

棄却文章断片の創造的文章作成時における活用可能性の検証

生田泰章^{†1} 高島健太郎^{†1} 西本一志^{†1}

概要: 筆者らはこれまで、文章作成においていったんは生成されたものの、最終稿には採録されずに棄却される文章断片の活用可能性に着目し、この棄却文章断片を収集するためのシステムである Text ComposTer を開発した。Text ComposTer は2つの異なる粒度の棄却文章断片を収集可能な文章作成支援システムであり、これまでの実験から粗粒度の棄却文章断片 (R-DTF) が特に高い活用可能性を有するという示唆を得ている。そこで、本稿においては、執筆者のアイデアが含まれる創造的文章を活用対象として、この創造的文章が作成される過程で R-DTF が実際に活用されるかどうかの検証を行った。検証実験の結果、R-DTF が創造的文章作成において有効に活用されることが確認できた。さらに本稿では、活用された R-DTF について詳しく分析を行う。

Are Unused Text Fragments Useful in Creating Another Creative Document?

HIROAKI IKUTA^{†1} KENTARO TAKASHIMA^{†1} KAZUSHI NISHIMOTO^{†1}

Abstract: We have focused on utilizing unused text fragments named DTFs that is generated once in writing but not in the final draft. Specifically, we developed a document composition support system Text ComposTer that can collect two kinds of DTFs with different granularity. We found that R-DTFs (Rough grain DTFs) has high possibility to (re)utilize through user studies. In this study, we investigate reusability of R-DTFs by conducting user studies of creating creative document which is contained author's ideas. As a result, it was shown that R-DTFs were effectively utilized in creating a creative document. Furthermore, we analyze utilized R-DTFs in detail.

1. はじめに

情報技術の発展・普及に伴い知識創造社会を迎えつつある現在において、新たな知識の創造や、創造された知識の活用を促進することは非常に重要な課題である[1]。また、文章は、知識を形式知として活用するための重要なメディアであり、文章化という行為は形式知の創造活動と言える。文章化は、様々な要素が複雑にからみあった行為であり、これまで執筆者による文章作成行為を表すモデルが複数提案されている[2][3][4]。これらのモデルには共通して推敲のプロセスが存在する。推敲プロセスの存在は、執筆者が文書作成作業の開始時点からいきなり最終版の文章を完成させることができるのではなく、試行錯誤を繰り返しながら徐々に文章を完成させていくことを表している。それゆえ、文章作成が行われる過程では、草稿には記載されているものの、最終稿においては棄却されてしまう本文の一部である棄却文章断片 (DTF: Deleted Text Fragment) が生じる。DTFは、最終稿 (ある形式知) の創造には不用と判断された知識断片であり、最終的に記録も記憶もされないままに消失してしまうことがほとんどである。

一方、新たな知識やアイデアは、それ自体が急にひらめくものではなく、過去の複数のアイデアの新たな組み合わせにより創造されるという主張が存在する[5][6]。この主張に従えば、企画書や研究計画書、科学技術論文等の執筆者のアイデアが含まれる文章 (本稿では創造的文章と呼ぶ)

を作成する際に、そのアイデアの着想の手掛かりとして過去のアイデアが活用可能であると考えられる。これは DTF についても同様であり、ある文章において不用と判断された知識断片である DTF は新たな創造的文章を作成する際の着想の手掛かりとして活用可能であると考えられる。

このような背景から、筆者らはこれまで、DTF を新たな創造的文章作成に活用すべく研究を行ってきた[7][8][9]。これらの研究の中で、DTF を粒度別に分別して収集することが可能な Text ComposTer を開発し、収集された棄却文章断片のうち、意味的内容を多く含む粗粒度の DTF (R-DTF: Rough-grain DTF) の活用可能性が高いという示唆を得ている。しかしながら、R-DTF が新たな文章作成に活用されるかどうかの検証は未実施であった。そこで、本稿では、新たな創造的文章作成において R-DTF が実際に活用されるかどうかの検証実験を行い、その実験結果を踏まえ、DTF が有効に活用可能であるかどうかの検討を行う。

2. 関連研究

従来、生成されたコンテンツを効率的に再利用するための Contents Reuse に関する研究が活発に行われてきた。Rockley は、企業などの組織で生成されたコンテンツをより効率良く活用するために XML 等を用いた統一的なコンテンツ管理の方法論を提案している[10]。Verbert らは、生成したコンテンツを効率よく再利用するために ALOCOM フレームワークを提案している[11]。ALOCOM フレームワークは、コンテンツを再利用しやすいように、図表や文章

^{†1} 北陸先端科学技術大学院大学
Japan Advanced Institute of Science and Technology

などのあらかじめ決められた構成要素に分解し、新たなコンテンツを生成するときに検索可能なように、分解した構成要素をデータベースに蓄積している。また、Contents Reuse が実際にどのように行われているかを調査する研究も行われている。Mejova らは、Contents Reuse が企業内のどの部署においてどの程度行われているかの実態を調査している[12]。また、Jensen らは、17 人のナレッジワーカーを対象に、個人における Contents Reuse の実態を調査している[13]。

