

トピックモデルを用いた年代別作文の時間に関する認識の分析

谷口巴¹ 上原泉¹ 川崎采香¹ 浅原正幸² 酒井裕³ 越智綾子² 小林一郎¹

¹お茶の水女子大学 ²国立国語研究所 ³玉川大学脳科学研究所

{g1620524,koba}@is.ocha.ac.jp {uehara.izumi,kawasaki.ayaka}@ocha.ac.jp

{masayu-a,a.ochi}@ninjal.ac.jp sakai@tamagawa.ac.jp

概要

児童から壮年の各年代の人々を対象に過去と将来に関する作文の収集を行い、作文中の記述内容の分析に基づき、各年代における時間認識の調査を行った。記述内容に関する分析には、潜在ディリクレ配分法 (LDA) [1] を使い、抽出された各トピックの年代別割合と単語分布を俯瞰することによって、各年代における時間の捉え方について考察を行う。

1 はじめに

新学術領域「時間生成学」¹⁾では、時間の意識（このころの時間）がどこからどのように生まれるのかについて、多角的な研究を推進している。領域においては、ヒトの脳活動データの分析を進めるほか、ヒトとの対照分析としての霊長類の行動分析も進めている。霊長類と異なり、ヒトは時間の意識を言語化することが可能であり、その言語表現を分析することで時間の捉え方を明らかにできる可能性がある。本研究では、ヒトならではの時間の捉え方、すなわち自己視点の入った時間認識を調べることを目的として、過去の個人的出来事の記憶（「過去」の出来事の記憶）と将来に自分が経験するであろう出来事の予想（「将来」の出来事の想像）の言語表現に焦点をあてることにした。「過去」や「将来」の出来事の記憶や想像のいずれにおいても、PTSD や記憶障がいに伴うと、詳細さを欠くとの報告があるが (e.g., Brown et al., 2012), 従来、その内容と時間認識に関する分析は十分なされていない。確かに、10代後半から30代初め頃の出来事を、特に高齢期は想起しやすいといった (e.g., Rubin, et al., 1986), 時間に関わる報告はある。しかし、過去と将来の経験や時間の捉え方に関して、経験時点や想起時点 (年代) による違いの有無などは、ほぼ検討されていない。これらの捉え方における、世代差・性差を明らかに

するために、幅広い年代から「過去」の出来事の記憶と「将来」の出来事の想像に関する作文を収集し分析することにした。なお、伴う感情により内容や時間的距離感にそもそも差があること、どの年代にとっても倫理的問題がなく書きやすい内容になっている必要があることを念頭に、作文をもとめるのは、過去と未来における最も楽しかった、最も楽しいと予想される自分に関する出来事のみとした。内容分析と併せて、「過去」の出来事の記憶と「将来」の出来事の想像の作文に表出する言語表現の傾向についても分析を行う。作文のデータは、小学生から80代までの幅広い世代の男女741名から収集し、人手による形態論情報付与を行った。言語表現においては、作文の形態素3-グラムに対するトピックモデルを適用し、トピック毎に出現する言語表現を検討するとともに、同トピックの言語表現を利用する世代差・性差について検討する。その結果、世代や性の違いにより、言語によって表出される出来事 (ライフイベント)・出来事の参加者が異なることを明らかにした。

2 年題別作文中の時間認識の分析

2.1 作文データ

本節では利用した成人および児童の作文データについて示す。

2.2 作文課題

調査対象者に、作文課題2問と記号問題を記した回答用紙1枚、作文を書くための罫線入り用紙2枚の計3枚を配布し回答をもとめた。作文課題はできる限り、用紙の半分以上を書くようにもとめた。記号問題は、本発表では分析対象とはしていないため説明を割愛する。本発表で分析対象としたのは、作文課題の2問への回答 (作文) であった。作文課題は (1) 過去経験した一番楽しかった出来事を書くよ

1) <https://www.chronogenesis.org/>

うもとめる課題（以下「過去作文」）、(2) 将来経験すると予想する最も楽しい出来事を書くようもとめる課題（以下「将来作文」）の2問であった。成人と児童で教示文は同じであった。

2.3 手続き

成人のデータは調査会社を通じて募集をかけ、参加を希望された方々に、依頼の手紙、作文課題と記号問題を記した回答用紙1枚と作文を記入するための回答用紙2枚（過去作文用と将来作文用）、返信用封筒をお送りし、回答用紙3枚を返信用封筒に入れて投函するようにもとめ、収集した。募集から回収までの作業を調査会社に委託した。児童のデータは、学校長の承諾のもと、学校を通じて協力いただける児童に配布し収集した。首都圏の公立小学校2校に依頼し、1校から小学4年生（70名）に参加いただき66名の分析対象データを得た。もう1校から、小学3年生から6年生（185名）に参加いただき181名の分析対象データを得た。なお本調査は、お茶の水女子大学人文社会科学研究所の倫理委員会の承認を経て実施された。

