Vol.2015-NL-220 No.15
2015/1/20

# 単一文書自動要約のための言語資源構築に向けて

浅原 正幸[1,a)]　加藤 祥[1]　今田 水穂[2]

**概要**：本稿では単一文書自動要約の新たな展開について言語資源と評価指標の観点から検討する。まず、最初に語順に対する順序尺度を含めた距離空間・類似度・相関係数・カーネルにより既存の自動評価指標の整理を行い、現在ある言語資源を用いてその指標空間の性質を明らかにする。次に自動要約の評価として必要な軸として、提供すべき情報の過不足と読みやすさの二つを考える。情報の過不足については、元文書の情報構造を言語生産者・言語受容者の双方の観点から分析し、システム要約・参照要約双方の情報の質を検討する。読みやすさについては、生成されたテキストの読み時間に基づいた定量的な評価方法について検討する。最後に語順・情報構造・読み時間の関係性について解説し、読み時間を用いた言語受容者毎の要約作成の可能性について議論する。

## 1. はじめに

本稿では『現代日本語書き言葉均衡コーパス』[1](以下BCCWJ) を基づいた単一文書自動要約のための言語資源構築について議論する。要約作成は、元文書の言語生産者 $A$ と元文書の言語受容者かつ要約文書の言語生産者 $B$ と要約文書の言語受容者 $C$ の 3 種類の言語使用者が介在する行為である。小さな組織においてはこの三者の認識がある程度統制することも可能であるが、完全に一致させることは難しい。

- $A$ による元文書の言語生産過程において複数回同じ課題を試行して完全に同一の文書が得られるだろうか
- $B$ による要約文書の言語生産過程において複数回同じ課題を試行して完全に同一の文書が得られるだろうか
- $B$ による元文書の言語受容過程において複数回同じ課題を試行して完全に同一箇所を重要視することがあるだろうか
- $C$ による要約文書の言語受容過程において複数回同じ課題を試行して完全に同一箇所を重要視することがあるだろうか

本稿は**言語の生産過程と言語の受容過程の非対称性**を定量的に評価することを出発点とする。

自動要約や機械翻訳ではシステム出力の内容評価を行うために参照要約 (翻訳) との類似度を評価するためのスコアがいくつか提案されている。上記、言語生産過程・言語受容過程をこれらのスコアにより評価することを検討したが、先行研究の様々な文献においてスコアに対して数理的な説明がなされておらず、引用や比較においてスコアに対する正しい認識がされていないことがわかった。上記評価を行う前に、距離・類似度・カーネル・順序尺度・相関係数などの多様な指標を用いて**先行研究の (類似度) スコアの数理的構造を説明**する。そのうえで一般化されたスコアを用いて言語の生産過程と言語の受容過程の不安定さの評価を試みる。$A$, $B$, $C$ の三者の認識の相違を示した上で、次に自動要約の内容評価と読みやすさ評価について検討する。

前者については、**情報構造 (Information Structure) のアノテーションによる内容評価**を試みる。情報構造は情報状態 (Information Status) と主題 (Topic) と焦点 (Focus) によって構成される。日本語は情報構造を「とりたて」や「提題」として形態論的に陽に表出する言語である。本稿では、情報構造うち最初の情報構造について文献 [2] に基づいて現代日本語記述文法の知見をとりいれながらアノテーションを行うことを試みる。情報状態は情報の新旧を、「言語生産者が旧情報として提示し言語受容者が旧情報と認識している対象」(old)、「言語生産者が旧情報としては提示しておらず言語受容者が新情報と認識している対象」(new)、「言語生産者が旧情報としては提示していないが言語受容者にとっては旧情報である対象」(accessible) の三つに分類し、主節の名詞句単位に付与することを試みる。自動要約の内容評価としては、各分野の情報の重要性によって評価するのではなく、文法的に表出する情報状態に基づいて評価することを提案する。もし文法的に表出する情報状態が要約文作成に有効であることが示されれば、自動要約器の

[1] 人間文化研究機構 国立国語研究所
NINJAL, Tachikawa, Tokyo 190–8561, Japan
[2] 文部科学省
MEXT, Chiyoda, Tokyo 100–8959, Japan
[a)] masayu-a@ninjal.ac.jp

最適化すべき方向性を示すことになる。

後者については、テキスト受容過程における**読み時間による読みやすさの評価**を検討する。人による文処理時間は連続値で扱う毎ができ、統計的に扱いやすい。コーパスに基づく言語処理は、言語生産過程の成果物である生テキストからの学習と、練度の高い言語受容過程の成果物であるアノテーションからの学習が一般的である。受容過程の分析においては少しずつ被験者実験の記録に基づく言語処理が検討されているが、依然として少ない。読み時間を用いたテキスト受容過程の定量評価の方法論を検討し、そのために必要な言語資源構築について議論する。

最後に、スコアのうちの語順の順序尺度と情報構造や、情報構造と読み時間の関連性についての先行研究について紹介しながら、読み時間を手がかりとした言語受容者毎の自動要約作成の可能性について検討する。

本稿の貢献は以下のとおりである:

- 既存の文書要約や機械翻訳の自動評価に利用される指標と距離空間・類似度・カーネル空間・順序尺度・相関係数など多様な指標との関係を整理した (§2.1-§2.5)
- 複数人の要約文書の言語生産者 $B$ 間で生成される文書のゆれを定量的に評価することを試みた (§2.7)
- 同一人の通常の言語生産者 $A$ の課題試行間で生成される文書のゆれと同一人の要約文書の言語生産者 $B$ の課題試行間で生成される文書のゆれ定量的に評価することを試みた (§2.7)
- 要約の内容的な評価を行うために要約元文書の情報構造を構成する情報状態のアノテーションを試みた (§3)
- 要約の読みやすさの評価を行うための読み時間の利用を提案し、その方法論について検討した (§4)
- 元文書の語順と読み時間と情報構造との関連についての先行研究について紹介し、読み時間を手がかりとした言語受容者毎の自動要約作成について検討した (§5)

本稿はまだアイデア段階の研究の紹介である。コメント・照会については第一著者まで。

## 2. 既存の自動評価指標の特性と言語生産・受容過程の多様性

### 2.1 本節の趣旨

まず、最初に語順に対する順序尺度を含めた距離空間・類似度・カーネル・相関係数により既存の自動評価指標の整理を行う。先行研究の文献では連続記号列を表す部分文字列 (substring) とギャップを許す部分列 (subsequence) との混同が見られ、定性的な議論が弱い。本稿では、大きく分けて一致部分文字列による尺度・一致部分列による尺度・ベクトル型順序尺度・編集型順序尺度の四つに分類し議論する。**自動評価指標のまとめのみについてのみ知りたい方は §2.5 の表 1 を参照されたい。**

次に言語生産・受容過程の多様性を 4 種類の尺度により評価する。複数人が同一課題を実施した場合の各尺度の分散や、同一人が同一課題繰り返し実施した場合の各尺度の分散などを検討する。生産過程においては口述・筆術・タイプ入力の 3 種類について評価し、課題においては要約・語釈・再話について評価する。

なお、本節で用いる用語や記号の定義は §A.1 にまとめてある。

### 2.2 LCStr, LCS

#### 2.2.1 記号列と文字列と部分文字列と部分列

評価尺度の議論を始める前に、記号列と文字列と部分文字列と部分列の違いについて確認する。

何らかの全順序が付与されている記号集合のことを**記号列**と呼ぶ。本稿では記号列ベクトル $s = \langle s_1, \ldots, s_m \rangle, t = \langle t_1, \ldots, t_m \rangle$ などで表現する。元文書、要約文書は、ともに文字 (character) ベースの記号列もしくは形態素解析後の形態素 (morpheme) ベースの記号列とみなすことができる。

評価する記号列上の連続列のことを**文字列 (string)** と呼ぶ。記号列の要素が文字 (character) である場合を「文字ベースの文字列 (character-based string)」、記号列の要素が形態素 (morpheme) である場合を「形態素ベースの文字列 (morpheme-based)」と呼ぶこととする。

記号列に対して隣接性と順序を保持した部分的記号列のことを**部分文字列 (substring)** と呼ぶ。長さ $n$ の部分文字列を特に n-gram 部分文字列と呼ぶ。記号列 $s$ の $i$ 番目の要素からはじまる n-gram 部分文字列を $s_{i,\ldots,i-n+1}$ で表現する。

記号列に対して順序を保持した部分的記号列のことを**部分列 (subsequence)** と呼ぶ。隣接性は保持しなくてよい。長さ $p$ の部分列を特に p-mer 部分列と呼ぶ。記号列 $s$ の p-mer 部分列を、インデックスベクトル $\vec{i} = \langle i_1, \ldots, i_p \rangle (1 \leq i_1 < i_2 < \cdots < i_p \leq |s|)$ を用いて、$s[\vec{i}]$ と表す。

#### 2.2.2 最長共通部分文字列 (Longest Common String: LCStr) 長

最長共通部分文字列 (Longest Common String) の abbreviation は LCS だが、一般には 2.2.2 に示す最長共通部分列 (Longest Common Subsequence) のことを LCS と呼ぶことが多い。本稿では前者を LCStr, 後者を LCS と呼び、区別する。

記号列 $s, t$ を与えた際の最長共通部分文字列を次式で定義する:

$$\mathrm{LCStr}(s,t) = \underset{s_{i,\ldots,i-n+1} | \exists j, s_{i,\ldots,i-n+1} = t_{j,\ldots,j-n+1}}{\arg\max} n$$

記号列 $s, t$ を与えた際の最長共通部分文字列長 (LCStr 長) を次式で定義する:

$$|\mathrm{LCStr}|(s,t) = \max_{\forall i,\forall j,s_{i,\ldots,i-n+1}=t_{j,\ldots,j-n+1}} n$$

これを [0,1] 区間に正規化すると以下のようになる:

$$\mathrm{Score}_{\mathrm{LCStr}}(s,t) = \frac{2\cdot|\mathrm{LCStr}|}{|s|+|t|}$$

