

## 編集操作の測定でアプローチする自然言語処理の提案

## Natural Language Processing by Writing Measurement

山口琢<sup>\*1</sup>  
Taku Yamaguchi大場みち子<sup>\*2</sup>  
Michiko Oba高橋慈子<sup>\*3</sup>  
Shigeko Takahashi小林龍生<sup>\*4</sup>  
Tatsuo Kobayashi高橋修<sup>\*2</sup>  
Osamu Takahashi<sup>\*1</sup>公立はこだて未来大学大学院  
Graduate School of Future University Hakodate<sup>\*2</sup>公立はこだて未来大学  
Faculty of Future University Hakodate<sup>\*3</sup>株式会社ハーティネス  
Heartiness Co., Ltd.<sup>\*4</sup>有限会社スコレックス  
Scholex Co., Ltd.

We propose to study Natural Language Processing by measuring human writing process. In that approach, we record text editing operations by that metrics and analyze the time series (point process) of those operations. In the approach, the input text and the names of the files are not recorded. The metrics are designed to be neutral from models of text logical structure and writing teaching methods. The metrics are designed to be means for comparing and verifying those models and teaching methods. We will show the examples of analysis and the outlook of AI application.

## 1. はじめに

書記体系 (writing system) のデジタル化が進み、多くの読み書きが IT に移行してきた。これによって「IT による読み書き行為」を、大規模・多様・きめ細かく測定できるようになってきた。これは電子書籍の普及によるところが大きい。本がデジタル化されることで、Web ページだけでなく長文の読みも IT ツールを使って行われるようになった。読みの測定は実用化・商用化が進んでいる。電子書籍端末は、読者の読みを測定し活用している。電子書籍フォーマットが EPUB によって標準化されたことで、測定の規模はさらに大きく、また長期に及ぶことが可能となった。教育分野では、EPUB を含む EDUPUB によって、測定の標準化も始まっている [IMS 15].

われわれは作文 (writing, または文章産出 text production) に着目している。本稿では、作文 (文章産出) 過程の測定・分析によって自然言語理解にアプローチする手法を提案する。IT によるテキスト編集操作をモデル化・測定し、操作の時系列 (点過程) を分析する。例として共起分析を紹介する。

## 2. 作文行為の測定

本稿では、作文行為の測定と分析を区別して考える (図 1)。測定結果のデータは、さまざまな目的の分析で使える方がよい。測定では、この中立性を重視する。文章産出過程モデル、文章論理構造モデルや作文指導法などからの中立性を重視し、これらの比較・検証手段となることを目指す。1 つ公開コーパスに対して、それを使った分析・研究が複数あるのと同様である。

作文行為の測定とは、いつ (時刻)、どこ (文章の単位) を、どのように編集したかを記録することである。「文章の単位」や「どのように」は測定が前提とする文章編集モデルに関わる [山口 14]。本稿では測定についてこれ以上は触れない。以降、「文章の単位」≡「段落」、「どのように」≡「挿入」「更新」「削除」と考えてよい。

なお、この測定では、入力・編集された文字や文章ファイル

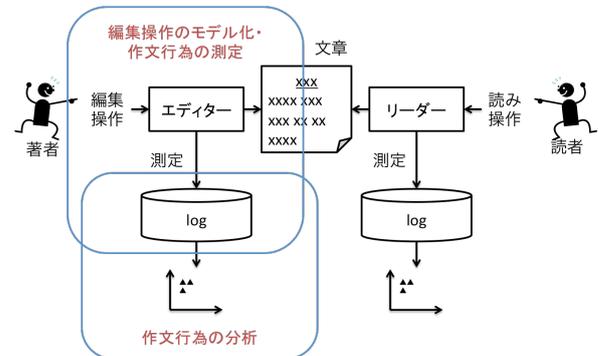


図 1: IT による作文行為のモデル化・測定と分析

名は記録していない。後で述べる分析でも、公開ワークシートの見出し以外、入力されたテキストを必要としない。

## 3. 作文行為の分析: 共起

分析の例として、テキスト分析での共起頻度分析にならって、作文行為すなわち編集操作の共起を考える。共起が起きる場所・区間は、時間軸、あるいは時刻を無視した編集操作列である。その時間軸・編集操作列のある区間で一緒に現れることを共起と呼ぶ。その編集操作の対象となった文章の単位に着目して、「共起する編集操作の対象である文章の単位は、互いに関係が深い」と考えるのが、作文行為に対する共起分析のアイデアである。

