

# 言及と志向：ヒトが {未来/過去} を想うとき

久保 圭<sup>1</sup> 仲村 哲明<sup>1</sup> 白田 泰如<sup>2</sup> 荒牧 英治<sup>1,3</sup>

<sup>1</sup>京都大学 <sup>2</sup>JST さきがけ

<sup>1</sup>kubo.kei.3r@kyoto-u.ac.jp, <sup>2</sup>tetsuakinakamura8@gmail.com

<sup>3</sup>usuda.yasuyuki.23n@st.kyoto-u.ac.jp, <sup>4</sup>eiji.aramaki@design.kyoto-u.ac.jp

## 1. はじめに

人間があることについて言及しているとき、その言及の内容と頻度によっては、その言及の背景にある心情や意図がどのようなものであるかを推定することは、そう難しくなく思われる。たとえば、誰かに執筆されたテキストの場合も、どのような言葉を用いているか、丁寧さの度合いはどのくらいか、どのようなことについて強調、あるいは繰り返して言及しているかなど、判断材料は多岐にわたるが、執筆者の心情を推定することは十分に可能であるだろう。

この点については、言語処理の分野においても、テキストから執筆者の心情を推定する技術が成熟しつつあり、さまざまな研究がおこなわれている。では、単体としての執筆者ではなく、執筆者の集団が属する国や言語圏でどのような事柄について頻繁に言及されているかを調査すれば、その国がどのような側面をもつかを推定することも可能ではないだろうか。

本研究では、テキスト上での世界と現実世界の間には何らかの関連性がみられるという考えのもとに、まず、日本国内や世界の国々で言及される《過去》と《未来》を表す語の使用頻度比について調査をおこなった<sup>1)</sup>。そして、その調査結果が現実世界における何を説明し得るものであるかについて分析をおこなった。

## 2. 本研究の位置づけ

### 2.1 センチメント分析：商品から言語圏へ

言語処理の発達に伴い、商品のレビューや個人のブログといったテキストから、特定の商品の評判を得る技術が成熟しつつある。このような技術はセンチメント分析 (sentiment analysis) とよばれ、テキスト執筆者の心情を推定するものとして位置づけられる[1]。

さらに、単体の執筆者を対象とするのではなく、SNS における執筆者集団でのやりとりを解析した研究[2]や、ある国において事件が発生したときの当事者の感情を解析した研究もおこなわれている[3]。

このアイデアを発展させると、ある言語で記述されたテキストを集めることで、その言語圏に関する情報を収集することも可能であると考えられる。

本研究は、日本語や英語といった言語における統計か

ら、執筆者の属する言語使用圏の情報を得ることに挑戦する研究として位置づけられる。

### 2.2 言及頻度と志向

現実世界での時間は、現在から未来に一方的に流れており、当然、過去に揺れ戻ることはない。しかし、思考という行為においては、我々は過去について想起し、未来についても自由にイメージすることができる。

ここで、人間の思考と言語が相互的に深く結び付くと考えるのであれば、未来のことをよくイメージするときには「未来についての言及」が多く、過去をよく想起するときには「過去についての言及」が多くなるであろうと推測できる。すなわち、過去や未来の言及頻度の多寡から、その人間の思考の方向性 (つまり志向) や心的な変化などを読み取ることも可能であるともいえる。

この考えを個人レベルから国レベルに拡大して適用した場合、各国において過去または未来のどちらについて多く言及されているか (以下、これを言及頻度とよぶ) を調査することで、その国の現状や変化、あるいは国家間における何かしらの関係性を読み取ることも可能ではないだろうか。

この言及頻度に着目した研究の一つとして、世界 45ヶ国の年号の言及比を調査したものがあ[4]。彼らは以下の定義に基づき、 $n$  年未来の年号と  $n$  年過去の年号の言及比 (Feature Oriented Index) の調査をおこない、FOI と GDP に  $r=0.53\sim 0.64$  の相関がみられたことを報告している。

$$FOI = \text{freq}(\text{year}+n) / \text{freq}(\text{year}-n)$$

本研究では、年号ではなく、《過去》と《未来》をあらゆる語の使用頻度比について調査をおこなう。

## 3. 語の選定と調査方法

### 3.1 語の選定

どのような言語においても、人称代名詞 (日本語では「私/君/彼女」など) や基本的な身体動作や生命維持に関わる行為 (持つ/歩く/食べる/寝るなど) をあらゆる基本語彙は共通して存在する。それは他の品詞においても同様であり、「水/空/人間/猫」といった名詞も普遍的にみられるものと推測される。