#### 2.2.3 最長共通部分列 (Longest Common Subsequence: LCS) 長と Levenshtein 距離

記号列 $s, t$ を与えた際の最長共通部分列 (Longest Common Subsequence: LCS) を次式で定義する:

$$\mathrm{LCS}(s,t) = \operatorname*{arg\,max}_{s[\vec{i}]\exists\vec{j},s[\vec{i}]=t[\vec{j}]} |\vec{i}|$$

記号列 $s, t$ を与えた際の最長共通部分列長 (LCS 長) を次式で定義する:

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

[0,1] 区間に正規化すると、以下のようになる:

$$\mathrm{Score}_{\mathrm{LCS}}(s,t) = \frac{2\cdot|\mathrm{LCS}|}{|s|+|t|}$$

なお、挿入のコストを 1、削除のコストを 1、代入のコストを 2(もしくは代入を禁止) した場合の Levenshtein 距離 (編集型) と LCS 長の関係は以下のようになる:

$$\mathrm{d}_{\mathrm{Levenshtein}}(s,t) = |s|+|t|-2\cdot|\mathrm{LCS}|$$

さらに LCS は §2.4.2.2 で示すとおり、対称群上の編集型距離のうちの Ulam 距離と深く関連し、一種の順序尺度であるとも考えられる。

#### 2.2.4 ギャップ加重最長共通部分列長によるスコア

部分列 LCS は部分文字列 LCStr と異なりギャップを伴う。ギャップの多い LCS に減衰させた値を割り当てるために、「LCS の記号列上の長さ」に対して加重を行うことができる。「LCS の記号列上の長さ」は参照要約側 ($|\mathrm{LCS}(C,R)|_R$) とシステム出力要約側 ($|\mathrm{LCS}(C,R)|_C$) とで異なるためにそれぞれ計算する必要がある。

$$|\mathrm{LCS}(C,R)|_R = \operatorname*{arg\,max}_{(j_{|\vec{j}|}-j_1)|\forall\vec{i},\forall\vec{j},C[\vec{i}]=R[\vec{j}]} |\vec{i}|$$

$$|\mathrm{LCS}(C,R)|_C = \operatorname*{arg\,max}_{(i_{|\vec{i}|}-i_1)|\forall\vec{i},\forall\vec{j},C[\vec{i}]=R[\vec{j}]} |\vec{i}|$$

参照要約側で重みを付けて正規化する再現率的なスコアを $\mathrm{R}_{\mathrm{WLCS}}(C,R)$ とし、システム出力要約側で重みを付けて正規化する精度的なスコアを $\mathrm{P}_{\mathrm{WLCS}}(C,R)$ とすると以下のようになる。

$$\mathrm{R}_{\mathrm{WLCS}}(C,R) = \frac{\alpha^{|\mathrm{LCS}|_R(C,R)-|\mathrm{LCS}|}\cdot|\mathrm{LCS}|}{|R|}$$

$$\mathrm{P}_{\mathrm{WLCS}}(C,R) = \frac{\alpha^{|\mathrm{LCS}|_C(C,R)-|\mathrm{LCS}|}\cdot|\mathrm{LCS}|}{|s|}$$

全体を正規化すると以下のようになる。

$$\mathrm{Score}^{(\gamma)}_{\mathrm{WLCS}}(C,R) = \frac{(1+\gamma^2)R_{\mathrm{WLCS}}(C,R)P_{\mathrm{WLCS}}(C,R)}{R_{\mathrm{WLCS}}(C,R)+\gamma^2P_{\mathrm{WLCS}}(C,R)}$$

### 2.3 既存の自動評価指標

次に自動要約と機械翻訳の自動評価指標をレビューするが、基本的には文単位の評価かつ参照要約/翻訳が一つであるという仮定をおく。

#### 2.3.1 要約の評価指標

##### 2.3.1.1 ROUGE-L [3]

ROUGE-L [3] は システム出力要約と参照要約の最長共通部分列 (LCS) 長をスコアとして正規化したものである。

$$\mathrm{Score}^{(\gamma)}_{\mathrm{ROUGE\text{-}L}}(C,R) = \frac{(1+\gamma^2)\cdot R_{\mathrm{LCS}}(C,R)\cdot P_{\mathrm{LCS}}(C,R)}{R_{\mathrm{LCS}}(C,R)+\gamma^2P_{\mathrm{LCS}}(C,R)}$$

ここで再現率に相当する $R_{\mathrm{LCS}}(C,R)$ と精度に相当する $P_{\mathrm{LCS}}(C,R)$ は以下のように定義する:

$$R_{\mathrm{LCS}}(C,R) = \frac{|\mathrm{LCS(C,R)}|}{|R|}$$
$$P_{\mathrm{LCS}}(C,R) = \frac{|\mathrm{LCS(C,R)}|}{|C|}$$

上記指標は文単位のものであり、文書レベルに拡張するために、システム出力要約中の文 $c_i \in C$ と参照要約中の文 $r_j \in R$ の LCS 記号列中の記号の集合和を用いて評価する。同様の議論が他の指標においても行われているが、以下本稿ではこの議論を省略する。

##### 2.3.1.2 ROUGE-W [3]

ギャップ加重最長共通部分列長に似た概念である。違いとしては「LCS の記号列上の長さ」を参照要約側とシステム出力要約側 $|\mathrm{LCS}(C,R)|_R + |\mathrm{LCS}(C,R)|_C$ でとった上で、加重関数 $f(x) : f(x+y) > f(x)+f(y), x>0, y>0, x\in N, y\in N$ ($N$ は自然数) を別に定義して「LCS の記号列上の長さ」に対して加重を行う。ROUGE-W の実装では $f(x)=x^\alpha$ という多項式を用いており、ギャップ加重最長共通部分列長 $\mathrm{Score}^{(\gamma)}_{\mathrm{WLCS}}(C,R)$ の一般化と考えることができる。

##### 2.3.1.3 ROUGE-N [3], [4]

ROUGE-N [3], [4] は n-gram の一致度をスコアとして用いるものである。

$$\mathrm{Score}^{(R)}_{\text{ROUGE-N}}(C,R)=\frac{\sum_{e\in\text{n-gram}_{\text{clip}}(C,R)}|e|}{\sum_{e\in\text{n-gram}(R)}|e|}$$

但し、$|e|$ は $e$ の要素数、n-gram$(C)$ はシステム要約 C に含まれる n-gram 集合、n-gram$(R)$ は参照要約 R に含まれる n-gram 集合とする。$\text{n-gram}_{\text{clip}}(C,R)$ はシステム要約に含まれる n-gram の、システム要約に含まれる出現頻度 $|e\in\text{n-gram}(C)|$ と参照要約に含まれる出現頻度 $|e\in\text{n-gram}(R)|$ の小さい方の集合とし、次式で定義する:

$$\text{n-gram}_{\text{clip}}(C,R)=\begin{cases}\text{n-gram}(C)\text{ if }|\text{n-gram}(C)|\le|\text{n-gram}(R)|\\ \text{n-gram}(R)\text{ otherwise}\end{cases}$$

#### 2.3.1.4 ROUGE-S(U) [3], [5]

ROUGE-S は 2-mer の部分列の一致度をスコアとして用いるものである。

$$\mathrm{Score}^{(\gamma)}_{\text{ROUGE-S}}(C,R)=\frac{(1+\gamma^2)P_s(C,R)R_s(C,R)}{R_S(C,R)+\gamma^2P_S(C,R)}$$

ここで精度に相当する $P_S(C,R)$ と再現率に相当する $R_S(C,R)$ は以下のように定義する:

$$P_S(C,R)=\frac{\sum_{e\in\text{2-mer}_{\text{clip}}(C,R)}|e|}{\sum_{e\in\text{2-mer}(C)}|e|}$$

$$R_S(C,R)=\frac{\sum_{e\in\text{2-mer}_{\text{clip}}(C,R)}|e|}{\sum_{e\in\text{2-mer}(R)}|e|}$$

但し、p-mer$(C)$:参照要約に含まれる p-mer 部分列集合、p-mer$(R)$:参照要約に含まれる p-mer 部分列集合とする。$\text{p-mer}_{\text{clip}}(C,R)$ はシステム要約に含まれる p-mer 部分列の出現頻度 $|e\in\text{p-mer}(C)|$ と参照要約に含まれる p-mer 部分列の出現頻度 $|e\in\text{p-mer}(R)|$ の小さい方の集合とし、次式で定義する:

$$\text{p-mer}_{\text{clip}}(C,R)=\begin{cases}\text{p-mer}(C)\text{ if }|\text{p-mer}(C)|\le|\text{p-mer}(R)|\\ \text{p-mer}(R)\text{ otherwise}\end{cases}$$

ROUGE-SU は上に ROUGE-S の $p=2$ を $p\le 2$ に拡張したものである。

#### 2.3.1.5 ESK [6]

ESK [6] は畳み込みカーネルの一つである拡張文字列カーネルのうち、ギャップ加重 p-mer 部分列カーネルを評価指標として定義したものである。

$$\mathrm{Score}^{\text{p-mer}}_{\text{ESK}}(C,R)=\frac{\sum_{u\in\text{p-mer}(C)}\sum_{v\in\text{p-mer}(R)}\lambda^{|e|-p}\delta(u,v)|u||v|}{\sqrt{\left(\sum_{u,u'\in\text{p-mer}(C)}\lambda^{(|e|-p)}|u||u'|\right)+\left(\sum_{v,v'\in\text{p-mer}(R)}\lambda^{(|e|-p)}|v||v'|\right)}}$$

文献 [6] では 2-mer の部分列に制限するほか、文単位にスコア比較し精度重視の指標と再現度重視の二つの調和平均を定義している。

### 2.3.2 翻訳の評価指標

#### 2.3.2.1 BLEU[7]

BLEU [7] は機械翻訳評価のための指標で、$n$ の値を変えた n-gram の精度系の指標の重み $(\omega_n)$ 付き相乗平均によりスコアを定義する。

$$P^{\text{n-gram}}_{\text{BLEU}}(C,R)=\frac{\sum_{e\in\text{n-gram}_{\text{clip}}(C,R)}|e|}{\sum_{e\in\text{n-gram}(C)}|e|}$$

$$\mathrm{Score}_{\text{BLEU}}(C,R)=BP(C,R)\cdot\exp\left(\sum_{n=1}^{N}\omega_n\log P^{\text{n-gram}}_{\text{BLEU}}(C,R)\right)$$

