

テキストマイニング・ツール TextVoice : テキストボイスで、

## 太宰治「走れメロス」を

テキストマイニングしてみた

マイボイスコム株式会社（東京都千代田区、代表取締役社長：高井和久）は、独自開発したテキストマイニング・ツール、TextVoice（テキストボイス）を用いて、「走れメロス」をテキストマイニングしましたので、ご報告いたします。

### ■分析データと分析手法

#### 【データ】

- ・青空文庫より 太宰治 著「走れメロス」

[http://www.aozora.gr.jp/cards/000035/files/1567\\_14913.html](http://www.aozora.gr.jp/cards/000035/files/1567_14913.html)

#### 【分析手法】

- ・TextVoice（テキストボイス）を用いた、テキストマイニング  
 サマライズ : 最大6つの言葉のつながりで内容を分類  
 セグメント : 属性ごとの関心点を抽出する属性分析

※TextVoiceには上記の他にも2つの分析機能があります。

- フォーカス : 強い言葉のつながりで、注目ポイントを抽出
- パースペクティブ : テキストの背後に隠れた意味を見つけ出すマッピング

### ■「私にはいのちの他には何も無い」「よくも私を信じてくれた」

サマライズ分析で内容をパターンに分けると、下記のような内容が抽出されました。

図1 サマライズ 出現量上位を抜粋

順番	出現量	ピース1	ピース2	ピース3	ピース4	ピース5	ピース6
J01	11	他	ない	何	命		
J02	10	私	信じ	くれた			
J03	7	殺し	人	自分			
J04	6	おまえ	わし	心	疑うの	くれた	
J05	6	すぐ	出発	市			

#### ※サマライズ

- ・どんな内容があるのか、最大6つのピース（言葉）の組合せで分類し、内容理解。
- ・該当する、ももとの原文の件数を数えて、ボリューム把握。原文の確認も可能。

■出現が多くなった内容を見ると、〈他には何も無い〉に代表されるような、「ない」ことや、否定を強調している文が多くあるようです。

例えば「他には、何も無い」「私にはいのちの他には何も無い。」「泳ぎ切るより他に無い」「私の命などは、問題ではない」「走るより他は無い」などです。

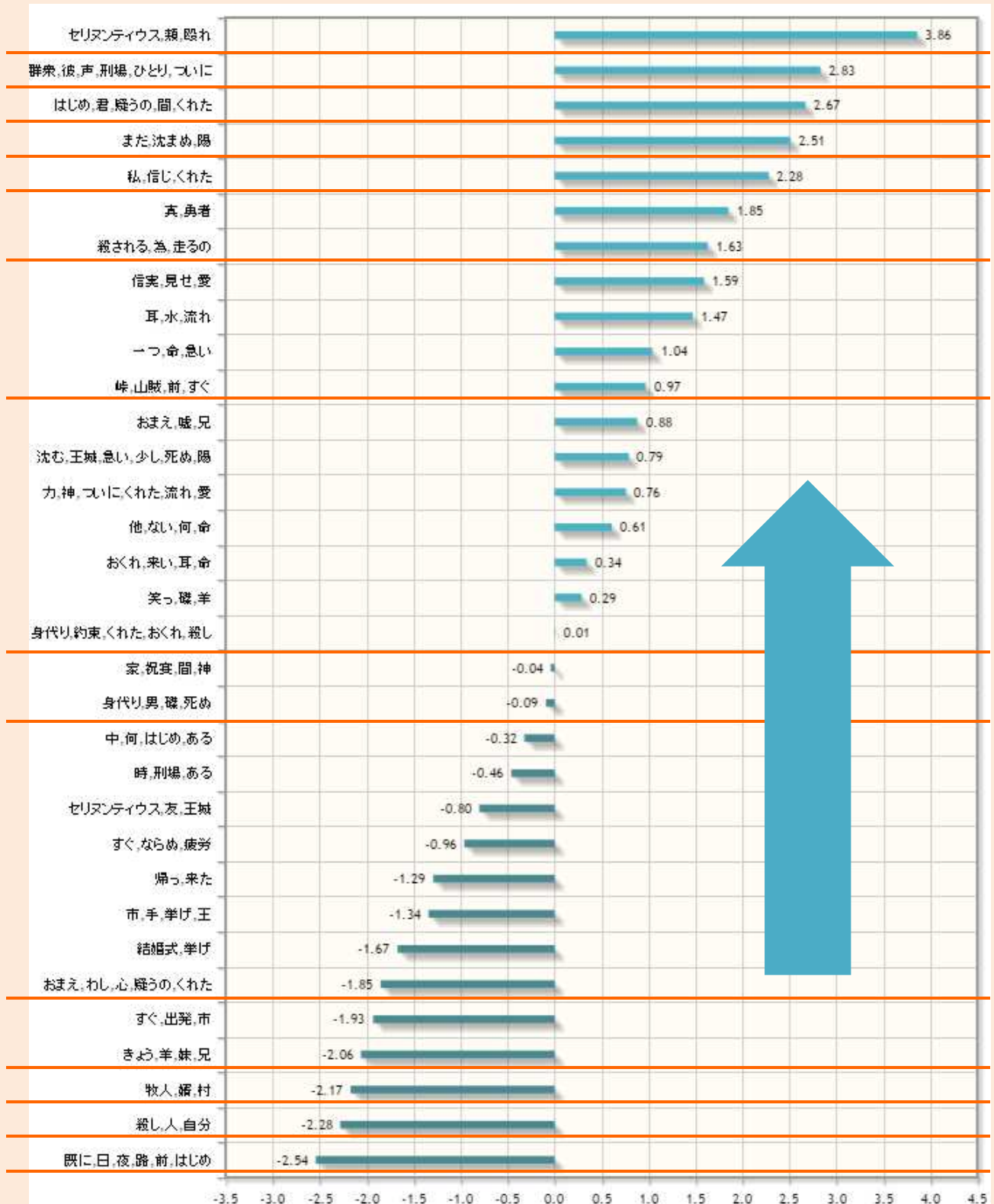
また、〈私を信じてくれた〉のような内容もポイントのようです。具体的には「よくも私を信じてくれた」「君だけは私を信じてくれるにちがいない」「私は、信じられている」などが挙がりました。