表1に調査対象者の属性の基礎統計を示す。

表1 調査対象者の属性

	小学生	20-30代	40-50代	60-80代	計
女性	127名	83名	89名	90名	389名
男性	120名	75名	79名	78名	382名
計	247名	158名	168名	168名	741名

3 実験

3.1 使用データセット

前述の作文データを将来に関する作文と過去に関する作文に分類し、解析を行う。LDAに入力する単語の単位に関しては、各年代ごとの単語使用の特性がより顕著に表れる3-グラムを用いる。これに基づき、日本語形態素解析器 MeCab を用いて形態素解析を行い、表層形を1単位として3-グラムを作成した。全文書中3回以上出現する3-グラムに限定し、文書を3-グラムからなる構成で再作成した。未来・過去作文データに関する文書数、3-グラム語彙数、1文あたりの平均単語数および3-gramの有効性を比較検証できるように文書をユニグラムで表現した際のパープレキシティを表2に示す。

表2 データ詳細

	文書数	3-グラム数	平均単語数	Perplexity
将来	722	3449	9	300
過去	738	4158	10	242

3.2 実験結果

トピック数 $K = 10$ とし、LDA を用いてトピッククラスタリングを行なった。潜在変数の推定には、ギブスサンプリング (GS) を採用し、十分収束するように、繰り返し数は1000と大きい値に設定した。また、モデルの評価指標としてパープレキシティを用いる。LDAにおけるパープレキシティは、学習されたトピック分布 $\theta_{d,k}$ と、トピック-単語分布 $\phi_{k,v}$ を用いて式1で示される。ここで、 D は文書数、 N_d は文書 d の単語数である。

$$perplexity(\mathbf{w}) = \exp \left(- \frac{\sum_{d=1}^D \sum_{n=1}^{N_d} \log(\sum_{k=1}^K \theta_{d,k} \phi_{k,w_{dn}})}{\sum_{d=1}^D N_d} \right) \quad (1)$$

パープレキシティは各文書における単語確率の幾何平均の逆数であるが、これは文書の正解単語を予測する際の単語の候補数であり、値が小さいほうがより優れたモデルといえる。

訓練時において、GSの各繰り返しごとに計算したパープレキシティをプロットすることにより、収束判定を行った（図1参照）。これにより $\theta_{d,k}$ および $\phi_{k,v}$ を取得した。

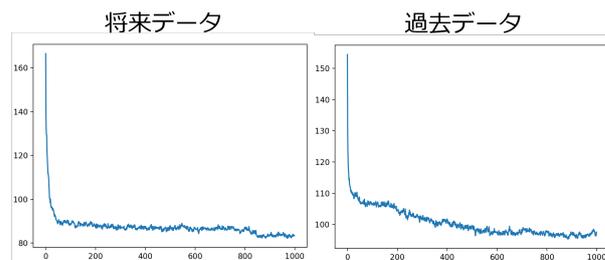
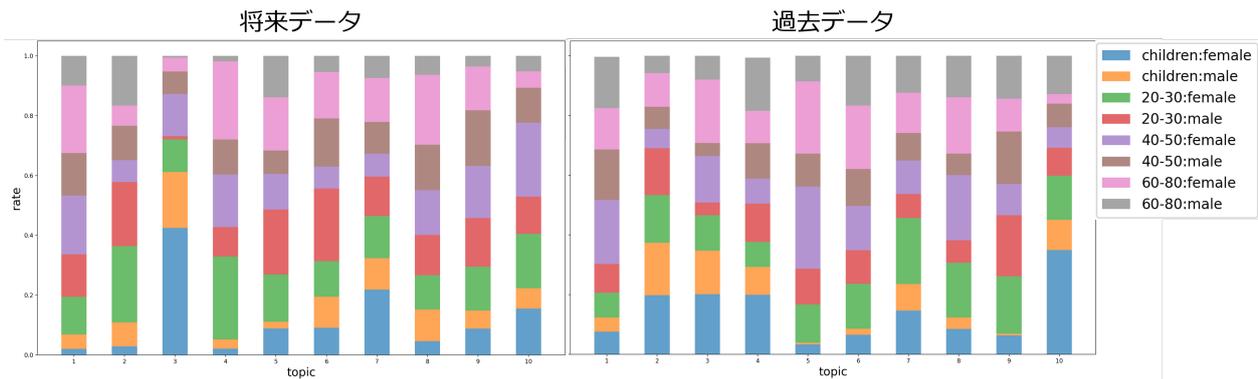


