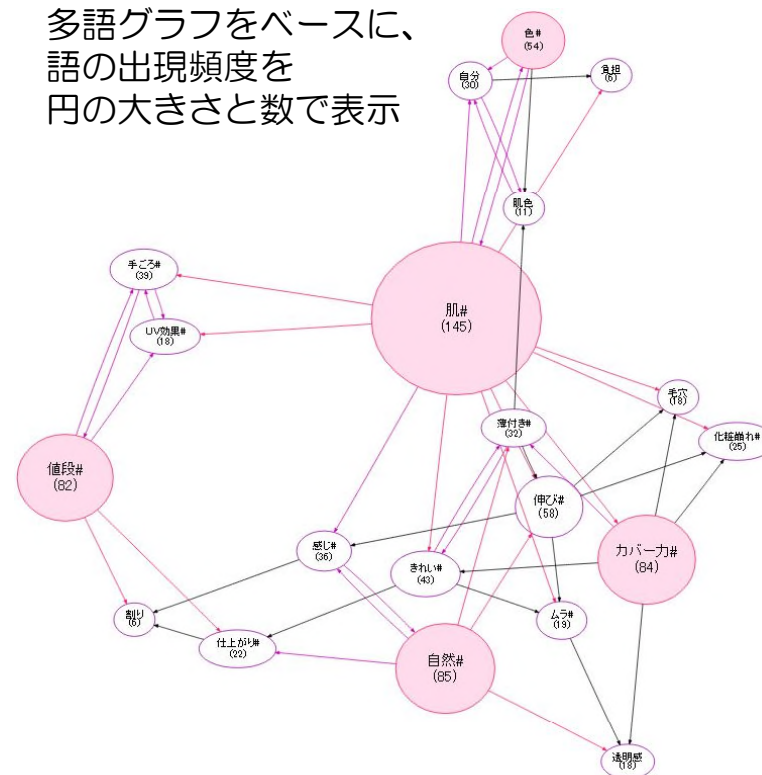


【凡例】

キーワードの重み  
 ● 多くの関係を持つ語のTOP5  
 ○ 多くの次に関係を多く持つ語  
 ○ 上記以外

## ウニグラフ

多語グラフをベースに、  
語の出現頻度を  
円の大きさで表示



キーワードの関連性  
 → ● のノードのTOP3からの矢印  
 ⇄ 一緒に語られることが多い関係を示す  
 → 上記以外

➡ 多く語られている内容の概観が可能



## 解釈支援表示

## 意味チャンク一覧 (語のつながり)

原文表  
(語による絞込)



特徴的な言葉の  
意味概要等を付加

20代女性

☐ きれい# (37)

☐見える (16)

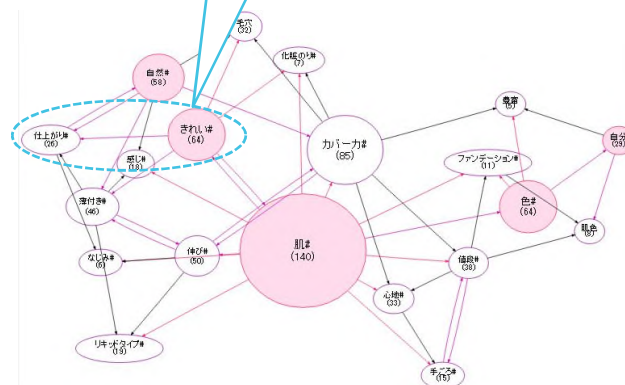
$$\boxed{+} \cdot \boxed{-} = (15)$$

自然\_で;肌\_が;きれい\_に;; 見える;; ; 791

…する;感じ\_で;見た目\_が;きれい\_に;;見える;;;; 954

好きなところ(FA)	肌	色	きれい	自然	自然
		#	い	#	な
薄付きなのに、肌が見えるところ、自然な感じに仕上がるところが好きです。	1	0	1	1	0
仕上がりが自然なところ。素肌よりきれいな見えるところ。	0	0	1	1	0
外装がしゃれている。きれいな肌に見える。	1	0	1	0	0
手軽に肌を綺麗に仕上げられる。使う量が少量ですむので、経済的。粒子が細かくて、軽いので、塗っても違和感がさほどない。	1	0	1	0	0
付け心地がよく、発色がよく、肌が見えるところ。	1	0	1	0	0
ぶっきゃんな肌でも、綺麗に肌に見えるようになるところ。薄付きで自然な肌になる。	1	0	1	1	0

きれいな  
仕上がり



40代女性

□ カバー力# (73)

□ ある (58)

$$\boxed{+} \cdot \text{か} \quad (49)$$

いる;適度;カバ-カ\_が;;ある;;;; 22

細かい;カバ-カ\_が;;ある;;;;26

好きなところ(FA)	肌#	自然#	値段#	カバー力#	色#
薄付きなのにカバー力が有り化粧崩れしにくく私には合っていると思います。	0	0	0	1	0
伸びが良くって薄つきだけで室外カバー力がある。化粧崩れしにくい。	0	0	0	1	0
粒子が細かくてカバー力があり、自然な感じに仕上がる。	0	1	0	1	0
オイルフリー、無香な点。伸びがよく、カバー力もあるところ。値段が手ごろなところ。インターネットでいつでも購入手ができるところ。	0	0	1	1	0
カバー力があるのに自然な仕上がりが好きで気に入っている	0	1	0	1	0
値段が安い割りに適当なカバー力もあって、手間がかからずに使いやすい。	0	0	1	1	0

肌の  
カバー力