測定データを利用する分析は他にもあると考えられるが、本稿では共起分析に絞って述べる。

## 4. ワークシートによる作文の共起分析

大学のロジカルライティングの講義で、「プレゼン評価」という課題作文を行った。このときの測定データを共起分析した例を示す。

受講生を複数グループに分けて、ある課題について提案ス



図 2: ワークシートによるプレゼン評価

ライドを作成して模擬プレゼンした。その後、公開ワークシートに従って他グループのプレゼンを評価するように求めた。図 2 は、ワークシートをロードしたエディター・アプリ画面のスナップショットである。「ロジカルなプレゼンテーションとは何か?」、「他のグループのプレゼンを聞いてのコメント」、「自分のグループのプレゼンテーションの企画・実施に関するコメント」と見出しのついた 3 つの記入欄が、左から右に並んでいる。

このワークシートへ記入する、1 つの作文を測定したデータの共起行列が図 3 である。「cc\_8」等は記入欄の ID であり、同じ公開ワークシートを使うと同じ ID になる。「cc\_8」は「ロジカルなプレゼンテーションとは何か?」、「cc\_10」は「他のグループのプレゼンを聞いてのコメント」、「cc\_12」は「自分のグループのプレゼンテーションの企画・実施に関するコメント」の記入欄 (段落) に対応する。左上から右下への対角線上のセルの数値は、続けて同じ段落に対して編集操作をした回数である。その右隣のセルは、右隣の記入欄 (段落) へ編集対象が移ったことに該当する。この対角線 2 線以上のセルは、ワークシート上で左から右に書き進んだことに該当する。この斜め 2 線から外れる操作が、右方向へジャンプしたり、左に戻ったりした編集行為に該当する。わざわざそのように編集したことから、該当する記入欄 (段落) 同士は書き手にとって関係が深いと考えられる。具体的には「cc\_8」と「cc\_12」、「ロジカルなプレゼンテーションとは何か?」と「自分のグループのプレゼンテーションの企画・実施に関するコメント」とが関係が深いであろう。

もっと大きな (長い) 文章の作文については、図 4 のような共起「散布図」を得ることができる。円の大きさが頻度に該当する。文章の最初から最後に向けて書いているが、ジャンプして先へ進んだり後へ戻ったりしている様子がわかる。

## 5. 共起頻度の解釈

このような共起頻度データを大量の作文行為について取得して統計処理することで、新たな知見が得られると期待できる。具体的には、ワークシートの比較、書き方の比較、ワークシートを処理させているエディター・アプリの比較、作文指導法の比較などである。

例えば斜め 2 線にしかデータが現れない場合、「左から右に書き進めば済むようにできている、良いワークシートだ」、あるいは「文章の各单位 (段落) 間の整合性を、あまり考えない書き手が多い」と判断できるかもしれない。指導前後のパターンを比較して指導効果の指標としたり、2 つのワークシートを比較したり、2 つのアプリ UI の比較をする応用が考えられる。

n \ n+1	cc_8	cc_10	cc_12	other
cc_8	2	1	1	0
cc_10	0	1	1	0
cc_12	1	0	2	0
other	0	0	0	0

図 3: ワークシート編集操作の共起行列

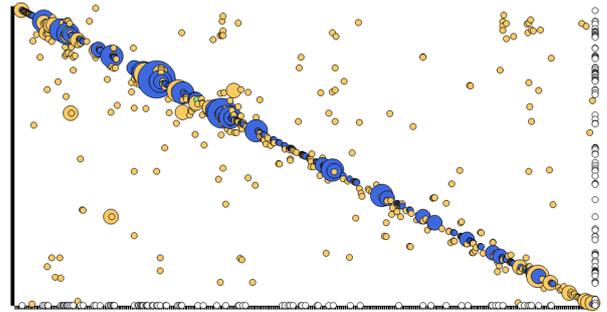


図 4: 編集操作の共起散布図

## 6. AI への応用

### 6.1 テキスト分析との連携

書き手の作文行為を分析した結果、文章内の離れた 2 つ段落の関係が深いと判断できれば、その結果をテキスト分析の結果と照合したり、テキスト分析の際の重み付けなどに利用できる可能性がある。これは読書の測定・分析についても同様である。

### 6.2 自動アドバイス

書き進め方の観点から、書き手にアドバイスすることも考えられる。「左から右 (上から下) に一直線に書きすぎ。背景と結論の整合性を見よ」などのアドバイスである。野球に例えて、テキスト分析が「ホームランかヒットか」の結果判定だとすれば、作文行為分析は「バッターボックスの後ろに立て、バットを短く持て」といったプロセス・アドバイスに該当すると考えられる。

## 7. まとめと今後

テキスト編集操作の測定・分析による自然言語理解のアプローチを提案し、例として編集操作の共起分析を紹介した。今後は、分析例を増やすとともに、測定データの公開に取り組む。

## 参考文献

- [IMS 15] IMS Global : EDUPUB Summit, <http://imglobal.org/feb2015UofPhoenix.html#Summit> (2015)
- [山口 14] 山口琢, 大場みち子, 高橋修: 相互運用可能な作文計測システムの設計, 第 60 回情報システム研究会, 電気学会 電子・情報・システム部門 (2014)