ここで相乗平均の計算を簡単にするために $\sum_n^N\omega_n=1$ という制約がある。

短いシステム翻訳に対して高い精度が出やすいこの精度系の指標に対し、精度と再現率の重み付き調和平均という方法を取らず、Brevity Penalty (BP) という項を入れて補正している。

$$\text{BP}(C,R)=\begin{cases}1 & \text{if }|C|>|R|\\ \exp(1-\frac{r}{c}) & \text{if }|C|\ge|R|\end{cases}$$

#### 2.3.2.2 IMPACT [8]

我々の理解が正しければ、IMPACT[8] は LCS に基づく指標ではなく、LCStr の再帰的な取得による指標である。

$$R_{IP}(C,R)=\left(\frac{\sum_{r=0}^{\text{RN}}(\alpha^r\sum_{e\in\text{LCStr}(C^{(r)},R^{(r)})}|e|^\beta)}{|R|^\beta}\right)^{\frac{1}{\beta}}$$

$$P_{IP}(C,R)=\left(\frac{\sum_{r=0}^{\text{RN}}(\alpha^i\sum_{e\in\text{LCStr}(C^{(r)},R^{(r)})}|e|^\beta)}{|C|^\beta}\right)^{\frac{1}{\beta}}$$

ここで $\alpha$ はイテレート回数 $r(r\le\text{RN})$ に対する重み

($\alpha < 1.0$)、$\beta$ は LCStr 長に対する重み ($\beta > 1.0$)、$C^{(1)} = C$、$R^{(1)} = R$、$C^{(r)} = C^{(r-1)} \setminus \{\mathrm{LCStr}(C^{(r-1)}, R^{(r-1)})\}$、$R^{(r)} = R^{(r-1)} \setminus \{\mathrm{LCStr}(C^{(r-1)}, R^{(r-1)})\}$ とする。

$$\mathrm{Score}_{\mathrm{IP}} = \frac{(1+\gamma^2)R_{\mathrm{IP}}P_{\mathrm{IP}}}{R_{\mathrm{IP}} + \gamma^2 P_{\mathrm{IP}}}$$

この指標は 2.4.1.1 節に示す文字列長加重全部分文字列カーネルに関連が深い。文字列長加重全部分文字列カーネルに対して、再帰的に LCStr を選択する際に既選択の LCStr を排除し、再帰の回数を RN で制限するという制約を入れたものである。

#### 2.3.2.3 RIBES [9]

RIBES [9] は、システム翻訳と参照翻訳のアラインメントをとったうえで、語順の編集型順序尺度を考慮したものである。

$$\mathrm{Score}_{\mathrm{RIBES}} = \left(d_{\mathrm{Kendall}}(\text{1-gram}_{\mathrm{align}}(C,R))\right) \cdot \left(P_{\mathrm{RIBES}}(C,R)\right)^{\alpha} \cdot \left(\mathrm{BP}(C,R)\right)^{\beta}$$

ここで $d_{\mathrm{Kendall}}(\mu,\nu)$ は 2.4.2.2 で定義する順位ベクトル $\mu,\nu$ に対する Kendall 距離、$\text{1-gram}_{\mathrm{align}}(\mu,\nu)$ は元論文 [9] の `wonder` で出力されるアラインメントされた二つの順序ベクトルの対を表す。左辺 2 項目は 1-gram(単語ベースのもの) 精度とよび $P_{\mathrm{RIBES}}(C,R) = \frac{|\text{1-gram}_{\mathrm{align}}(C,R)|}{|C|}$ とする。$|\text{1-gram}_{\mathrm{align}}(\mu,\nu)|$ は `wonder` で出力されるアラインメントされた順序ベクトルの長さ (二つ出力されるが等しい)。

$\alpha$ は記号精度に対する重み、$\beta$ は BLEU で用いられた BP に対する重みである。

なお、$P_{\mathrm{RIBES}}(C,R)$ は、それぞれの記号列に重複する記号がない場合、以下が成り立つ:

$$\begin{aligned} P_{\mathrm{RIBES}}(C,R) &= \mathrm{Score}^{(P)}_{\text{ROUGE-1}}(C,R) \\ &= \frac{\sum_{e \in \text{1-gram}_{\mathrm{clip}}(C,R)} |e|}{\sum_{e \in \text{1-gram}(R)} |e|} \end{aligned}$$

#### 2.3.2.4 LRscore [10]

LRscore [10] も同様に、アラインメントをとったうえで、語順の順序尺度を考慮したものである。順序尺度としてベクトル型である Hamming 距離と編集型である Kendall 距離を用いている。

$$\mathrm{Score}^{\mathrm{Hamming}}_{\mathrm{LRscore}}(C,R) = \alpha \cdot BP(C,R) * d_{\mathrm{Hamming}}(\hat{C},\hat{R}) + (1-\alpha)\mathrm{Score}_{\mathrm{BLEU}}$$

$$\mathrm{Score}^{\mathrm{Kendall}}_{\mathrm{LRscore}}(C,R) = \alpha \cdot BP(C,R) * d_{\mathrm{Kendall}}(\hat{C},\hat{R}) + (1-\alpha)\mathrm{Score}_{\mathrm{BLEU}}$$

### 2.4 関連するカーネル・順序尺度

上に述べた指標は、基本的には以下のカーネルおよび順序尺度の組み合わせで構成することができる。以下では、各種指標に関連するカーネルおよび順序尺度について確認する。

#### 2.4.1 カーネル・距離 (文字列の共有)

畳み込みカーネルのうち系列データに対するカーネル [11] は、共通する可能な部分文字列・部分列を数え上げる。いずれも効率よく計数する方法が提案されている。また、適切に正規化することにより部分文字列・部分列の共有についての距離やスコアを規定することができる。

様々なカーネルの説明に入る前に、スコア化 ([0,1] 区間正規化) について示す。カーネルのスコア化はカーネルの研究分野でよく用いられており以下の式により行われる:

$$\mathrm{Score}_{K_-}(s,t) = \frac{K_-(s,t)}{||K_-(s,s)||||K_-(t,t)||}$$

各種指標のように、再現率-精度間の重み $\gamma$ を入れたい場合には以下のようにする:

$$\mathrm{Score}^{(\gamma)}_{K_-}(s,t) = \frac{(1+\gamma^2)K_-(s,t)}{\sqrt{(K_-(s,s))^2 + \gamma^2(K_-(t,t))^2}}$$

##### 2.4.1.1 全部分文字列カーネルと文字列長加重全部分文字列カーネル

全部分文字列カーネル (All String Kernel or Exact Matching Kernel) は共通する全ての部分文字列の数を数える。

長さ $n$ の部分文字列 $u$ を座標とする特徴量空間 $F_{\text{all\_str}}$ を考える。

$$\begin{aligned} \Phi^*_{\mathrm{str}} : \sigma^* &\to F_{\text{all\_str}} \sim R^{|\sigma|^*} \\ \Phi^*_{\mathrm{str}} &= (\phi^*_u(s))_{u \in \sigma^*} \\ K_{\text{n-gram}}(s,t) &= \langle \Phi^*_{\mathrm{str}}(s), \Phi^*_{\mathrm{str}}(t) \rangle_{F_{\text{all\_str}}} \\ &= \sum_{u \in \sigma^*} \phi^*_u(s)\phi^*_t(s) \\ \phi^*_u(s) &= |\{i | s_{i\ldots *} = u\}| \end{aligned}$$

カーネル関数を直接計算すると以下のようになる:

$$K_{\text{all\_seq}}(s,t) = \sum_{n=1}^{\min(|s|,|t|)} \sum_{i=1}^{|s|-n+1} \sum_{j=1}^{|t|-n+1} \delta(s_{i\ldots i+n-1}, t_{i\ldots i+n-1})$$

このカーネルは、提案された 2002 年ごろではバイオインフォマティクスなど特定の分野以外では有効な用途が

提案されていない。言語処理の場合、得られる n-gram に対して加重をかけることが一般に行われている。例えば、文字列長に対して加重をかけたものを文字列長加重全部分文字列カーネル (Length Weighted All String Kernel or Length Weighted Exact Matching Kernel) と呼ぶ。

$$K_{\text{all\_seq}}(s,t) = \sum_{n=1}^{\min(|s|,|t|)} \sum_{i=1}^{|s|-n+1} \sum_{j=1}^{|t|-n+1} \omega_{|s|}\delta(s_{i\ldots i+n-1}, t_{i\ldots i+n-1})$$

ここで $\omega_n$ は長さ $n$ に対する重みを表す。

§2.3.2.2 で述べた IMPACT はこのカーネルの特殊形とみなすことができる。

このカーネルと次の n-スペクトラムカーネルは Suffix Tree を用いて効率よく計算する方法が提案されている。

#### 2.4.1.2 n-スペクトラムカーネル

n-gram スペクトラムカーネル (Spectrum Kernel) は共通する長さ $n$ の部分文字列 (n-gram) の数を数える。

長さ $n$ の部分文字列 $u$ を座標とする特徴量空間 $F_{\text{n-gram}}$ を考える。

$$\begin{aligned}
\Phi^n_{\text{str}} : \sigma^* &\rightarrow F_{\text{n-gram}} \sim R^{|\sigma|^n} \\
\Phi^n_{\text{str}} &= (\phi^n_u(s))_{u\in\sigma^n} \\
K_{\text{n-gram}}(s,t) &= \langle \Phi^n_{\text{str}}(s), \Phi^n_{\text{str}}(t)\rangle_{F_{\text{n-gram}}} \\
&= \sum_{u\in\sigma^p} \phi^n_u(s)\phi^n_t(s) \\
\phi^n_u(s) &= |\{i | s_{i\ldots i+n-1} = u\}|
\end{aligned}$$

直接計算すると以下のようになる:

$$K_{\text{n-gram}}(s,t) = \sum_{i=1}^{|s|-n+1} \sum_{j=1}^{|t|-n+1} \delta(s_{i\ldots i+n-1}, t_{j\ldots j+n-1})$$