これらの Contents Reuse に関する研究は、生み出されたコンテンツの一部または全部を直接的に再利用することで、コンテンツ生成の効率化を図るためのものである。対して、我々は、新たなコンテンツを生み出す機会の創出を目的として、この機会創出のトリガとなる可能性を秘めているにも関わらず、今まで見過ごされていた棄却文章断片を収集し、活用することを目指している。本稿においては、棄却文章断片が新たな創造的文章作成時にどのように活用されるかを観察することで、その活用方法を見出すことを目指している。

3. Text ComposTer

3.1 システム概要

図 1 は、本稿の検証実験で使用する Text ComposTer の操作画面である。Text ComposTer は、文章断片が記入されたカード状の要素の配置位置にしたがって本文全体を形成することにより、執筆者の文章作成を支援するシステムである。Text ComposTer は、本文全体を表示する表示領域と、要素の配置が可能な配置領域を備える。また、配置領域は、反映領域と非反映領域を有しており、反映領域に配置された要素内の文章断片は表示領域に反映され、非反映領域に配置された要素内の文章断片は表示領域に反映されない。ここで、表示領域に反映される文章断片の順序は、Nakakoji らの Art#001[14]と同様に、反映領域の上下方向における要素の配置位置に対応している。すなわち、反映領域の最上部に配置された要素内に記入された文章断片が表示領域の最先に表示され、反映領域の最下部に配置された要素内に記入された文章断片が表示領域の最後に表示される。ユーザは、要素の生成、要素内への文章断片の記入、反映領域または非反映領域への要素の配置を行うことによって、本文全体を作成する。

Text ComposTer は、以下の各機能を有している。



図 1. Text ComposTer の操作画面
 Fig. 1. Snapshot of Text ComposTer.

- (1) Generate 機能：要素の生成
- (2) Split 機能：要素の分割
- (3) Save 機能：操作画面内の作業環境の保存
- (4) Load 機能：Save 機能で保存された作業環境の再現
- (5) Done 機能：表示領域内のテキストを外部ファイルとして出力（作業完了時等に実行される機能）

図 1 の要素は、メモ欄と本文欄の 2 つのテキスト入力可能な領域を有している。本文欄に記入されたテキストの内容は、要素の配置位置によって表示領域への反映・非反映を変化させる。メモ欄への記入内容は、要素が反映領域、非反映領域のいずれに配置されていたとしても、表示領域に反映されない。メモ欄と本文欄を 1 つの要素に設けることによって、ユーザがメモ欄に文章作成のためのアイデア・キーワードを記入し、そのアイデア・キーワードを基にした本文を本文欄に記入することを可能とする。

3.2 R-DTF とその特徴

Text ComposTer は、粗粒度と細粒度の 2 つの粒度別に棄却文章断片を収集する機能を有している。粗粒度の棄却文章断片 (R-DTF) は、Done 機能が実行されたときに非反映領域内に配置された要素それぞれに記入されている文章断片である。細粒度の棄却文章断片 (F-DTF: Fine-grain Deleted Text Fragment) は、各要素内の文章断片の編集に削除された文字列のことである。

前述のような GUI および機能を備える Text ComposTer は、一般的なテキストエディタのように単に文章作成過程の最終状態を表示するだけでなく、文章作成過程全体を支援する機能を持つ。Text ComposTer を用いることで、文章

作成の上流工程（文章の構想・構成段階）で創造されたものの、最終的に本文に採用されなかった棄却文章断片を R-DTF として収集し、文章そのものに対する誤字の訂正や表現の修正等の編集操作によって生じる棄却文章断片を F-DTF として分別収集することができる。

これまでの被験者実験から、R-DTF は、文章構成においていったん書かれたものの最終的に削除された文章塊であることが多く、後に活用される可能性が F-DTF に比べて大幅に高いという結果を得ている[8]。一方、Text ComposTer の使用方法によっては、R-DTF の一部に、本文に採用された文章断片の内容とほとんど同じ内容が含まれる場合があるという示唆も得ている[8]。

このように R-DTF の特徴を知るための被験者実験を行ってきたが、新たな創造的文章の作成時に、実際に R-DTF が活用されるかどうかの検証は未実施であった。そこで、本稿では 4 章および 5 章で示す実験を行うことで R-DTF の活用可能性を検証する。具体的には、R-DTF を収集するための事前実験（4 章参照）を経た後、収集された R-DTF が創造的文章作成時に活用されるかどうかを検証する活用実験（5 章参照）を行う。

4. 事前実験

4.1 実験設定

4 名の被験者それぞれに図 1 の操作画面を有する Text ComposTer を用いて文章作成を行ってもらい、R-DTF を収集する実験を行った。被験者はすべて、筆者らの所属する大学院の修士課程の学生であり、以下の 4 つのテーマについて文章作成を行ってもらった。本稿においては、作成される文章が創造的文章となるように各テーマを設定した。

- テーマ 1：10 年後の公園がどのようになっているか。
- テーマ 2：50 年後の公園がどのようになっているか。
- テーマ 3：未来のレストランがどのようになっているか。
- テーマ 4：未来のファッションがどのようになっているか。