図1 訓練時のパープレキシティ。

過去と将来の作文において、LDAにより学習された $\phi_{k,v}$ から、各単語（3-グラムからなる単語）について正規化自己相互情報量 (NPMI) [2] を計算した上位10単語を表3および表4に、それぞれ示す。ここで、NPMIとは自己相互情報量 (PMI) を-1から1の間に正規化したものであり、式2で示される。

$$\begin{aligned} NPMI(k, v) &= \frac{PMI(k, v)}{-\log p(k, v)} \\ &= -1 + \frac{\log p(k)p(v)}{\log p(k, v)} \end{aligned} \quad (2)$$

図2 トピック別年代の割合.



トピックに対する単語の出現を観察するのに $\phi_{k,v}$ ではなく、NPMI を用いることで単語の出現頻度の影響を受けず、トピックを代表する単語を観察することができる。また、各トピックにおける年代別割合と確率上位2文書を、それぞれ図2、表5、表6に示す。

3.3 考察

図2より、年代別による頻出トピックが確認できる。表3、表4、表5、表6と照らし合わせながら、将来データと過去データについてそれぞれ考察を述べる。

将来データ

トピック1 40代以上の女性が多く割合を占めている。トピックの内容としては子育て後の楽しみなどを説明する単語が多く割り当てられている。

トピック2 20代から30代の働き盛りの男女や60代から80代の男性の割合が大きくなっているが、仕事における夢であったり、定年後の未来に関する文書が多く見受けられた。

トピック3 児童が半分以上の割合を占めている。トピック1と比べて遠い将来の夢や願望を表す単語が多く割り当てられていることが確認できる。児童が成人に比べて遠い将来に関する願望を書く傾向があるのは妥当な結果であるといえる。

トピック4 対照的に児童の割合が最も小さく、成人女性の割合が高くなっている。自分の子供や孫の成長を楽しみとする文書が多く見受けられた。

トピック5 成人男性がやや大きな割合を占めている。ライフイベント(結婚や家の購入)に関するものが多い。

トピック6 成人男性がやや大きな割合を占めている。趣味に関しての文書が目立つ。

トピック7 児童の割合が相対的に多く、形式名詞「こと」が多用されている。日常的な学校や友人に関するトピックである。

トピック8 40代以上の割合が相対的に多く、学業や旅行、投資などの中高年の夢や願望を含む割合が高いトピックである。

トピック9 トピック8と同じく40代以上の割合が相対的に多い。上位文書には結婚生活など家庭に関する記述が多くみられた。

トピック10 トピック1と同じような単語の分布が見受けられる。トピック1と比較して年齢層は若く、子供の成長やこれからの家族の行く末を思う文書が多くみられた。

全体を通して、児童の割合が小さいトピックには逆接や順接が含まれる単語や過去形の単語が多く割り当てられていた。成人は将来の話をする前に、理由や経験として過去の話をするなど、論理的に話す傾向がみられた。

過去データ

トピック1 児童の割合が小さく、中高年が割合を大きく占めている。家族や旅行に関する単語が見受けられ、遠い過去の思い出が多くみられる。

トピック2 家族との国内の旅行に関する記述が多く見受けられ、主に若年世代が想起する思い出から成っている。

トピック3及びトピック4 トピック2と分布が似ているが、学校生活に関しての記述が目立つトピックであった。

トピック5 児童の割合が小さく、中高年女性が割合を大きく占めている。家族や健康に関する単語が多く割り振られている。将来データと同様、男性に比べて女性の作文には他人が主役であったり、自分に関するイベントであっても家族や友人に関する

表3 未来データ：トピック別3-グラム上位10語.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
予定して	として	したいです	いるので	になり、	ていまし	もらいたい。	たが、	になると	イラストを
いる姿を	は、定年	はいりたい	子供達が	には、	ができれば	はとても楽	をすること	と思われ	かく
もあります	経験すると	です。旅行	したいと	中学時代か	ことができ	なケーキを	。ぼくは	おばあちゃ	の時の
ていた	都道府	、お父	ができるよ	、休日に	ママとパパ	てもらいた	そうです。	なること。	いるかも
0歳を	の管理職	、これも	に行く事	ないが、	していき	の誕生日	はあります	20年	になった
、これは	に帰った	、その土地	出来ました	と共に	ほいくえん	ます。自分	が出て	。子供が	思うと、
、そのため	、これから	、それも	が3人	、しょうら	、食べるこ	お店に	になりました	、おいしいも	、海外旅行
、それは	のは、	、どこか	です。昨年	、その日	しての	、いつも	に住んで	、そして孫	を想像する
、とても楽し	を感じられ	、わかりませ	ます。孫	、それを	年生に	、この先	ていませ	、それで	」という
み、ゆっくりと	なると	、一度	楽しみたい	、上の	、その時	、そこで	を食べ、	、という	、これが