ROUGE-N は、分子に $K_{\text{n-gram}}(C,R)$ より小さい値を持ち、分母に参照要約ののべ出力 n-gram 数を持つことから、再現率として正規化する。通常の正規化した $K_{\text{n-gram}}(s,t)$ は再現率と精度の調和平均と解釈できる。

また 1-gram スペクトラムカーネルは 1-mer 部分列カーネルと同値で、これらは近似的に BLEU などで利用されている BP 相当の値を計算すると考える。

#### 2.4.1.3 全部分列カーネル

全部分列カーネルは共通するすべての部分列の数を数える。

任意の長さの部分列 v を座標とする特徴量空間 $F_{\text{all\_seq}}$ を考える。

$$\begin{aligned}
\Psi^*_{\text{seq}} : \sigma^* &\rightarrow F_{\text{all_seq}} \sim R^{|\sigma|^\infty} \\
\Psi^*_{\text{seq}}(s) &= (\psi^*_v(s))_{v\in\sigma^*} \\
\psi^*_v(s) &= |\{\vec{i} | s[\vec{i}] = v\}| \\
K_{\text{all_seq}}(s,t) &= \langle \Psi^*_{\text{seq}}(s), \Psi^*_{\text{seq}}(t)\rangle_{F_{\text{all_seq}}} \\
&= \sum_{v\in\sigma^*} \psi^*_v(s)\cdot\psi^*_v(t)
\end{aligned}$$

ここで $\psi^*_v(s) = |\{\vec{i} | s[\vec{i}] = v\}|$ とする。

$K_{\text{all\_seq}}(s,t)$ は以下のように再帰的に計算することにより $O(|s||t|)$ で計算することができる。$\epsilon$ を空記号列とすると $K_{\text{all\_seq}}(s,\epsilon) = K_{\text{all\_seq}}(t,\epsilon) = 1$ とし、$K_{\text{all\_seq}}(s,t)$ が求まると $K_{\text{all\_seq}}(s\cdot a,t) = K_{\text{all\_seq}}(s,t) + \sum_{1\le i\le|t|, j:t_j=a} K_{\text{all\_seq}}(s,t_{i\ldots j-1})$ と $s$ 再帰的に定義できる。さらに $\tilde{K}_{\text{all\_seq}}(s\cdot a,t) = K_{\text{all\_seq}}(s,t_{i\ldots j-1})$ とすると、$\tilde{K}_{\text{all\_seq}}(s\cdot a, t\cdot b) = \tilde{K}_{\text{all\_seq}}(s\cdot a,t) + \delta(a,b)K(s,t)$ と $t$ 再帰的に定義できる。

#### 2.4.1.4 固定長部分列カーネル

固定長部分列カーネルは共通する長さ $p$ の部分列 (p-mer) の数を数えあげる。

長さ $p$ の部分文字列 $v$ を座標とする特徴量空間 $F_{\text{p-mer}}$ を考える。

$$\begin{aligned}
\Psi^p_{\text{seq}} : \sigma^* &\rightarrow F_{\text{p-mer}} \sim R^{|\sigma|^p} \\
\Psi^p_{\text{seq}}(s) &= (\psi^p_v(s))_{v\in\sigma^*} \\
\psi^*_v(s) &= |\{\vec{i} | s[\vec{i}] = v\}| \\
K_{\text{p-mer}}(s,t) &= \langle \Psi^p_{\text{seq}}(s), \Psi^p_{\text{seq}}(t)\rangle_{F_{\text{p-mer}}} \\
&= \sum_{v\in\sigma^p} \psi^p_v(s)\cdot\psi^p_v(t)
\end{aligned}$$

ここで $\psi^p_v(s) = |\{\vec{i} | s[\vec{i}] = v\}|$ とする。

ROUGE-S は、分子に $K_{\text{2-mer}}(C,R)$ より小さい値を持ち、分母に参照要約ののべ出力 2-mer 数を持つことから、再現率として正規化する。ROUGE-SU は、分子に $K_{\text{1-mer,2-mer}}(C,R)$ より小さい値を持ち、分母に参照要約ののべ出力 1-mer, 2-mer 数を持つことから、再現率として正規化する。通常の正規化した $K_{\text{p-mer}}(s,t)$ は再現率と精度の調和平均と解釈できる

#### 2.4.1.5 ギャップ加重部分列カーネル

ギャップ加重部分列カーネル: p-mer の部分列の数え上げの際に隣接性を考慮して重み $\lambda$ を加重する。ESK [6]、このカーネルを用いたスコアである。

長さ $p$ の部分列 $v$ を座標とする特徴量空間 $F_{\text{p-mer}}$ を考える。

**表 1** 指標・スコア・距離・カーネル・相関係数の関係まとめ

| | 指標 (要約系) | (翻訳系) | スコア $[0,1]\uparrow$ | 距離 $[0,\infty]\downarrow$ | カーネル $[0,\infty]\uparrow$ | 相関係数 $[-1,1]\uparrow$ |
|---|---|---|---|---|---|---|
| 部分文字列系 | | IMPACT §2.3.2.2 [8] | $\mathrm{Score}^{(\gamma)}_{K\text{all\_str}}$ | | (加重) 全部分文字列 §2.4.1.1 | |
| (n-gram) | ROUGE-N §2.3.1.3 | BLEU §2.3.2.1[7] | $\mathrm{Score}^{(\gamma)}_{K\text{n-gram}}$ | | n-スペクトラム §2.4.1.2 | |
| | | LRscore§2.3.2.4[10] | | | | |
| 部分列系 | | | $\mathrm{Score}^{(\gamma)}_{K\text{all\_seq}}$ | | (加重) 全部分列 §2.4.1.3 | |
| (p-mer) | ROUGE-S(U) §2.3.1.4 [3], [5] | | $\mathrm{Score}^{(\gamma)}_{K\text{p-mer}}$ | | p-mer 部分列 §2.4.1.4 | |
| | ESK §2.3.1.5[6] | | $\mathrm{Score}^{(\gamma)}_{K\text{gap\_p-mer}}$ | | 加重 p-mer 部分列 §2.4.1.5 | |
| 順序系 §2.4.2.1 | | | $\mathrm{Score}_{\|\mathrm{rank}\|_\theta}$ | | | |
| (ベクトル型) | | | $\mathrm{Score}_{\mathrm{footrule}}$ | $d_{\mathrm{footrule}(\theta=1)}$ | | |
| | | RIBES? §2.3.2.3 [9] | $\mathrm{Score}_{\mathrm{Spearman}}$ | $(d_{\mathrm{Spearman}(\theta=2)^2})$ | | Spearman's $\rho$ |
| | | | | | | $\subset$ Pearson's |
| | | LRscore§2.3.2.4[10] | $\mathrm{Score}_{\mathrm{Hamming}}$ | $d_{\mathrm{Hamming}}$ | | |
| 順序系 §2.4.2.2 | | RIBES§2.3.2.3[9] | $\mathrm{Score}_{\mathrm{Kendall}}$ | $d_{\mathrm{Kendall}}$ | | Kendall's $\tau$ |
| (編集型) | | LRscore§2.3.2.4[10] | | | | |
| | | | | $d_{\mathrm{Caylay}}$ | | |
| | | | | $d_{\mathrm{Ulam}}$ | | |
| (最長一致部分列長) | ROUGE-L§2.3.1.1 | | $\mathrm{Score}_{\mathrm{LCS}}$ | $d_{\mathrm{Levenshtein}}$§2.2.3 | | |
| (加重最長一致部分列長) | ROUGE-W §2.3.1.2[3] | | $\mathrm{Score}^{(\gamma)}_{\mathrm{WLCS}}$ | | | |
| (最長一致部分文字列長) | | | $\mathrm{Score}_{\mathrm{LCStr}}$ | | | |

Kendall $d_{\mathrm{Kendall}}((1,4,3,2),(1,2,3,4))=3$

$$\begin{pmatrix}1&4&\underline{3}&\underline{2}\\1&4&\underline{2}&\underline{3}\end{pmatrix}\Rightarrow\begin{pmatrix}1&\underline{4}&\underline{2}&3\\1&\underline{2}&\underline{4}&3\end{pmatrix}\Rightarrow\begin{pmatrix}1&2&\underline{4}&\underline{3}\\1&2&\underline{3}&\underline{4}\end{pmatrix}$$

$d_{\mathrm{Kendall}}((2,3,1,4),(1,2,3,4))=2$

$$\begin{pmatrix}2&\underline{3}&\underline{1}&4\\2&\underline{1}&\underline{3}&4\end{pmatrix}\Rightarrow\begin{pmatrix}\underline{2}&\underline{1}&3&4\\\underline{1}&\underline{2}&3&4\end{pmatrix}$$

Caylay $d_{\mathrm{Caylay}}((1,4,3,2),(1,2,3,4))=1$

$$\begin{pmatrix}1&\underline{4}&3&\underline{2}\\1&\underline{2}&3&\underline{4}\end{pmatrix}$$

$d_{\mathrm{Caylay}}((2,3,1,4),(1,2,3,4))=2$

$$\begin{pmatrix}\underline{2}&3&\underline{1}&4\\\underline{1}&3&\underline{2}&4\end{pmatrix}\Rightarrow\begin{pmatrix}1&\underline{3}&\underline{2}&4\\1&\underline{2}&\underline{3}&4\end{pmatrix}$$

Ulam $d_{\mathrm{Ulam}}((1,4,3,2),(1,2,3,4))=2$

$$\begin{pmatrix}1&\boxed{4&3}&\underline{2}\\1&\underline{2}&\boxed{4&3}\end{pmatrix}\Rightarrow\begin{pmatrix}1&2&\boxed{4}&\underline{3}\\1&2&\underline{3}&\boxed{4}\end{pmatrix}$$

$d_{\mathrm{Ulam}}((2,3,1,4),(1,2,3,4))=1$

$$\begin{pmatrix}\boxed{2&3}&\underline{1}&4\\\underline{1}&\boxed{2&3}&4\end{pmatrix}$$