また、各被験者には 1 日に 1 テーマの文章作成を行ってもらった。このとき、カウンターバランスを考慮して各被験者にテーマを割り振った。文章作成時間は 30 分程度に設定し、各文章の作成文字数を 100 字以上 400 字以内となるように被験者に教示した。

以上の設定のもとで被験者それぞれが各テーマに関する文章を執筆した後、本文に関する情報および R-DTF に関する情報を収集した。具体的には、反映領域に配置されたエレメントの情報を本文に関する情報として収集し、非反映領域に配置されたエレメントの情報を R-DTF に関する情報として収集した。ここで、上述したように、収集した R-DTF の中には、本文に含まれる文章断片とほぼ同内容のものが含まれている場合がある。そこで、本稿では、本文 d に対する R-DTF の文字列 s の含有率 $R_{s,d}$ を下記式によ

表 1. 本文を構成するエレメント数。

Table 1. Number of elements of the body of documents.

被験者	テーマ1	テーマ2	テーマ3	テーマ4	計
S1	6	6	4	5	21
S2	3	3	8	5	19
S3	4	5	5	7	21
S4	8	6	4	5	23
計	21	20	21	22	84

表 2. 収集された R-DTF の数。

Table 2. Number of collected R-DTFs.

被験者	テーマ1 (真/偽)	テーマ2 (真/偽)	テーマ3 (真/偽)	テーマ4 (真/偽)	計
S1	2(1/1)	3(2/1)	1(0/1)	0(0/0)	6(3/3)
S2	4(4/0)	4(4/0)	7(7/0)	6(4/2)	21(19/2)
S3	8(5/3)	5(3/2)	4(1/3)	11(9/2)	28(18/10)
S4	11(5/6)	5(5/0)	11(10/1)	9(5/4)	36(25/11)
計	25(15/10)	17(14/3)	23(18/5)	26(18/8)	91(65/26)

て算出し、 $R_{s,d}$ が閾値未満の R-DTF を真の R-DTF、 $R_{s,d}$ が閾値以上の割合の R-DTF を偽の R-DTF として区別する。なお、R-DTF の文字列 s は、本文 d の執筆中に生成されたものとする。

$$R_{s,d} = \frac{|LCS(s,d)|}{|s|}$$

ただし、 $|LCS(s,d)|$ は、R-DTF の文字列 s と本文 d の最長共通部分列（LCS: Longest Common Subsequence）の長さであり、次式で定義される[15]。

$$|LCS(s,d)| = \max_{\forall i, \forall j: s[i]=d[j]} |\vec{i}|$$

ただし、 \vec{i} は、R-DTF の文字列長 $|s|$ 以下の長さのインデックスベクトルであり、 $s[\vec{i}]$ は、R-DTF の部分列を表す。また、 \vec{j} は、本文の文字列長 $|d|$ 以下の長さのインデックスベクトルであり、 $d[\vec{j}]$ は、本文 d の部分列を表す。なお、部分列は、記号列に対して順序を保持した部分的な記号列を表し、隣接性は問わない[15]。また、本実験では、閾値を 0.7 として R-DTF の真偽を区別した。

4.2 実験結果

表 1 は、反映領域に配置されたエレメントの数を被験者およびテーマ別にまとめた表であり、表 2 は、非反映領域に配置されたエレメントの数、つまり R-DTF の数を被験者およびテーマ別にまとめた表である。また、表 2 には、真の R-DTF および偽の R-DTF の内訳も含まれている。

表 1 に示されたように、本文を作成するに当たって生成されたエレメントの数は、被験者間においても、テーマ間においても大きな差はなく、統計的な傾向も有していなかった。一方、表 2 に示されたように、計 91 個の R-DTF が収集され、そのうち真の R-DTF は 65 個であった。また、

テーマ間で R-DTF の数に統計的な傾向はなかったが、被験者間においては、Tukey 法による多重検定を実施したところ、被験者 S1 および S3 と、被験者 S1 および S4 との間で収集された R-DTF の数に有意差があった。

5. 活用実験

4 章で収集した本文に関する情報および R-DTF に関する情報を用いて、R-DTF が創造的文章作成時に実際に活用されるかどうかの検証を行う。ここで、本稿では、R-DTF を活用する主体は、事前実験の被験者とは異なる被験者を対象とする。事前実験の被験者は、当該実験にて不用と判断した文章断片を R-DTF として棄却している。事前実験とは異なる新たな創造的文章を作成する場合であったとしても、その被験者が過去に R-DTF を不用なものであるとして棄却した記憶が残っている状態では、その R-DTF が活用される可能性が低いことが考えられる。そのため、事前実験の被験者を活用主体とする場合、事前実験の実施後から十分な期間が経過後、活用実験を実施する必要があると思われる。そのため、本稿では、まず R-DTF を生成した本人以外が活用主体となるかどうかを検証する実験を行う。

5.1 実験設定

活用実験においては、事前実験の被験者とは異なる 12 名の被験者に収集実験のテーマ 2 について文章作成を行ってもらった。被験者は、筆者らの所属する大学院の修士課程の学生であり、収集実験で採用した被験者とは異なる。また、被験者を 4 人一組の 3 グループに分け、各グループで以下の異なる条件にて文章作成を行ってもらった。

- 条件 1：何も見ない状態で文章作成を行う。
- 条件 2：同一テーマの完成形の記事を参照可能な状態で文章作成を行う。
- 条件 3：R-DTF が参照可能な状態で文章作成を行う。

条件 1 および条件 2 は、新たな創造的文章の作成時における R-DTF の活用態様を観察するに当たっての比較対象としての条件であり、条件 3 が本稿における主たる観察対象としての条件である。なお、条件 2 と条件



図 2. 条件 1 の被験者が使用する操作画面

Fig. 2. Snapshot of the system for the experiments of the condition 1.