表4 過去データ：トピック別3-グラム上位10語.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
、20	、みんな	をかって	に、お	であり、	いたの	ができた	遊んでい	旅行をし	お母さん
時が一番	のは、	行いました	フリースタ	中学校に	200	のような	ので、	と行きまし	た。わたし
を購入し	ジェット	の頃と	24日	に赴任し	部の人	なったこと	動物園に	、子供の	いとこたち
た	た出来事は	気の合う	夢のよう	過ごしまし	だったが	た。妹	た時」	で行ったり	。お母
ながらも、	たと思う	に住んで	観に行っ	も、とても	自分のため	、家族と	体験した	来た。	！」と
が私を	三年生	ことが出来	。でも	に声を	もらえた。	。そして、	でしょうか。	16年	それでも
海外での	今年の夏休	、家族4	ており、	あると思	したの	2月1	出来事では	がんばって	と思った
の歌を	た時は	、お互い	、これが	いるのだ	幸せだった	の海外で	、お互い	、これが	事に感謝
、おかし	10日間	、この時	、その前	、いままで	さんとの	6月に	、この時	、その前	。子供達
、お金	に行ったり	、そのとき	、それまで	、こんなに	のだと	の会社に	、そのとき	、それまで	4年頃

記述が多く見受けられる。

トピック6 他のトピックと比べて男性の割合が多いトピックである。トピック確率が高い文書では旅行関連の記述が多い。また、当時の西暦や年齢など、数字を用いての表現も多くみられた。

トピック7 若年世代の女性と中高年による友人に関する内容の占める割合が比較的高く、文章では連用中止の表現が多い。

トピック8 女性の割合が大きい。他者より自分中心のレクリエーション関連の出来事が多く見られる。

トピック9 主に海外旅行に関するトピックであり、一人旅など家族や友人に関する要素が薄く、20代から30代の男女の割合が大きくなっている。

トピック10 学校生活での思い出に関してのみで構成されており、女子児童が割合の大半を占めている。

過去データでは旅行関係のトピックが多く見られ、家族や学校の要素が加わることでトピックが細かく分かれていることが確認できた。

4 まとめと今後の課題

年代別に作成された作文を収集し、LDAを用いてトピッククラスタリングを行った。時間認識に関す

る顕著な特徴としては、過去といったときに中高年は遠い過去（学生時代など）に思いを馳せ、将来といったときに児童は遠い将来（主に若年成人期）を思い描く傾向がみられ、推測の域を出ないが、世代を問わず共通して、青年期から若年成人期を人生の中心とする主観的な時間軸を有している可能性が考えられる。とはいえ、詳細までは把握しきれなかったが、内容、時間認識とも世代差や性差が本データにおいて示唆されており、今後はデータ数を増やし、Dynamic Topic Modelなどを用いて年代の移り変わりによるトピックの変化を分析することなどが課題として挙げられる。

謝辞

本研究は、科学研究費補助金・新学術領域研究「時間生成学」（18H05521,18H5524）の支援を受けて行った。

トピック解析において統計数理研究所持橋大地氏から有益なアドバイスを頂きました。ここに深謝いたします。

参考文献

- [1] David M. Blei, Andrew Y. Ng, Michael I. Jordan, and John Lafferty. Latent Dirichlet Allocation. **Journal of Machine Learning Research**, Vol. 3, pp. 994–1022, 2003.
- [2] Gerlof Bouma. Normalized (Pointwise) Mutual Information in Collocation Extraction. **Proceedings of GSCL**, pp. 31–40, 2009.
- [3] 浅原正幸, 川崎采香, 上原泉, 酒井裕, 須藤百香, 小林一郎, 越智綾子. 「過去」「未来」を主題にした作文の文体分析. 日本語学会 2021 年度秋季大会, No. A-7, pp. 37–42, 2021.