**図 1** 対称群上の編集型距離

$$K_{\text{gap\_p-mer}}(s,t) = \langle \Psi_{\text{seq}}^{gap\text{-}p}(s), \Psi_{\text{seq}}^{gap\text{-}p}(t) \rangle_{F_{\text{p-mer}}} = \sum_{v \in \sigma^p} \psi_v^{gap\text{-}p}(s) \cdot \psi_v^{gap\text{-}p}(t)$$

ここで $\psi_v^{gap\text{-}p}(s) = \sum_{\vec{i}:v=s[\vec{i}]} \lambda^{l(\vec{i})}$ とし、$l(i) = |s_{i_1,\ldots,i_{|v|}}|(\vec{i} = \langle i_1, \ldots, i_{|v|} \rangle)$ とする。

#### 2.4.2 順序尺度

以下では順序尺度について考えるが、文献 [12] が詳しい。基本的には同じ長さ $m$ の二つの順位ベクトル $\mu, \nu \in S_m$ に対する 2 種類の距離を考える。

##### 2.4.2.1 順位ベクトル型距離

一つ目の距離は「順位ベクトル型」の距離で順位ベクトルを $m$ 次元空間中の点を表すベクトルとみなし、ベクトル空間上の距離を定義する。ベクトル空間を $\theta$-ノルム採用すると以下のようになる:

$$d_{||\text{Rank}||_\theta}(\mu,\nu) = (\sum_{i=1}^{m} |\mu(i) - \nu(i)|^\theta)^{1/\theta}$$

ここで $\theta = 1$ の場合、特に Spearman footrule と呼ぶ。

$$d_{\text{Footrule}}(\mu,\nu) = (\sum_{i=1}^{m} |\mu(i) - \nu(i)|)$$

$\theta = 2$ の場合は通常の Euclid 距離だが、この Euclid 距離を 2 乗したものを特に Spearman 距離と呼ぶ。

$$d_{\text{Spearman}}(\mu,\nu) = (\sum_{i=1}^{m} |\mu(i) - \nu(i)|^2)$$

Spearman 距離は、距離の公理のうち対称性と正定値性を満たす。しかし、Euclid 距離を 2 乗したものなので三角不等式を満たさないが、慣習的として距離として扱われる。さらに [-1, 1] 区間に正規化したものは Spearman の順位相関係数 $\rho$ として知られている。

$$\text{Spearman's } \rho = 1 - \frac{6 \cdot d_{\text{Spearman}}(\mu,\nu)}{m^3 - m}$$

この値は順序尺度に基づく二つの順位ベクトル $\mu, \nu$ の Pearson 相関関係と等しい [*1]。

その他、順位ベクトルの同一順位のものが同じ要素である要素数を数えた Hamming 距離がある。

$$d_{\text{Hamming}}(\mu,\nu) = \sum_{i=1}^{m} \delta(\mu(i), \nu(i))$$

Hamming 距離は文字列上で代入 (コスト 1) のみを許した編集距離としても解釈できる。

[*1] ここで順序尺度とは、間隔に意味がある間隔尺度を順位のみに変換していることを前提にしている。

##### 2.4.2.2 対称群上の編集型距離

二つ目の距離は「編集型」の距離である。

順序ベクトルを記号列とみなした場合、順位ベクトル $\mu$ をもうひとつの順位ベクトル $\nu$ に変換するために必要な最小操作数を Levenshtein 距離について述べた。以下では、順序ベクトルを対称群とみなした場合の編集型距離について述べる。編集に許される操作によっていくつかの距離のバリエーションがある。図 1 に順序ベクトルによる置換により表現した編集型距離を示す。

- Kendall 距離:
  Kendall 距離 $d_{\text{Kendall}}$ は順序ベクトルを対称群とみなした際に隣接互換によって置換する最小回数によって定義される。言い換えると隣接する対象対を交換 (Swap) する操作の最小回数を用いたものである。Kendall 距離は、二つの順位ベクトル中の $\frac{m(m-1)}{2}$ 個の対象対のうち逆順になっている対の数に等しい。

$$d_{\text{Kendall}} = \min(\arg\max_q \delta((\Pi_{q=1}^{q} \pi_2(k_q, k_q{+}1)) \cdot \mu, \nu))$$

$$d_{\text{Kendall}} = \sum_{i=1}^{m} \sum_{j=i+1}^{m} \chi(i,j)$$

ここで $\chi$ は対象対 $\langle i,j \rangle$ が同順のとき 0、逆順のとき 1 を返す指示関数:

$$\chi = \begin{cases} 1 & \text{if } (\mu(i) - \mu(j))(\nu(i) - \nu(j)) < 0, \\ 0 & \text{if } (\mu(i) - \mu(j))(\nu(i) - \nu(j)) \geq 0 \end{cases}$$

これをスコアとして使いやすくするために [0,1] 区間の範囲に正規化すると以下のようになる:

$$\text{Score}_{\text{Kendall}} = 1 - \frac{2 \cdot d_{\text{Kendall}}(\mu,\nu)}{m^2 - m}$$

これを [-1,1] 区間の範囲に正規化したものは Kendall の順位相関係数 $\tau$ として知られている。

$$\text{Kendall's } \tau = 1 - \frac{4 \cdot d_{\text{Kendall}}(\mu,\nu)}{m^2 - m}$$

- Cayley 距離:
  Cayley 距離 $d_{\text{Caylay}}$ は順序ベクトルを対称群とみなした際に隣接互換によって置換する最小回数によって定義される。言い換えると隣接していなくても良い対象対を交換 (Swap) する最小回数を用いたものである。

$$d_{\text{Caylay}} = \min(\arg\max_q \delta((\Pi_{q=1}^{q} \pi_2(k_q, l_q)) \cdot \mu, \nu))$$

- Ulam 距離:
  Ulam 距離 $d_{\text{Ulam}}$ は順序ベクトルを対称群とみなした際に連続した順序ベクトル部分列 $i, i+1, \ldots, j-1, j\rangle$

の巡回置換の操作のみによって置換する最小回数によって定義される。これは「本棚の本の入れ換え」で例えられる。順位ベクトル $\mu$ で並んでいる本棚の本を順位ベクトル $\nu$ に並び替えるために、ある要素を抜いて別の場所に挿入するということを行う。
Ulam 距離は同じ要素が記号列に存在しないという前提のもと、最大共通部分列距離と以下の関係にあることが知られている。

$$d_{\mathrm{Ulam}}(\mu,\nu) = m - |\mathrm{LCS}(\mu,\nu)|$$

これを [0,1] 区間の範囲に正規化すると以下のように正規化最大共通部分スコアと同じになる:

$$\begin{aligned}\mathrm{Score}_{\mathrm{Ulam}}(\mu,\nu) &= 1 - \frac{d_{\mathrm{Ulam}}(\mu,\nu)}{m} \\ &= \frac{|\mathrm{LCS}(\mu,\nu)|}{m} \\ &= \mathrm{Score}_{\mathrm{LCS}}(\mu,\nu)\end{aligned}$$

以下は、我々の意見だが、言語生産時の編集作業において [13] の swap に代表されるような Kendall 距離的編集よりも Ulam 距離のような編集の方が自然なのではないかと考える。

#### 2.4.2.3 順序尺度間の関係

ベクトル型の Spearman's $\rho$ と Kendall's $\tau$ との間には以下の Daniels の不等式が成立する:

$$-1 \leq \frac{3(m+2)}{m-2}\tau - \frac{2(m+1)}{m-2}\rho \leq 1$$

$m \to \infty$ の極限をとると $-1 \leq 3\tau - 2\rho \leq 1$ が成り立つ。このことから二つの相関係数の間には高い相関があることが示される。

距離の観点からは、$d_{\mathrm{Caylay}} \leq d_{\mathrm{Kendall}}$ が成り立つ。さらに Footrule 距離と Kendall 距離と Cayley 距離の間に以下の不等式が成り立つ (Diaconis-Graham inequality):

$d_{\mathrm{Kendall}} + d_{\mathrm{Caylay}} \leq d_{\mathrm{Footrule}} \leq 2 \cdot d_{\mathrm{Kendall}}$

また Spearman 距離と Kendall の距離の間には以下の不等式が成り立つ (Durbin-Stuart inequality):

$\frac{4}{3} d_{\mathrm{Kendall}}(1 + \frac{d_{\mathrm{Kendall}}}{m}) \leq d_{\mathrm{Spearman}}$

スコアのデザインにおける順序尺度の選択による効果は、あくまでこれらの不等式の範囲によって抑えられる。

### 2.5 スコアの一般化

以上、指標・スコア・距離・カーネル・相関係数を議論してきた。まとめると表 1 のようになる。

各スコアと人手の評価結果という観点からすると、[14] のように、表 1 にあげたすべてのスコア $\mathrm{Score}_{\_} \in \{\mathrm{Score}_*\}$ の加重相乗平均 (下式) を考え、加重 $\omega_{\_}$ と各スコアに付随するパラメータを各指標の従属性や相関に注意しながら人手の評価指標との回帰により求めれば良い。

$$\overline{\mathrm{Score}_*} = \sum \omega_{\_} \sqrt[\omega_{\_}]{\Pi \mathrm{Score}_{\_}^{\omega_{\_}}}$$

$$\log \overline{\mathrm{Score}_*} = \frac{1}{\sum \omega_{\_}} (\sum w_{\_} \cdot \log \mathrm{Score}_{\_})$$

このスコアのあり方については議論すべき点がいくつかある。

- substring(部分文字列: n-gram 系) と subsequence(部分系列: p-mer 系) との違いを踏まえる。
- 最長一致部分文字列は対称群上の編集型距離である Ulam 距離と深く関連する。
- 順序に対する順位ベクトル型距離と編集型距離の間には 2.4.2.3 節に示される関係が成り立つ。

本稿では**スコアの一般化についてはこれ以上踏み込まない**。次節以降各スコアがさまざまな言語資源上でどのような振る舞いをするのかについてみていきたい。

### 2.6 評価に用いる言語資源

本稿では次節以降に述べるように人手の評価結果の再構築を視野に入れているため、ここでは研究室で有する言語資源のテキスト対のスコアを検証することにより、各スコアがとらえようとしているものが何なのかを分析する。