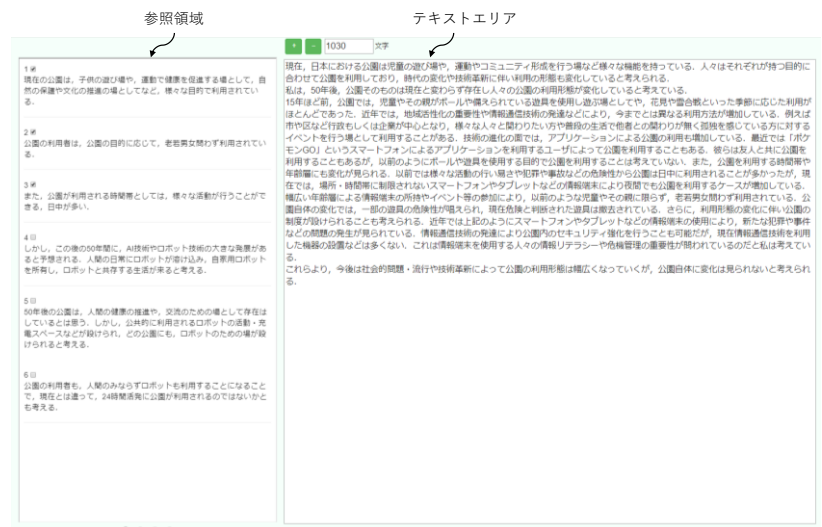


図 3. 条件 2 の被験者が使用する操作画面

Fig. 3. Snapshot of the system for the experiments of the condition 2.



図 4. 条件 3 の被験者が使用する操作画面

Fig. 4. Snapshot of the system for the experiments of the condition 3.

3 の被験者に対しては、それぞれ参照可能な外部知識（条件 2 については本文の文章断片を指し、条件 3 については R-DTF を指す）がどのようにして生成されたものかは教示

表 3. 完成した文章の文字数.

Table 3. Character length of obtained documents

条件1				条件2				条件3			
被験者1	被験者2	被験者3	被験者4	被験者5	被験者6	被験者7	被験者8	被験者9	被験者10	被験者11	被験者12
1437	1462	1171	1702	1612	1097	1003	1033	1427	1050	1421	1366

していない。

各条件では異なる操作画面を有するシステムを用いて文章作成を行ってもらった。条件1のグループ(被験者1-4)には図2の操作画面内のテキストエリアに文章を記入してもらった。条件2のグループ(被験者5-8)には、図3の操作画面を用いて文章作成を行ってもらった。図3の操作画面は、文章を記入するためのテキストエリアと、参照領域を有する。参照領域には、事前実験にて4名の被験者が作成したテーマ2の本文それぞれを参照可能に構成されている。また、各本文はText ComposTerのエレメントごとに区切られ、各区分にはチェックボックスが付与されている。つまり、表1のテーマ2の列に記載された対応するエレメントの数だけ、各本文が区切られている。条件2の被験者には、文章作成の際に参照領域の本文を参考にしてもよい旨教示し、本文の一部を参考にした場合、その部分が含まれる区分のチェックボックスにチェックを入れるように教示した。条件3のグループ(被験者9-12)には、図4の操作画面を用いて文章作成を行ってもらった。操作画面には、R-DTFリストとテキストエリアが設けられている。R-DTFリストには、事前実験で収集された91個のR-DTFが順に記載されており、各R-DTFには識別番号とチェックボックスが付されている。被験者は、R-DTFリスト内のR-DTFを参照しながらテキストエリアに文章を記入することができる。条件3の被験者には、文章作成時にR-DTFを参考にしてもよい旨教示し、さらにR-DTFを参考にした場合にはそのR-DTFに付されたチェックボックスにチェックを入れるように教示した。

本実験においては、文章作成時間を無制限に設定し、最低1000文字以上の文章を作成することを被験者に要求した。ただし、各被験者には、最大3時間分の謝金が作成時間に応じて支給される旨伝えている。このような実験設定において、各被験者が文章作成を行う様子を実験用システムの画面を録画することで観察した。また、文章作成後の各被験者にインタビューを実施した。インタビューにおいて、どのように操作画面を使って文章作成を行ったかということを中心に質問した。また、条件2の被験者にはどのような場面で参照領域内の本文を参照したかということを中心に質問し、条件3の被験者にはどのような場面でR-DTFを参考にしたかということを中心に質問した。

5.2 実験結果

以上の実験設定において行った実験結果を、完成した文章、文章作成プロセス、および文章作成時に参照された外

表 4. 文章間類似度の対応表

Table 4. Corresponding table of similarities of documents.