また、各国の文化や慣習は個別的であっても、起床することで一日がはじまり、何らかの活動や思考をおこない、一日の終わりに睡眠をとるといった基本的な生活パターンに関しては、大差ないものと考えられる。

では、世界中の人々が、上記のような生活パターンに沿って暮らし、それにともなって生じる思考を表現するために言葉を発すると考えるのであれば、基本語彙の使用率はどの言語圏においても一律であると考えられる。しかし、実際には、基本語彙の使用率は言語ごとに異なりがみられる。その主な理由として、各言語に固有の性質が考えられる。

基本語彙の使用率は、各言語における文法的要請や、ある語について同義語がどの程度の多様性をもって存在するかなど、その言語に固有の性質によって影響を受ける。例えば、日本語においては、人称代名詞を含む主格はたびたび省略が可能であるが、英語ではそのような省略は基本的に認可されず、主語を明示することが要求される。また、日本語では一人称代名詞のバリエーションが多く、最も基本的である「私」の他にも「僕／あたし／自分」などがあり、発言者の性別や年代、丁寧さの度合いによっても使い分けられ、それぞれが広く用いられている。しかしながら、英語の場合は一人称単数の主格は「I」にほぼ限られ、yours truly などの表現は存在するものの、ほぼ用いられない。

このような個別的性質を考慮すれば、日本語における「私」の使用率と英語における「I」の使用率を比較した場合、前者が少なくなると予想される。

また、ある語の使用頻度を複数の言語間で比較しようとしても、各言語の母語話者数が異なっているため、絶対頻度による比較は意味をなさない。よって、まず調査対象とする語  $x$  と別の語  $y$  との割合を調査し、それ  $x/y$  によって比較をおこなうという手続きが必要となる。

この  $x$  と  $y$  にどのような語を当てはめるかについては、当然、上に挙げた言語個別的な性質による影響を受けないように選定する必要がある。

本研究では、過去をあらわす表現の一例として《昨日》を、未来をあらわす表現の一例として《明日》を調査対象として選定し、それらの言及頻度比（以降では TYR: Tomorrow / Yesterday Ratio とよぶ）を調査した。

$$TYR = \text{freq}(\text{tomorrow}) / \text{freq}(\text{yesterday})$$

この《昨日》と《明日》は「過去」と「未来」を端的にあらわす基本語彙であることに加えて、同義語がほぼ存在しない。また、これらの語を必要に応じて省略することを認可する言語的制約があるとは考えにくいことから、本研究の目的において適例である。

### 3.2 調査方法

本研究では、検索サイト Google<sup>2)</sup>を用いたある語の完全一致検索によって得られた検索ヒット数をその語の使用頻度として定める。さらに、言語と国を限定して検索

をするという目的により、検索オプション内の「検索結果の絞り込み」機能における「言語」と「地域」を用いた。例えば、日本における TYR を調査する場合、「昨日」と「明日」を「言語：日本語」、「地域：日本」に設定して完全一致検索をおこなった。

## 4. 通時的分析：日本編

日本時間の2011年3月11日14時46分、東日本大震災が発生し、日本国内のみならず国外でも大きなニュースとなった。これは日本周辺における観測史上最大規模の地震であり、強い揺れとそれにともなう津波、そして原発事故によって発生した被害は多方面にわたり、そのいずれもが甚大であった。

本研究では、日本における《過去》と《未来》に関する言及頻度、ひいては日本人の志向の変化を捉えるために、この東日本大震災（以降、震災とよぶ）をひとつの契機として捉え、日本国内における震災以前と以後の TYR を調査した。なお、調査日は2013年11月27日であり、調査の対象とした期間は2008年7月11日から2013年11月10日までの64ヶ月間である。TYR に関しては、1ヶ月区切りで扱うものとした。