表 2 に利用する言語資源について示す。まず言語生産の目的として、要約 (BCCWJ-SUMM) と語釈 (GROSS) と再話 (RETELLING) の 3 種類の言語資源を準備する。要約と語釈については、クラウドソーシングにより安価で大量にデータを得る手法 (タイプ入力) と実験室にて被験者に繰り返し同一課題を依頼してデータを得る手法 (筆述) の 2 種類の方法を用いた。再話のデータについては既存のデータを用いた。再話については、言語生産形態として筆述による形態と口述による形態のデータを準備した。

以下各言語資源について解説する。

#### 2.6.1 BCCWJ-SUMM_C

BCCWJ-SUMM_C は BCCWJ の新聞記事の要約を Yahoo! クラウドソーシング (15 歳以上の男女) により被験者実験的に作成したものである。

BCCWJ の 1 サンプルには複数の記事が含まれており、それを記事単位に分割したうえで元文書集合 19 文書を構築した。元文書集合は BCCWJ コアデータ PN サンプル (優先順位 A) から選択した。40 文字毎に改行した元文書を画像として提供し、実験協力者に 50-100 文字に要約せよという指示で収集した。実験協力者の環境は PC 環境に限定した。元文書毎に約 100〜200 人の実験協力者が要約に従事した。実験実施時期は 2014 年 9 月である。

得られたデータには、文字数制限を守っていないもの・実験の趣旨を理解していないもの・既に実験を行った実験協力者から同一回答を提供されたと考えられるものなどが

**表 2** 指標評価に使う言語資源

| 言語資源名 | 収集場所 | 生成過程 | 繰り返し | 取得人数 | 摘要 |
|---|---|---|---|---|---|
| BCCWJ-SUMM_C | クラウドソーシング | タイプ入力 | なし | 100-200 | 19 文書の要約 |
| BCCWJ-SUMM_L | 実験室 | 筆述 | 3 回 | のべ 47 | 8 文書の要約 |
| GROSS_C | クラウドソーシング | タイプ入力 | なし | 71,111,113 | 鶏・兎・象の語釈 |
| GROSS_L | 実験室 | 筆述 | 4 回 | 7,6,3 | 鶏・兎・象の語釈 |
| RETELLING_I | 実験室 | 口述 | 10 回 | 5 | インタビュー |
| RETELLING_K | 実験室 | 口述 | 3 回 | 3,3,3 | 怪談 3 種の再話 |
| RETELLING_M | 実験室 | 筆述 | 4 回 | 10 | 物語「桃太郎」の再話 |

含まれており、これらを排除したものを有効要約とする。統計分析においてこの有効要約のみを用いる。

得られたデータ 19 文書の統計は表 3 のとおり。収集要約数はクラウドソーシングで得られたファイルの総数で、有効要約数は要約以外の意見陳述などのファイルを排除して、規定の文字数を満たしているものの総数。

**表 3** BCCWJ-SUMM_C データ概要

| FileID | 有効要約数 | 収集要約数 |
|---|---|---|
| A_01 | 106 | 198 |
| A_02 | 112 | 195 |
| B_02 | 98 | 149 |
| B_03 | 74 | 100 |
| C_01 | 63 | 100 |
| C_02 | 63 | 99 |
| C_03 | 53 | 100 |
| D_01 | 55 | 100 |
| D_02 | 55 | 100 |
| D_03 | 48 | 99 |
| D_05 | 55 | 99 |
| E_01 | 58 | 99 |
| E_02 | 46 | 98 |
| E_03 | 54 | 100 |
| E_04 | 60 | 99 |
| E_05 | 48 | 100 |
| E_06 | 56 | 98 |
| F_01 | 57 | 100 |
| F_02 | 58 | 100 |

### 2.6.2 BCCWJ-SUMM_L

BCCWJ-SUMM_L は BCCWJ の新聞記事の要約を実験室環境で筆述により作成したものである。BCCWJ-SUMM_C で用いた元文書を印刷紙面で提供し、実験協力者に 50-100 文字に要約せよという指示で収集した。一つの元文書に対して、3 回まで繰り返して要約文作成を行った。繰り返しに際しては、特別に「前と同じ要約文を作成してください」などといった指示は行わず、質問された場合にも「自由に要約文を作成してください」と教示した。実験協力者は原稿用紙上で筆述 (鉛筆と消しゴム利用) で要約を行い、そのデータを電子化した。

現在のところデータは 8 文書のべ 47 人分に限定した。得られたデータの概要は表 4 のとおり。

**表 4** BCCWJ-SUMM_L データ概要

| FileID | 有効要約数 | 被験者数 |
|---|---|---|
| A_01 | 16 | 6 |
| A_02 | 15 | 5 |
| B_02 | 15 | 5 |
| B_03 | 18 | 6 |
| C_01 | 15 | 5 |
| C_02 | 15 | 5 |
| C_03 | 15 | 5 |
| Q | 30 | 10 |

本実験の実験参加者は要約作業前に要約元文書の読み時間のデータも取得している。さらに 4.3 節に述べる被験者の特性 (最終学歴・語彙数・言語形成地・記憶力) などのデータが利用できる。実験実施時期は 2014 年 8 月〜10 月であるが、今後このデータは引き続き拡充していく予定である。

統計分析においては、同一課題について、異なる被験者間のスコア (1 回目のみを評価: BCCWJ-SUMM_L(P)) と、同一被験者の回数間のスコア (BCCWJ-SUMM_L(T)) の両方を評価する。

### 2.6.3 GROSS_C

GROSS_C は語釈文を Yahoo! クラウドソーシング (15 歳以上の男女) により被験者実験的に作成したものである。

「その動物を知らない人がどのようなものかわかるように説明してください」と教示し、同意した実験協力者は兎 (単語親密度 6.6)・鶏 (6.4)・象 (同 6.0) の 3 種類から対象物を選択回答した [*2]。150 文字以上 250 文字以内で 3 文字以上の同文字連続は認めない設定とした。実験協力者 300 名を募集したところ得られた解答数は、鶏:71・兎:111・象:113(295/300) であった。

### 2.6.4 GROSS_L

GROSS_L は語釈文を実験室環境で筆述により収集したものである。

実験協力者 8 名 (20 代-50 代の男女) に、GROSS_C と同様に「その動物を全く知らない人がどのようなものかわかるように説明してください」と教示した。実験協力者は、10 分間で兎 (単語親密度 6.6)・鶏 (6.4)・象 (同 6.0) の 3 種類から 2 種類の対象物を選択回答した。目安として 5 分経

[*2] 単語親密度は [15] による。

過時にブザー音を鳴らした。選択した対象物について同様に記述を繰り返すことを 4 回行った。得られた解答数は、兎 7 人分× 4 回、鶏 6 人分× 4 回、象 3 人分× 4 回である。平均 145 文字 (max 227 文字, min 85 文字) を得た。

統計分析においては、同一課題について、異なる被験者間のスコア (1 回目のみを評価: GROSS_L(P)) と、同一被験者の回数間のスコア (GROSS_L(T)) の両方を評価する。

#### 2.6.5 RETELLING_I

最初の再話のデータは「独話 Retelling コーパス」[16], [17] である。このコーパスは [18] でも用いられている。

実験協力者は 5 名で、同一人が同内容をそれぞれ 10 回独話を繰り返した。就職活動を前提とした模擬面接の設定で、実験協力者は自ら予め用意した「学生生活で力を入れてきたこと (3 分間程度)」についての独話を行った。同内容を繰り返すことや何回依頼するかは知らせていない。5 人分× 10 回 (50 話分) の独話を取得した。面接官 (聴衆) は有無を交互とした。奇数回 (1・3・5・7・9 回) は聴衆なしの独話、偶数回 (2・4・6・8・10 回) は聴衆に対する独話である。聴衆には、聴いていることを表すために頷くことのみを許可しており、話者への質問や意見など、発話は一切行わなかった。収録は録音と録画を行い、音声データを書き起こした。

被験者によってインタビュー内容が異なるために、統計分析においては同一被験者の回数間のスコア (RETELLING_I(T)) のみを評価する。

#### 2.6.6 RETELLING_K

次の再話のデータは怪談を繰り返し口述したものであり、先行研究 [19] によるものである。

実験協力者は 3 名 [*3] で、実験は 1 名ずつ個別に行った。実験協力者は怪談を聞いたのち、その怪談について 3 回の再話を行った。怪談は 3 種類を用意したため、各人 9 回の語りを行った。語りに関しては、「怪談として他の人に伝えるよう話す」との指示をした。既存の物語では、個人の記憶による先入観の影響が予測されたため、4 分間程度の新規な怪談を 3 本作成した。

実験環境は図 2 のように、ビデオカメラと録音機により、録音と録画を行った。聴衆の影響を除去するために、聴衆は設置しなかった。実験協力者は以下の配置で録音機に向かって話した。

↓□ (ビデオカメラ)
↓■ (録音機)
↑○ (実験協力者)

**図 2** RETELLING_K データの収録環境

本稿では音声データを書き起こしたものを用いる。

統計分析においては、同一課題について、異なる被験者間のスコア (1 回目のみを評価: RETELLING_K(P)) と、同一被験者の回数間のスコア (RETELLING_K(T)) の両方を評価する。

#### 2.6.7 RETELLING_M

最後の再話のデータは桃太郎の物語を筆述で繰り返し記述したものであり、先行研究 [20] によるものである。

実験協力者 10 名 (20 代-50 代の男女) に,「桃太郎の物語を全く知らない人に向けて記述してください」と教示し、実験協力者は 10 分間で記述 (筆述) した。同様に記述を繰り返すことを 4 回行った。平均延べ 284 語 (min:150 語・max:451 語)、異なり語 107 語 (min:74 語・max:152 語) の「桃太郎」 10 人分× 4 回 (40 話分) を取得した。

統計分析においては、同一課題について、異なる被験者間のスコア (1 回目のみを評価: RETELLING_M(P)) と、同一被験者の回数間のスコア (RETELLING_M(T)) の両方を評価する。