	被験者1	被験者2	被験者3	被験者4	被験者5	被験者6	被験者7	被験者8	被験者9	被験者10	被験者11	被験者12
被験者1	1	0.405	0.415	0.527	0.402	0.421	0.249	0.354	0.43	0.443	0.439	0.397
被験者2	0.405	1	0.33	0.39	0.327	0.414	0.375	0.467	0.458	0.307	0.385	0.403
被験者3	0.415	0.33	1	0.562	0.506	0.5	0.234	0.441	0.459	0.518	0.593	0.502
被験者4	0.527	0.39	0.562	1	0.485	0.483	0.23	0.352	0.48	0.526	0.522	0.444
被験者5	0.402	0.327	0.506	0.485	1	0.465	0.258	0.452	0.497	0.525	0.492	0.491
被験者6	0.421	0.414	0.5	0.483	0.465	1	0.247	0.371	0.415	0.454	0.423	0.39
被験者7	0.249	0.375	0.234	0.23	0.258	0.247	1	0.424	0.301	0.213	0.274	0.286
被験者8	0.354	0.467	0.441	0.352	0.452	0.371	0.424	1	0.646	0.31	0.575	0.411
被験者9	0.43	0.458	0.459	0.48	0.497	0.415	0.301	0.646	1	0.319	0.649	0.397
被験者10	0.443	0.307	0.518	0.526	0.525	0.454	0.213	0.31	0.319	1	0.408	0.415
被験者11	0.439	0.385	0.593	0.522	0.492	0.423	0.274	0.575	0.649	0.408	1	0.407
被験者12	0.397	0.403	0.502	0.444	0.491	0.39	0.286	0.411	0.397	0.415	0.407	1

部知識のそれぞれについて述べる。

5.2.1 完成した文章について

表3は、完成した文章の文字数を被験者別にまとめた表である。表3に示すように、被験者間および条件間で完成した文章の文字数に大きな差はなく、統計的な傾向も見られなかった。

表4は、完成した文章の類似度をまとめた表である。表4に記載の類似度は、対応する2つの文章のコサイン類似度を計算したものである。具体的には、対応する2つの文章それぞれについて名詞、動詞、形容詞、未知語をMecab[16]を用いて抽出し、各文章に対応するBag-of-wordsを生成する。そして、この2つのBag-of-wordsのコサイン類似度を計算することで、2つの文章間の類似度を算出する。なお、本稿では各文章で頻出する「公園」の文字列をストップワードとしてMecabにて対応する文字列を抽出した。表4に示すように、各文章間の類似度は0.213から0.649の値まで広く分布していた。

5.2.2 文章作成プロセスについて

次に、活用実験における被験者の文章作成プロセスについて説明する。実験の結果、被験者2および被験者3を除くすべての被験者が、本文を作成する前に構成を思案する様子が観察された。具体的には、被験者2および被験者3は、実験開始直後に一連の文章の記述を始めたのに対し、その他の被験者は、テーマに関連する単語や短文の羅列や、構成案について発話しているなどの様子が観察された。表5には、各被験者が文章作成に要した時間と、事前の構成に費やした時間が示されている。条件1-3の構成時間についてTukey法により多重検定を行ったところ、条件1および3と条件2および3において有意差があった。つまり、

表 5. 文章作成全体に要した時間および本文作成に当たって事前の構成に費やした時間

Table 5

	条件1				条件2				条件3			
	被験者1	被験者2	被験者3	被験者4	被験者5	被験者6	被験者7	被験者8	被験者9	被験者10	被験者11	被験者12
作成時間(分)	85	113	54	120	88	66	145	125	186	143	126	145
構成時間(分)	7.6	0	0	4.2	18.4	6	24	4.8	27.5	29.3	20.6	37.5

条件 3 の被験者は他の条件における被験者に比べて構成に費やす時間が有意に長かったことを示している。次いで、インタビューの結果と動画分析の結果を踏まえ、各条件別に被験者の文章作成プロセスの概要、特に本文の作成を開始するまでのプロセスを中心に説明する。

(1) 条件 1

被験者 1 は、実験開始後、文章構成のためのメモ（構成メモと呼ぶ）の作成に 7.6 分間を費やした。その後、このメモを残した状態で、本文の作成を開始していった。被験者 2 は、実験開始後、いきなり本文の作成を開始した。そして、ある程度本文を書き上げたところで、その本文の一部を分解し、加筆修正を加えていながら文章全体を構成していった。被験者 3 も被験者 2 同様、実験開始後にいきなり本文の作成を開始した。そして、ある程度本文を書き上げたところで、その本文から構成を抜き出して、構成メモを書き出した。その後、本文と構成メモを行き来しながら、本文の形成と構成メモの内容の拡充を行っていった。被験者 4 は、実験開始後、構成メモの作成を開始し、4.2 分後に文章作成を開始した。その後の文章作成プロセスは被験者 1 と同様である。