### 4.1 調査結果：日本国内にみられる TYR 推移

図1は、調査によって得られた「日本における TYR の推移」をまとめたものである。

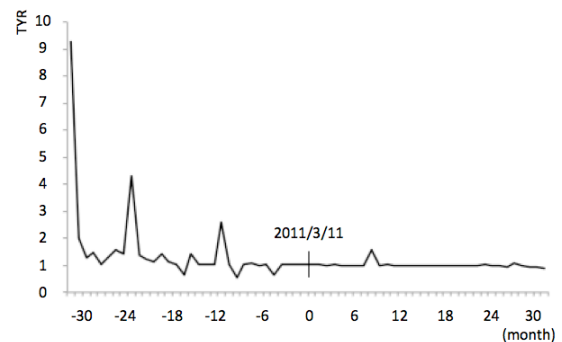


図1：日本における TYR の推移

この TYR の調査結果に関して、震災以前 (before) の TYR 平均値を  $x$ 、震災以後 (after) の TYR 平均値を  $y$  として、 $x$  と  $y$  に関する  $t$  検定をおこなった。その結果、両者の間に有意差はみられなかったが、有意な傾向がみられた ( $t(31.32)=1.93, p=0.06$ )。

### 4.2 考察：TYR 推移は何に起因するのか

調査結果に関して、震災以前と以後で極端な変化はみられなかったが、長期的にみれば TYR が下降気味になっていることがわかった。この原因が昨日に関する言及頻度が増加したことによるのか、あるいは明日に関する言及頻度が減少したことによるのか、それとも現在に意識を置くようになったことによるのかは判別しがたい。し

かし、いずれが主因であるとしても、日本人の言及、ひいては思考が「未来志向」から「過去志向」へと緩やかに変わりつつあるということは観測できる。また、地震という自然災害が突発的に発生するものであり、完全には御しがたいものであるということからも、明日はどうかかわからないと考えるようになった、もしくは、現在にも及ぶ深刻な状況を目の当たりにしながら、震災の発生以前、つまり過去について振り返るようになったという可能性が考えられる。

以上、日本国内における TYR と震災の関係性について分析と考察をおこなった。次節では、さらに視野を拡げて、世界の国々における TYR を調査し、それが何を説明し得るものであるかについて述べる。

## 5. 共時的分析：世界編

本節では、調査によって得られた世界の国々における TYR が、各国のどのような関係性を説明し得るものであるかについて述べる。

まず、日本を含む 73 ケ国を調査対象として、各国の TYR を算出した。次に、それによって得られた結果と、各国の様々な指標との相関を調査した。なお、調査対象とする国の選定については、以下の基準によっておこなった。

### 調査対象とする国の選定基準

まず、ある単一の言語が公用語として認定されている国を選出した。次に、Google の検索オプション「言語」で対象外になっている言語を公用語とする国を除外した。さらに、Google の検索オプション「地域」で対象外とされている国を除外した。その結果、日本を含む 73 ケ国が調査対象として得られた。なお、5 節における調査では 64 ヶ月間を 1 ヶ月区切りにして調査をおこなったが、本節では、調査日（2013 年 11 月 21 日）時点における《昨日》と《明日》の言及頻度を調査し、TYR を算出している点に注意されたい。

### 5.1 調査結果：73 ケ国における TYR と統計指標

調査によって得られた各国の TYR を抜粋したものを以下の表に示す。

表：世界各国の TYR (抜粋)

Region	TYR	Region	TYR
Angola	0.226	Nicaragua	0.620
Brazil	0.489	Oman	0.732
China	0.972	Panama	0.957
Egypt	0.467	Russia	0.496
Germany	1.030	Spain	1.349
Indonesia	0.486	<b>Togo</b>	<b>0.172</b>
<b>Japan</b>	<b>2.077</b>	United States	0.684
Kuwait	0.351	Vietnam	1.339
Lithuania	1.046	Yemen	0.361
Mexico	1.177	Zambia	0.204