### 2.7 評価

本節では前節で述べたコーパスを用いて文書間距離がどのように振る舞うかを観察する。利用する文書間距離は以下の 30 種類である。

- n-gram スペクトラム (1,2,3,4) (char/mrph)
- n-gram 以下スペクトラム (≤2,≤3,≤4) (char/mrph)
- p-mer 部分列 (2,3,4) (char/mrph)
- p-mer 以下部分列 (≤2,≤3,≤4) (char/mrph)
- 1-gram スペクトラム+Footrule (char/mrph) (=Spearman)
- 1-gram スペクトラム+Kendall (char/mrph)

表 A·1,A·2 にそれぞれの距離空間によるスコアの平均値 (Mean) と標準偏差 (SD) を示す。スコアについて “_c” は文字単位の記号列として評価したもの、“_m” は形態素単位の記号列 (MeCab-0.98+IPADIC-2.7.0 による) として評価したものである。シャピロ・ウィルク検定の結果、ほとんどの場合 p 値が 0.05 未満であり、正規分布とはいえない傾向が見られた。

#### 2.7.1 スコアのグラフ

図 3 に形態素単位に評価した、n-gram(1),n-gram(2),p-mer(2),Kendall のスコアのグラフを示す。

見た目のレベルだが、unigram(n-gram(1)) を用いた場合、要約と語釈は中程度、再話はかなり高いスコアを達成している。GROSS_L(T) がほぼ再話と同程度のスコアで一方、BCCWJ-SUMM_L(T) が低いことから、要約を繰り返す際の言語生産の特殊性が見られる。要約を繰り返す際には、回数毎に文章中の重要箇所を変更するサンプル・被験者が存在し、標準偏差も高くなっている。

Bigram(n-gram(2)), skip-bigram(p-mer(2)) を用いた場合、異なる被験者間のスコアと繰り返し間のスコアとの間に差が見られるようになる。これは何らかの個々人の文体

[*3] 実験協力者 1 20 代・女性・東京都、実験協力者 2 30 代・女性・茨城県、実験協力者 3 20 代・女性・神奈川県

差が形態素の連接に影響を与えているのではないかと考える。

Bigram(n-gram(2)) と skip-bigram(p-mer(2)) の間の差として、語釈の場合のみ bigram のスコアが下がることがわかる。語釈という課題の都合上、物語や要約と異なり、情報の提示順が変わることも考えられる。しかし、順序尺度である Kendall のスコアでは bi-gram のスコアほど顕著な差が見られなかった。単語の隣接性が語釈のみ下がるというスコアの振る舞いについては今後検討していきたい。

クラウドソーシングと研究室内被験者実験との差 (BCCWJ-SUMM_C ⇔ BCCWJ-SUMM_L(P), GROSS_C ⇔ GROSS_L(P)) については、各スコア・各課題 (要約・語釈) で差が見られなかった。

#### 2.7.2 課題間の評価

以下、課題間を比較するために、6 種類の評価軸を分析する。殆どの場合、正規分布であることも等分散であること (F 検定による) も仮定できない。ここではウィルコクソンの順位和検定 (0.05 未満で 2 群の代表値が左右にずれている) を行う。 *4

- 実験室における複数人の課題間の違いの評価
  BCCWJ-SUMM_L(P) ⇔ GROSS_L(P) ⇔ RETELLING_K(P) ⇔ RETELLING_M(P)
  - BCCWJ-SUMM_L(P) ⇔ GROSS_L(P)
    文字単位の評価の場合 n-gram(2,3,4)_char, Kendall_char に有意差が見られた。
    形態素単位の評価の場合 n-gram(2,3,4,≤2,≤3,≤4)_mrph, Footrule_mrph, Kendall_mrph に有意差が見られた。
  - BCCWJ-SUMM_L(P) ⇔ RETELLING_K(P)
    n-gram(3,4)_mrph **以外**で有意差が見られた。
  - BCCWJ-SUMM_L(P) ⇔ RETELLING_K(M)
    全てのスコアについて、有意差が見られた。
  - GROSS_L(P) ⇔ RETELLING_{K,M}(P)
    全てのスコアについて、有意差が見られた。
  - RETELLING_K(P) ⇔ RETELLING_M(P)
    n-gram(≤3,≤4)_mrph,p-mer(3,4,≤3,≤4) で有意差が見られた。

  要約 ⇔ 語釈間は n-gram(1) で有意差が見られなかった。同じ文字・同じ形態素を使うという観点では一致度のレベルが等しいが、語の連接や順序尺度が入ると有意差が見られることがわかった。グラフの見た目から語釈の方が語の連接や順序尺度の一致度が低い。これは語釈の目的としては情報の提示順に重要性がないことが伺える。

  要約 ⇔ 再話、語釈 ⇔ 再話の間においては有意差が見られた。再話は同じ話をするという特性から、一致度が高くなる一方、要約・語釈は目的を達成するがために同じ表現を用いなければならないという制約がなく、低くなる傾向にある。
- 実験室における単一人の回数間距離の課題間の違いの評価
  BCCWJ-SUMM_L(T) ⇔ GROSS_L(T) ⇔ RETELLING_I(T) ⇔ RETELLING_K(T) ⇔ RETELLING_M(T)
  - BCCWJ-SUMM_L(T) ⇔ GROSS_L(T)
    文字単位の評価の場合 n-gram(2,3,4)_char, Kendall_char に有意差が見られた。
    形態素単位の評価の場合 n-gram(2,3,4,≤2,≤3,≤4)_mrph, Footrule_mrph, Kendall_mrph に有意差が見られた。
  - BCCWJ-SUMM_L(T) ⇔ RETELLING_{I,K,M}(T)
    全てのスコアについて、有意差が見られた。
  - GROSS_L(T) ⇔ RETELLING_{I,K,M}(T)
    全てのスコアについて、有意差が見られた。
  - RETELLING_I(T) ⇔ RETELLING_K(T)
    文字単位の評価の場合 n-gram(1,4,≤2)_char, p-mer(2,≤2)_char に有意差が見られた。
    形態素単位の評価の場合、全てのスコアに有意差が見られた。
  - RETELLING_I(T) ⇔ RETELLING_M(T)
    Kendall_char **以外**について有意差が見られた。
  - RETELLING_I(T) ⇔ RETELLING_M(T)
    文字単位の評価の場合 n-gram(2,≤2,≤3,≤4)_char, p-mer(2,3,4,≤2,≤3,≤4)_char に有意差が見られた。
    形態素単位の評価の場合、n-gram(1,2,≤2,≤3,≤4)_mrph, p-mer(2,3,4,≤2,≤3,≤4)_mrph に有意差が見られた。

  複数人間の評価ではなく、複数回間の評価でも、前項と同じ傾向が見られる。
  再話課題の間については、形態素単位の評価においては、三課題のうちどの二つ組においても有意差が出る傾向にある。口述による再話 (RETELLING_{I,K}) の方が筆述による再話 (RETELLING_M) より一致度が高くなる。また口述による再話においては、自身の体験に基づく再話 (RETELLING_I) の方が、他者から聞いた話の再話 (RETELLING_K) よりも一致度が高くなることが認められた。
- クラウドソーシングにおける課題間の違いの評価
  BCCWJ-SUMM_C ⇔ GROSS_C について、全てのスコアについて、有意差が見られた。
  クラウドソーシングにおける課題間の違いについて

*4 コルモゴロフ=スミルノフ検定 (0.05 未満で 2 群は異なる分布から取り出されたことを示す) も行ったが、ほぼ同等の結果が得られたために省略する。

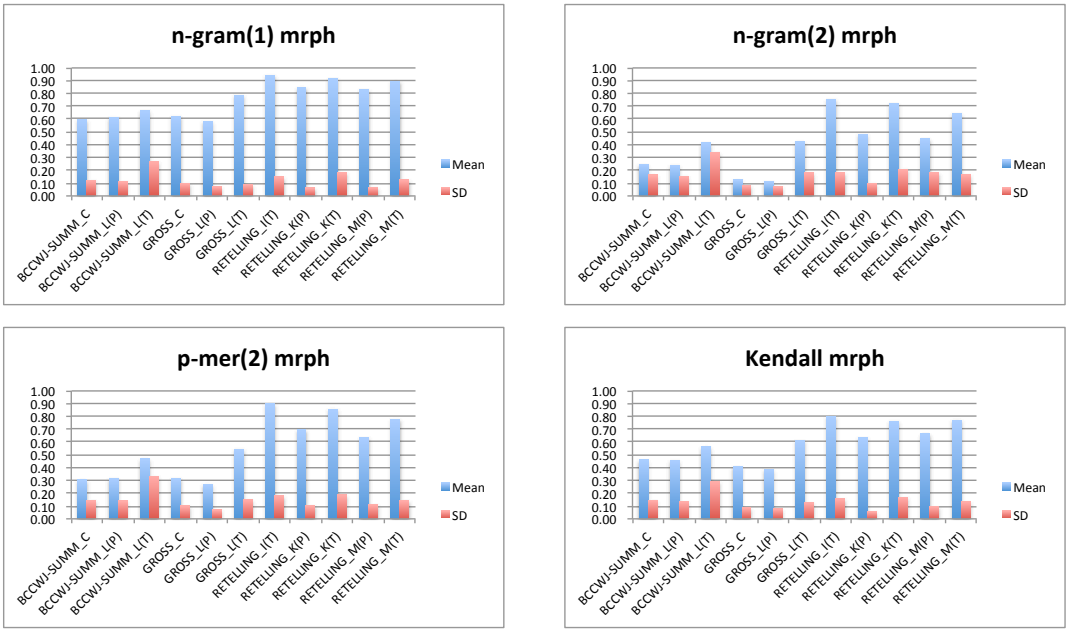


**図 3** 課題とスコア (n-gram(1),n-gram(2),p-mer(2),Kendall: 形態素単位)

も、前項と同じ傾向が見られる。

- 要約課題においてクラウドソーシングと実験室との違いを評価する (複数人間)
  BCCWJ-SUMM_C ⇔ BCCWJ-SUMM_L(P) について、n-gram(2)_char, n-gram(3)_char, n-gram(4)_char にのみ有意差が見られた。
  これは、タイプ入力 (BCCWJ-SUMM_C) と筆述 (BCCWJ-SUMM_L(P)) とで、表記ゆれの統制の差がでたのではないかと考える。
- 語釈課題においてクラウドソーシングと実験室との違いを評価する (複数人間)
  GROSS_C ⇔ GROSS_L(P) について、n-gram(2,3,4)_char, n-gram(2,3,4)_mrph, Footrule_mrph, Kendall_mrph **以外**について有意差が見られた。
  語釈においては、クラウドソーシングの場合 wikipedia や辞書サイトからのコピーが行われる傾向にある一方、実験室の場合は特にリファレンスもなく筆述で行うために差が出たのではないかと考える。
- 複数人間距離と単一人の回数間距離の違い
  BCCWJ-SUMM_L(P) ⇔ BCCWJ-SUMM_L(T), GROSS_L(P) ⇔ GROSS_L(T), RETELLING_K(P) ⇔ RETELLING_K(T), RETELLING_M(P) ⇔ RETELLING_M(T) について、全てのスコアについて有意差が見られた。
  基本的に単一人が実施したほうが一致度が高いと考えられるが、統計分析の結果からもそれが確認できる。