(2) 条件 2

被験者 5 は、実験開始後、まず構成メモの作成に取り掛かった。そして、構成メモの作成にあたり、適宜参照領域に記載の文章を参照しながら内容の拡充を行っていった。そして、実験開始から 18.4 分後に本文の作成を開始した。被験者 6 は、まず初めに参照領域に記載の文章全体に目を通した上で、それらの文章を参照しながら発話にて文章構成の思案を行い始めた。そして、実験開始から 6 分後に本文の作成を開始した。被験者 7 は、被験者 6 と同様に初めに参照領域に記載の文章全体に目を通した。その後、構成メモの作成に取り掛かった。このとき、被験者 7 はテーマ 2 に対する自身の意見を中心に構成メモの作成を行い、参照領域中の文章は参考程度にしか参照しなかった。そして実験開始から 24 分後に本文作成に取り掛かった。被験者 8 は、実験開始後、参照領域中の文章を参照しながら、文章構成の思案を行い始めた。このとき、被験者 8 は構成メモを取らずに頭の中で構成を形成していた。そして、実験開始から 6 分後に本文の作成を開始した。

(3) 条件 3

被験者 9 および被験者 10 は、実験開始後、まず構成メモの作成に取り掛かった。そして、構成メモの作成にあたり、

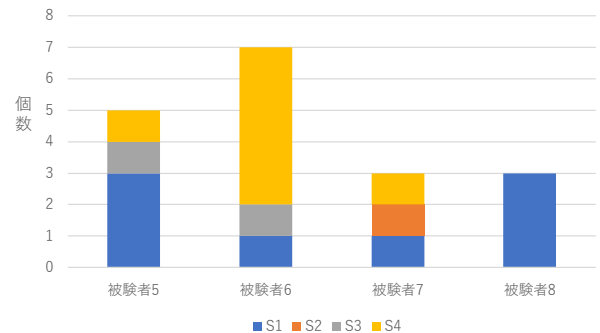


図 5. 文章断片の活用個数

Fig. 5. Number of utilized text fragments.

適宜 R-DTF リストを参照しながら内容の拡充を行っていった。その後、被験者 9 は実験開始から 27.5 分後、被験者 10 は実験開始から 29.3 分後に本文の作成を開始した。被験者 11 は、まず R-DTF リストを参照しながら、テーマ 2 に関する自身の意見を構成メモとして箇条書きした。そして、実験開始 20.6 分後に本文の作成をいったん開始した。しかしながら、実験開始 48 分後にこれまで作成した本文をすべて削除し、もう一度新たに本文の作成を開始し始めた。このとき、新たな本文に記載のテーマ 2 に対する被験者 11 の意見は、削除された本文の内容に記載されていたものとは異なる内容であった。インタビューより、被験者 11 は R-DTF リストを参照しているうちにテーマ 2 に対する意見が変わったため、一から本文作成を開始したと実験を振り返っていた。このとき、被験者 11 は特定の R-DTF によって意見が変わったのではなく、複数の R-DTF に影響されたと述べていた。被験者 12 は、まず R-DTF リストを参照しながら気になったキーワードを構成メモとして抽出した。その後、抽出したキーワードを組み合わせる短文を複数生成し、自身がテーマ 2 について書き出したいことの探索を始めた。そして実験開始から 37.5 分後に本文作成を開始した。

5.2.3 文章作成時に参照された外部知識について

(1) 条件 2

図 5 は、条件 2 における被験者が文章作成の際に活用した文章断片の活用個数を、事前実験の被験者別に集計した棒グラフである。図 5 に示すように、被験者 5-8 は、文章作成をするに当たり、事前実験で得られた本文を参照し、そのうち 3 個から 7 個の文章断片を活用した。各被験者が活用した文章断片は様々であったが、特に、被験者 S1 お

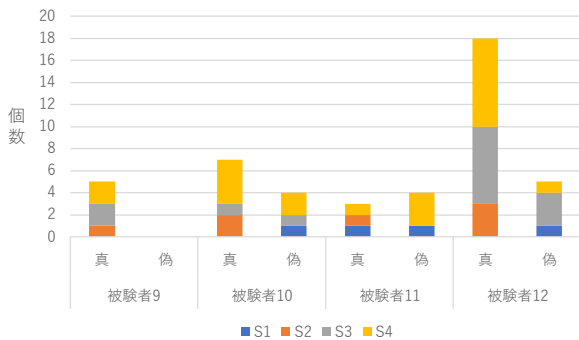


図 6. R-DTF の活用個数

Fig. 6. Number of utilized R-DTFs.

表 6. 活用した R-DTF の重複数

Table 6. Duplicate number of utilized R-DTFs.