次に、この TYR と各国における様々な指標（出生率、

福祉充実度、平均 GDP、平均 GDP 成長率、平均寿命、Happy Planet Index) との相関を調査した<sup>3)</sup>。指標によっては、数値が公開されていない国がみられ、その場合は TYR の調査対象である 73 ケ国から除外して相関を調査した。本稿では、Happy Planet Index と平均 GDP との分析結果について述べる。

### 指標 1：Happy Planet Index (HPI)

HPI (地球幸福度指数) は国の幸福度を測るための指標であり、イギリスの環境保護団体である Friends of the Earth が発表している<sup>4)</sup>。この指標は、平均寿命、国民満足度、生態系リソースの消費量などから算出される。また、HPI は経済的発展のみを捉えた指標ではなく、人間的な活動を支える文明の持続可能性を考慮に入れた国の幸福度を示すものであると紹介されている。

本研究では、2012 年に発表された HPI のリストを用いて、71 ケ国の TYR との相関を調べた。HPI と TYR に関する散布図を図 2 の左図に示す。

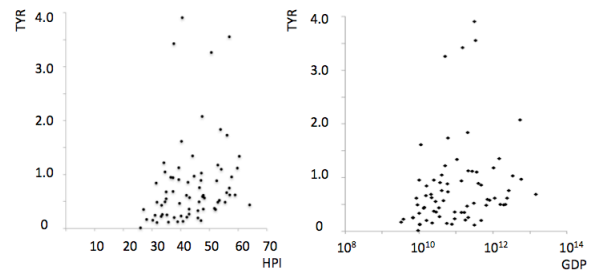


図 2：TYR と HPI (左) と TYR と GDP (右)

TYR と HPI に関する Pearson の相関係数を求めた結果、両者の間には 5%水準で有意な相関がみられた ( $r=0.25$ ,  $df=69$ ,  $p=0.04$ ).

### 指標 2：国内総生産 (GDP)

GDP (国内総生産) は、ある国の市場で取引された商品やサービスなどの付加価値の総額を示しており、各国の経済を総合的に把握するための指標のひとつである。

本研究では、2008 年から 2011 年までの 4 年間の平均 GDP を用いて、73 ケ国の TYR との相関を調べた。GDP と TYR に関する散布図を、図 2 の右図に示す。

TYR と GDP に関する Pearson の相関係数を求めた結果、両者の間に有意な相関はみられなかった ( $r=0.06$ ,  $df=71$ ,  $p=0.04$ ).

そこで、GDP ランクによって、調査した 73 ケ国を 3 群 (top24, middle24, bottom25) に分けた。たとえば、top24 には、調査をおこなった国のうち GDP ランク上位の 24 ケ国が含まれている。この各群の TYR 平均値の差に関する分散分析をおこなった結果、これらの群の間に 5%水準で主効果がみられた ( $F(2, 70)=3.72$ ,  $p=0.03$ ) ため、多重比較 (Ryan's method) をおこなった。その結果を以下の図 3 に示す。

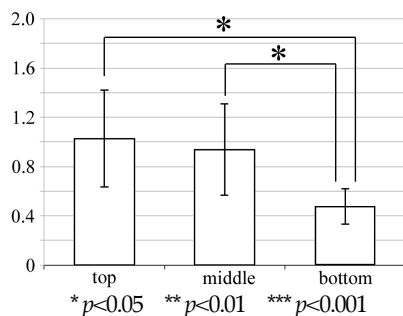


図3：top24, middle24, bottom25の多重比較の結果

その結果, top24 と bottom25 の間に有意差がみられ ( $p=0.01$ ), また, middle24 と bottom25 の間にも有意差がみられた ( $p=0.04$ ). しかし, top24 と middle24 の間には有意差がみられなかった ( $p=0.69$ ).

## 5.2 考察：HPI や GDP と TYR

前節の結果から, TYR が HPI や GDP を説明し得る可能性があることがわかった. しかし, それらの関係性がどのような要因によってもたらされているものであるかについては, 本稿では明らかにできなかった.