A 付録

表5 将来データ：トピック別上位2文書

Topic	文書	年代
1	中学校の生活。	children:male
	5年後くらいに、子供を独立させて、会社を辞めて、温かい南の島の別荘を買い、嫁さんと...	60-80:male
2	2018年に結婚したため、妻や将来生まれてくるであろう子供と、休日に公園やテーマパークで...	40-50:male
	これから経験すると思う楽しい事は、車中泊にて、日本一周することです。まだまだ日本には、見た...	60-80:male
3	わたしは、大人になったらパティシエのおしごとが楽しみです。みおさんとゆなさんといっしょに...	children:female
	わたしのしょうらいのゆめは、びようしになることです。わたしは大学生になったらかなうといいな...	children:female
4	私の夢は歌を唄う人になる事なので、そういった未来をよく想像します。近い将来、バンドを...	20-30:female
	仕事ができるようになったら楽しいのかも知れない。大病したあとはなにも手付かずだったので...	60-80:female
5	時期はわかりませんが、国内旅行がしたいです。時間とお金の許す範囲で限られてしまうと...	40-50:female
	マイホームでの日常が最も楽しいのではと感じられます。結婚し、子どもが生まれ、学生時代...	40-50:male
6	今、行っている夫との年末やお盆の時をさけた毎月のドライブ旅行を、今後も続けて...	40-50:female
	最も楽しいと思う事は、マイホームを持ち、その家を、自分の好きな様にカスタマイズ...	40-50:male
7	しょうらい公園をさんぼしている時いろいろな人に「おはようございます。」「こんにちは。」...	children:female
	わたしはひっこしてきたから、夏休みに友達とあうことがたのしみです。	children:female
8	5年ほど前から運用している海外株式投資がうまく利益が乗り、自分が定年65歳を迎える...	40-50:male
	私の子供(娘)ですが、昨年、成人式を迎え自分的にもうれしく思えました。いずれいい男性...	60-80:female
9	夫婦二人で旅行に行っています。いつまで健康でいられるかわかりませんが優しくなった夫...	40-50:female
	本当にただ予想しているだけで、確定ではないのですが、もしかしたら結婚して子どもが...	20-30:female
10	10年後位、私達にも孫が出来て、子供達も独立して、夫婦二人になった時、一年の半分位...	20-30:male
	自分の両親、姉妹の家族、自分の家族でワイワイと旅行に行き、のんびり散歩をしたり...	20-30:male

表6 過去データ：トピック別上位2文書

Topic	文書	年代
1	2012年から2017年にかけて、仕事の都合で海外駐在をしていました。赴任地において...	40-50:male
	今までで一番楽しかった出来事は初めての子供に恵まれた時だ。32年前、彼女は私達...	60-80:male
2	沖縄石垣島で1994年から過ごした3年間の間、1997年2月からの1年間です。働いて...	60-80:female
	家族と外出したこと。今まで、飛行機で旅行に行った事や、休日はショッピングモールへ行った...	children:female
3	30才前後の3年6ヶ月。主人の仕事でオーストラリアのバースに住みました。当時3才...	40-50:female
	家族旅行。47才の時、娘が高校受験を終えての春休みにヨーロッパ旅行に出かけた事。海外旅行...	40-50:female
4	高校3年生(1962年)の3月に兵庫県芦屋市の実家で東京大学理科I類現役合格を...	60-80:male
	15年ぐらい前の夏、千葉の海、ボディーボードスクールの合宿に参加した時のこと。友人夫婦...	60-80:female
5	今までで一番楽しかった出来事は、高校生になってはじめて迎えた正月旅行で、母と...	20-30:male
	今から10年前、単身赴任をしていた主人が命に関わる大病で手術をしました。幸い...	40-60:female
6	今までで一番楽しかった出来事は、昨年ホーチミンへ学会のため訪問したことである。研究室の後輩...	20-30:female
	今が一番良いと信じて生活しています。過去の事をいろいろ振り返っても終わった事は終わった事...	60-80:male
7	1990年頃は、いまの妻と知りあい、特に不安など考えた事ありませんでした。毎日、今の妻...	60-80:female
	今までで一番楽しかった出来事は、数え切れないけれど、それらの集積が「今」だと考えます...	40-50:female
8	大学時代は、はじめての一人暮らしができるということで自由な時間をめいっぱい楽しみました。親...	20-30:female
	私が今までで一番楽しかったのは、一番好きなアーティストのライブに初めて行った時のことです...	20-30:female
9	高校3年生の時、秋の体育祭でダンスを踊る種目があり、その時男女ダンスのペアになった子が好きに...	40-50:male
	平成19年11月の新婚旅行。妻とイタリアに1週間の新婚旅行に行きました。最初にミラノでダヴィ...	60-80:female
10	わたしは、3年生になってえんそくにいったことがたのしかったです。こうこうはくぶつかんと...	children:female
	それは、1963年8月末の事でした。残暑きびしい中、我市立旭中学校野球部は県大会の決勝戦へ駒...	60-80:male