	被験者9	被験者10	被験者11	被験者12
被験者9	0	1	0	4
被験者10	1	0	2	9
被験者11	0	2	0	3
被験者12	4	9	3	0

および被験者 S4 が生成した文章断片が多く活用されていた。

インタビューの結果、被験者 5 および被験者 8 は自身の文章作成における内容を補完する目的で文章断片を活用していた。被験者 5 は、活用した 5 つの文章断片のうち、2 つの文章断片の内容の一部を直接自身の文章の内容に組み込み、残りの 3 つについては、文章断片に記載された単語を見て、文章に記載すべき内容を発想するきっかけとして活用していた。被験者 8 は、3 つの文章断片全てについて、自身が文章を書く上での発想のきっかけとして活用していた。

被験者 6 および被験者 7 は、まず被験者 S1-S4 が作成した本文を参照し、他者がテーマ 2 についてどのようなことを書いているかを把握した後、自身が書くべき内容を思案していた。そして、被験者 6 および被験者 7 はそれぞれ、その内容を発想するきっかけとなる文章断片についてチェックボックスにチェックを入れていた。被験者 6 はさらに、文章作成時に自身の記述内容に近い文章断片を探し出し、内容として使えそうな文章断片にチェックを入れていた。被験者 7 は、文章作成時に使えそうな記述がないかを網羅的に探索していたが、結果として活用するには至らなかった。

インタビューと動画分析の結果、被験者 6 は文章断片を積極的に活用する姿勢を見せていたが、その他の 3 人の被験者は、テーマ 2 に対する自身の意見に対して、他者の本文はあくまで参考程度のもので活用していた。

(2) 条件 3

図 6 は、条件 3 における被験者が文章作成の際に活用し

た R-DTF の数を、事前実験の被験者別に集計した棒グラフである。また図 6 は、真の R-DTF と偽の R-DTF を区別して集計してある。図 6 に示すように、条件 3 における被験者は、少なくとも 3 つは真の R-DTF を活用していた。また、条件 2 においては、被験者 S2 および S3 が生成した文章断片が活用された数が非常に少なかったのに比べ、R-DTF においては、図 6 に示すように S2 および S3 が生成したのもも活発に活用されていることが分かる。特に、S2 が生成した R-DTF のうち、活用されたものはすべて真の R-DTF であった。さらに、事前実験で収集した真の R-DTF の割合は、約 71% であったのに対し、条件 3 にて活用された真の R-DTF の割合は約 72% であり、事前実験における R-DTF の真偽の割合と同様の割合にて R-DTF が活用されていることが分かった。

表 6 は、被験者それぞれが活用した R-DTF のうち、被験者間で重複しているものの数を集計した表である。表 4 において、被験者 9 が作成した文章と被験者 11 が作成した文章の類似度は約 0.65 と比較的高い。一方、表 6 に示すように、被験者 9 と被験者 11 が活用した R-DTF の重複数は 0 である。また、表 6 に示すように被験者 10 と被験者 12 が活用した R-DTF の重複数は 9 である。一方、表 4 に示すように、被験者 10 が作成した文章と被験者 12 が作成した文章の類似度は 0.5 を下回っている。したがって、作成される文章は、R-DTF の活用起因して似通うわけではない。

また、インタビューの結果から、被験者 9 および被験者 10 は、自身の構成メモの作成に当たり欠けている視点を、R-DTF を活用することで補っていた。このとき、両者は主に R-DTF に記載の単語を補う視点として活用していた。さらに、被験者 9 および被験者 10 は、文字数が多すぎる R-DTF は、活用しづらいと回答した。また、被験者 10 は、文章作成時には自身の文章表現と同内容の R-DTF を探し、活用できないかどうかを検討していた。さらに、条件 3 におけるすべての被験者が、R-DTF を活用していたのは、主に構成メモを作成する段階であった。また、すべての被験者は、テーマ 2 とは明らかに異なるテーマ、すなわちテーマ 3 および 4 の R-DTF は参照時に活用可能性が低いとして読み飛ばしていたと回答した。

6. 議論

5 章の実験結果から、新たな創造的文章の作成時における R-DTF の活用可能性について議論する。まず、図 6 に示したように新たな創造的文章作成時において R-DTF は実際に活用された。条件 3 の被験者は真の R-DTF についても偽の R-DTF についても同様に活用していることから、他者が本文に採用するのに有用と判断した文章断片（偽の R-DTF）であろうが、不用と判断した文章断片であろうが（真の R-DTF）、分け隔てなく活用されることが分かった。

一方、R-DTF が活用されたからと言って、作成される文

章の内容については、他条件、特に条件2で作成された文章と異なる内容を示すわけではない。具体的には、表4中、条件2の被験者と条件3の被験者に対応する類似度群は、その他の条件の組み合わせによる類似度群と有意な差がない。これは、R-DTFが活用される主なタイミングが構成メモを作成する段階であったことに起因すると考えられる。つまり、条件3の被験者は従来のContents Reuseのように、R-DTFを本文作成に当たって直接的に活用したわけではないため、本文の内容に条件2には存在しない新たな内容としてのR-DTFの特徴が直接的には反映されない。

しかしながら、インタビューの結果から、他者の本文を参照するよりもR-DTFを参照した方が、執筆者に与える影響が大きいことが分かった。具体的には、条件2の被験者の大半は、参照領域内の本文は参考程度にしか活用しなかったのに対し、条件3においては、被験者11がR-DTFによって自身の意見を変化させ、被験者12がR-DTFを用いることによって自身の意見を探索していた。これは、R-DTFが執筆者の文章作成行動に強く影響を及ぼしていることを表している。