国に対する不満を募らせている, あるいは貧しい経済状況のなかで日々を過ごしている人々が, 明日をどう過ごすかなどについて考えることが比較的少ないということなのか, それとも, 過去を振り返ることが多いということなのだろうか.

これを調べるために, 人が「昨日」や「明日」について言及するときには, 具体的にどのような話題について述べているかについて調査をおこなった. 以下の図4は, Google N-gram Search を用いて調査した, 日本語テキストにおける「昨日」と「明日」の上位100件の内訳である.

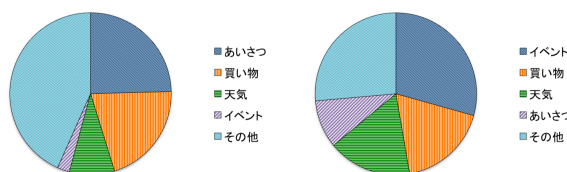


図4：「昨日」(左)と「明日」(右)の内訳

この「昨日」と「明日」の内訳については, そのカテゴリ分類が困難となる場合も多くみられ, カテゴリは多岐にわたると考えられるが, 本稿では上位4つの大きなカテゴリに「その他」を加えた, 5つのカテゴリを挙げる.

それぞれの内訳をみると, 「あいさつ」や「天気」といったカテゴリが目立つが, このようなカテゴリに含まれるテキストの量は, 国の内情や世界情勢によって大きく変化するものではないと考えられる. しかし, 「買い物」や「イベント」に関するテキストや, 「その他」のカテゴリに多く含まれている個人ブログのテキストの多寡や内容は, 国内や世界の景気傾向などによって左右さ

れる可能性が多分にある. これらのカテゴリに含まれるテキスト量の変化を観察すれば, 本稿の分析結果として得られた HPI や GDP と TYR との関係性を紐解く手がかりが得られるかもしれないが, それは今後の課題である.

## 6. おわりに

本研究では, テキスト上の世界と現実世界との関連性を捉えることを目的として, 日本国内や世界各国での《過去》と《未来》をあらわす語の言及頻度比 (TYR) を用いた分析をおこなった. 具体的表現として, 《昨日》と《明日》を選定し, 通時的分析と共時的分析の両方をおこなった.

まず, 通時的分析として, 日本国内における東日本大震災以前と以後の TYR 推移を観察した. その結果, 震災以前と以後で, 日本人の志向に緩やかな変化が生じていることがわかった.

次に, 共時的分析として, 世界各国における TYR とさまざまな世界統計指標との関係性について調査をおこなった. その結果, TYR は HPI と GDP を説明し得るものであることがわかった.

しかしながら, 今回の言及比による分析では, 上記の結果の背景にある因果関係までは明らかにできなかったため, 言及されている《昨日》と《明日》の内実がどのようなものなのかを調査し, どのような内容のテキストが増減しているかを分析することが今後の課題として挙げられる.

### 注

- 1) 本稿において《 》で示す語は, 語そのものではなく概念をあらわす. 各言語によって具体的な語は変わる.
- 2) <https://www.google.co.jp/>
- 3) 各指標の数値について, Happy Planet Index 以外の指標は <http://www.worldbank.org/> 内で公開されているデータを用いた.
- 4) <http://www.happyplanetindex.org/>

### 参考文献

- [1] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques, *Proceedings of EMNLP*, 79-86.
- [2] Preslav Nakov, Zornitsa Kozareva, Alan Ritter, Sara Rosenthal, Veselin Stoyanov, and Theresa Wilson. 2013. In Semeval-2013 Task 2: Sentiment Analysis in Twitter. *Proceedings of the 7th International Workshop on Semantic Evaluation*, 312-320.
- [3] Back, M. D., Kufner, A. C. P., and Egloff, B. 2010. The emotional timeline of September 11, 2001. *Psychological Science*, 21, 1417-1419.
- [4] Tobias Preis, Helen Susannah Moat, H. Eugene Stanley, Steven R. Bishop. 2012. Quantifying the Advantage of Looking Forward, *Scientific Reports*, 2: 350.