## ■クライマックスは、〈セリヌンティウス、頬、殴れ〉

次に、だいたいのストーリーの流れ、あらすじを、原文を読まなくても何となく把握できるようにできないかと考え、〈セグメント〉分析を応用して、作品の最初から最後に向けて、出てきやすい言葉のつながりを並べてみました。

次のようなグラフでその傾向が分かります。グラフの上にある表現ほど、作品の後半にあり、グラフの下にあるほど、作品の前半にある表現の傾向ということを示しています。

図2 セグメント 上になるほど、作品後半の表現パターン



グラフを下から見てみると、夜のシラクスの路を歩いていることを表す、<既に、日、夜、路、前、はじめ>から始まり、人を殺すことがテーマのひとつになっていることがうかがえる<殺し、人、自分>と続きます。具体的な文を上げると、

「もう既に日も落ちて、まちの暗いのは当りまえだが、けれども、なんだか、夜のせいばかりでは無く、市全体が、やけに寂しい」

「路で逢った若い衆をつかまえて、何かあったのか、二年まえに此の市に来たときは、夜でも皆が歌をうたって、まちは賑やかであった筈だが、と質問した」

と、不吉さを暗示する文章が続き、答えを求めて、近くにいた老人に質問したところ、返ってきたのは、

「王様は、人を殺します」

の一言でした。

その後で、<おまえ、わし、心、疑うの、くれた>に表れる、「疑うのが、正当の心構えなのだ、わしに教えてくれたのは、おまえたちだ」

という王様の発言や、妹と牧人の結婚式が行われることを示す表現が続きます。

続いて、<身代り、男、磔、死ぬ>や、<身代り、約束、くれた、おくれ、殺し>では、身代わりというテーマがあらわれ、具体的には、

「人は、これだから信じられぬと、わしは悲しい顔して、その身代りの男を磔刑に処してやるのだ」  
「そうして身代りの男を、三日目に殺してやるのも気味がいい」  
「おくれたら、その身代りを、きっと殺すぞ」

といった、王様の発言がみられます。

また、物語の中ほどでは、峠で山賊も登場し、盛り上がりつつ、山場に向かうことがうかがえます。

最後に近づくと、  
<殺される、為、走るの>（「殺される為に走るのだ」「身代りの友を救う為に走るのだ」）  
<私、信じ、くれた>（「君だけは私を信じてくれるにちがいない」「私は、信頼に報いなければならぬ」）  
<まだ、沈まぬ、陽>（「いや、まだ陽は沈まぬ」「私は陽の沈まぬうちに王城へ行かなければならぬ」）

そして最後には、

<群衆、彼、声、刑場、ひとり、ついに>  
（「と大声で刑場の群衆にむかって叫んだつもりであったが、喉がつぶれて嘎れた声か幽かに出たばかり、群衆は、ひとりとして彼の到着に気がつかない」）

<セリヌンティウス、頬、殴れ>  
（「セリヌンティウスは、すべてを察した様子で首肯き、刑場一ぱいに鳴り響くほど音高くメロスの右頬を殴った」「メロスは腕に唸りをつけてセリヌンティウスの頬を殴った」）

でグラフは終わります。

このように、TextVoiceを使えば、ポイントになる内容を抽出してすることができます。

## TextVoice（テキストボイス）のご紹介

### ■ 導入事例 ～利用されたお客様のご意見～

- ・ 課題発見に期待できそう（調査会社様）
- ・ アウトプットに納得感がある（調査会社様）
- ・ 肌感にあったものが出てくる（飲料メーカー様）
- ・ 手間の軽減にもなるし、深い結果を引き出せていると思う（食品メーカー様）
- ・ 最大6語のつながりがとても分かりやすい（食品メーカー様）
- ・ アウトプットが分かりやすいので、そのまま報告書に使える（食品メーカー様）
- ・ チューニングなしでも、素早く、ざっと傾向が把握できて、納得感もある（情報関連会社様）
- ・ 操作が簡単で、分かりやすい（コンサルティング会社様）
- ・ セグメントでの傾向確認が分かりやすい（広告代理店様）
- ・ このようなアウトプットが出るのであれば、テキストを分析するのも有効だと思う（製造業様）
- ・ テキストをきれいに表現できるのがいい（IT関連企業様）
- ・ 目視を軽減できる（IT関連企業様）
- ・ 分析結果を関係者で共有して、納得が得られる（食品メーカー様）
- ・ 自分たちの分析目的、分析の方向性、レベル感にあっている（製薬会社様）
- ・ 社内で課題として感じていたことが、ピタリピタリと抽出されたことに驚いた（製薬会社様）

などのお声をいただいております。