このような事象が生じた原因の一つに、参照する外部知識の液状化[17]の程度が関係すると思われる。液状化は、堀らの研究グループが提唱した概念であり[17]、知識がより大きく分節している状態を指す。本稿の事象で言えば、条件2の本文よりも条件3のR-DTFの方が、知識の液状化の度合いが高いため、執筆者の知識と結合しやすいと思われる。つまり、知識を液状化した状態で提示されたR-DTFは、完成した文章を提示するよりも、より活用されやすい状態であると考えられる。

以上より、R-DTFは新たな創造的文章を作成するに当たり、有効に活用可能であると結論付けることができる。

7. まとめ

本稿では、R-DTFを収集する事前実験と、収集されたR-DTFが創造的文章の作成で活用されるかどうかを検証する活用実験を行った。結果として、R-DTFが新たな創造的文章において活用されることを確認した。また、R-DTFは、一般的な活用対象である文章よりも液状化の程度が高いために、執筆者の文章構成により強い影響を及ぼすという示唆を得た。

さらにR-DTFを有効に活用可能とするためには、液状化の粒度を調整し、R-DTFと執筆者の知識が結晶化して新たな知識創造を行うことができるようになる方策を検討する必要がある。また、R-DTFは文章構成段階で活用されることが主であるため、5章の実験における条件3のようなインタフェースよりも、文章構成段階を支援可能なインタフェース(例えばText ComposTerのインタフェース)にR-DTFの提示機能を追加することによって、より活用しやすくなると思われる。さらに、文章作成時のテーマとかけ離れた

テーマから生成されたR-DTFは、活用対象から活用されにくいことが、5章の実験より判明した。そのため、このようなR-DTFが活用されるためには、かけ離れたテーマであることを執筆者に気付かせないようにする処理が必要である。例えば、テーマ固有の単語(トピック)をR-DTFから除外することで、より有効に活用されるものと思われる。今後は、これらの実験と検討の結果を踏まえ、実用的な棄却文章断片の活用環境の構築を目指す。

謝辞 本研究は、JSPS 科研費 JP15K12093 の助成を受けたものです。本稿の執筆に当たり実験に協力下さった被験者の方々に謝意を表します。

参考文献

- [1] 湯川 抗: IT を活用した知識創造社会の実現にむけて-プラットフォームとしてのコミュニティ-, Economic review. 7(1), pp. 108-134, (2003)..
- [2] Neuwirth, C., Kaufer, D., Chimera, R. and Gillespie, T.: The Notes Program: A Hypertext Application for Writing from Source Texts , Proc. of the ACM Conference on Hypertext, pp. 121-142, (1987).
- [3] Smith, J. B., Weiss, S. F. and Ferguson, G. J.: A Hypertext Writing Environment and Its Cognitive Basis (Panel Session) , Proc. of the ACM Conference on Hypertext , pp. 195-214 (1987).
- [4] Hunter, W. J. and Begoray, J.: A Framework for the Activities Involved in the Writing Process, The Writing Notebook, vol. 7, No. 3, pp. 40-42 (1990).
- [5] Boden, M. A.: The Creative Mind: Myths and Mechanisms, 2nd ed. London: Routledge (2004)
- [6] Jensen. S. : Where Good Ideas Come From: The Natural History of Innovation, Riverhead Books (2010)
- [7] 生田泰章, 才記駿平, 西本一志: 文章作成過程における棄却文章断片の活用に関する一検討. インタラクション2016 論文集. 2016, 1B35, p. 302-305.
- [8] Ikuta, H. and Nishimoto, K.: Wasting “Waste” is a Waste: Gleaning Deleted Text Fragments for Use in Future Knowledge Creation, Proc. of The Tenth International Conference on Advances in Computer-Human Interactions, pp.193-199, 2017.
- [9] 生田泰章, 西本一志: Con-Text ComposTer: 棄却テキスト断片の活用機会を創出する知識創造活動支援システム, インタラクション2017 論文集, 2-508-41, pp. 529-534, 2017.
- [10] Rockley, A. (2002). Managing Enterprise Content: A Unified Content Strategy (Indianapolis: NewRiders)
- [11] Katrien Verbert, Xavier Ochoa, and Erik Duval. The alocom framework: Towards scalable content reuse, Journal of Digital Information, 9, 2008.
- [12] Mejova, Y., Schepper, K. D., Bergman, L., and Lu, J.: Reuse in the wild: An empirical and ethnographic study of organizational content reuse, Proc. of the 2011 Annual Conference on Human factors in Computing Systems, pp. 2877– 2886, (2011).
- [13] Jensen, C., Lonsdale, H., Wynn, E., Cao, J., Slater, M., Dietterich, T. G.: The Life and Times of Files and Information: A Study of Desktop Provenance, CHI '10.
- [14] Nakakoji, K., Yamamoto, Y., Reeves, B.N., and Takada, S.: Two-Dimensional Positioning as a Means for Reflection in Design, Proc. of Design of Interactive Systems, pp.145-154, (2000).
- [15] 浅原 正幸, 加藤 祥. 文書間類似度について. 自然言語処理. 23 卷 (2016) 5 号 p. 463-499
- [16] Kudo, T, Yamamoto, K., Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis, Proc. of the 2004 Conference on Empirical Methods in Natural Language Processing, pp.230-237 (2004.)
- [17] 堀浩一: 創造活動支援の理論と応用, オーム社 (2007).