### 2.7.3 スコア毎の特性

前節の課題間の議論から考えられるスコア毎の特性について論じる。

- 文字 n-gram はタイプ入力と筆述入力の差として認められることから、表記ゆれレベルで一致度が下がる特性があると考える。
- 形態素 n-gram は再話と繰り返しで顕著に高くなることから、個々人の言い回しや文体などを反映していると考える。
- p-mer, Footrule, Kendall などは語順などを反映していると考えるが、情報の提示順が重要な要約・再話で一致度が高い一方、語釈などにおいては低い傾向にあることがわかった。
- n-gram, p-mer ともに $n, p$ の値が高くなるにつれてスコアが低くなる。このために有意差が出にくくなる傾向にある。
- n-gram, p-mer ともに $n$ (or $p$) 以下のスコアとして設定した場合に、より低い $n$ (or $p$) の方が一致が多くなる傾向にあるために、より高い $n$ (or $p$) の差異が見られなくなる傾向がある。これはスコアの自然な解釈であると考えられるが、何らかの用途で長い n-gram, p-mer を重要視する場合には加重を行う必要があるだろう。
- n-gram(1)_* と Kendall_* と比較した場合、n-gram(1)_*では有意差が出るが、順序尺度を入れた Kendall_* では有意差が出ないスコアの組み合わせがいくつかあった。これは文字順・語順の一致度が低い場合に、順序尺度を掛けあわせたがために全体の一致度の差がなくなったことが考えられる。

### 2.8 自動評価指標の特性のまとめ

本節では、まず自動要約・機械翻訳で用いられている評価指標の数理的構造を説明した。評価指標がどのカーネル・距離・相関係数と対応しているのかを説明し、n-gram 系、p-mer 系、順序尺度の三つに抽象化した。次に様々な言語資源を用いて各指標で用いられているスコアの特性を明らかにした。要約・語釈・再話からなる 7 種類の言語資源を用いて、課題・多人数産出・複数回産出・産出手段 (口述・筆述・タイプ) の軸を用いて、どのような分散が観察されるかを確認した。

逆の観点からいうと、これらの評価指標を用いて、整備している被験者実験に基づく要約データを評価していることになる。

しかしながら、スコアが捉える言語の特性については明らかにしたが、本来自動要約に必要な内容評価と読みやすさの観点については何も言っていないに等しい。3 節では内容評価の評価方法について示し、4 節では読みやすさの評価方法について示す。

## 3. 情報構造を用いた要約文の評価に向けて

本節では要約の評価における内容評価に関して、各課題における情報の有用性の観点からではなく、言語の談話構造の観点からの評価方法について検討する。具体的には情報構造 (Information Structure)[21] に基づいて、情報の新旧 (情報状態: Information Status)、主題 (Topic) や焦点 (Focus) などをコーパスにアノテーションし、作成された要約文が言語学的に分析された情報構造のどの部分を抽出しているかなどを検討することを試みる。

情報構造は、文法的には構成素の左方移動 (もしくは右方移動) などにより表出するほか、特に日本語はとりたて詞などの存在により形態論的に明示的に表出する場合もある。

本節では、多言語に適用されている Götze[2] の情報構造アノテーションと関連研究を紹介し、BCCWJ-SUMM の元文書に対するアノテーションの試みについて報告する。

### 3.1 情報構造アノテーションの先行研究

Götze[2] は言語非依存で特定の言語理論によらない信頼性のあるアノテーションを行うためにアノテーションガイドラインを策定した。ガイドラインは、コアアノテーションスキーム、拡張アノテーションスキームからなる表 5 に Götze の情報構造タグ (コアアノテーションスキーム) を示す。

情報状態のアノテーションにおいては、談話要素 (discourse referents) の談話中の情報状態をアノテーションすることを目的とする。談話要素は個体、場所、時間、事象、状況などの様々なタイプのエンティティにより構成され、何らかの照応表現により参照される。

**表 5** Götze の情報構造のタグ (コアアノテーションスキーム) [2]

| Layers | Tags | Description |
|---|---|---|
| 情報状態 | giv | Given(旧情報) |
| (Information Status) | acc | Accessible(補完可能) |
| | new | New(新情報) |
| | cat | cataphor (後方参照) |
| | nil | non-referential (指示対象ではない) |
| 主題 | ab | Aboutness topic |
| (Topic) | fs | Frame setting topic |
| 焦点 | nf | New Information Focus |
| (Focus) | cf | Contrastive Focus (対比的焦点) |

情報状態 (information status) は先行詞もしくは参照すべき実体を認定する困難さを表す認定可能性 (retrievability) を規定する。“giv(en)” は先行文脈に明示的に規定されているもの、“acc(essible)” は先行文脈に明示的に規定されてはいないが言語生産者と言語受容者の間で共有される世界知識などにより推論によって規定できるものを表す。“new” は先行文脈によって明示的に規定されておらず、推論によっても参照すべき実体が仮定できないものを表す。

この情報状態の分類は、Prince [22] の情報状態の分類を元にしている。Prince は、情報の新旧を テキストの談話構造の状況に基づく { 既出 (discourse-old)・未出 (discourse-new)} と受容者の状況に基づく { 既知 (hearer-old)・未知 (hearer-new)} に分割して、四つのタイプに分類した (表 6)。

Prince [22] の分類では、談話中の状態と生産者が受容者側に仮定する知識の観点から、“giv”=(既出, 既知)、“acc”=(未出, 既知)、“new”=(未出, 未知) の三つに分けられる。なお、 Prince は、(既出, 未知) にあたる表現は、成立している談話中に出現しないとしている。

主題 (topic) は言語受容者側で既知のもので、文もしくは節によって説明される中心的な対象に対してアノテーションする。Jacobs[23] はアバウトネス主題 (aboutness topic) とフレームセット主題 (frame setting topic) の 2 種類の違いについて論じている。前者は文が何について論じているか (“what the sentence is about”)、後者は文の中に内在するフレーム (“the frame within which the sentence holds”) としており、フレームは以下の通り定義している:

> Frame-setting ([23], p .656) (X,Y) において、X がY のフレームである ⇔ Y によって表現される命題が制限される可能な現実世界のドメインを、X が明確に指定する

焦点 (focus) は言語受容者側で未知のもので、言語生産者側が新情報を伝える要素を指す。焦点のうち他の談話要素と対比的に述べられているものを対比的焦点 (contrast focus) と呼ぶ。

この Götze[2] のスキームにより他言語においてアノテーションが進められている。Cook ら [24] はドイツ語の新聞

表 6 Prince の情報状態の分類

| 情報状態 | Prince の分類 | 談話構造 | 受容者 | 摘要 |
|---|---|---|---|---|
| giv(旧情報) | evoked | 既出 | 既知 | 生産者が「受容者が既知である」と仮定し、先行談話に出現しているもの |
| acc(認定可能) | unused | 未出 | 既知 | 生産者が「受容者が既知である」と仮定し、先行談話に出現していないもの |
| - | 存在しない | 既出 | 未知 | 生産者が「受容者が未知である」と仮定し、先行談話に出現しているもの |
| new(新情報) | brand-new | 未出 | 未知 | 生産者が「受容者が未知である」と仮定し、先行談話に出現していないもの |

記事 588 文について情報構造のうちの “aboutness topics” についてアノテーションを試行的に行い、文のタイプによって (Fleiss' $\kappa$)0.19 と 0.57 のアノテーション一致度を確認した。

### 3.2 日本語に対する情報状態アノテーションのスキーマ

以下では、情報構造アノテーションの出発点として、新聞記事に対する情報状態アノテーションについて現在までに検討したアノテーション単位とアノテーションタグ集合について示す。

#### 3.2.1 アノテーション単位

今回は、文書中の各文の主節の名詞句に対して、情報状態を付与することを目標に行った。主節の名詞句とは、主語・補語・連用修飾語などである。以下の例では、主節のガ格 NP、デ格 NP を付与対象とする。

| 連文節 | 情報状態 |
|---|---|
| 地方自治体が運営する公営地下鉄二十六路線のうち二〇〇〇年度決算で経常損益が黒字なのは、札幌市南北線など四路線にとどまったことが、 | new |
| 公営交通事業協会が十日まとめた報告書で | new |
| 分かった。 | - |

述語は付与対象としないが、名詞述語は補語名詞句を含むので付与対象とする。名詞句を修飾する語句 (連体修飾語) は名詞句の一部と見なし、付与対象としない。次の例では、主節のガ格名詞句と述語名詞句を付与対象とする。連体修飾語「東京都大江戸線の」は述語名詞句の一部なので付与対象としない。

| 連文節 | 情報状態 |
|---|---|
| 赤字額が最も多いのは | acc-inf |
| 東京都大江戸線の三百十一億円だった。 | new |

主節述語にかかる連用修飾節は付与対象としない。次の例では、「〜赤字で」は連用修飾節なので付与対象としない。

| 連文節 | 情報状態 |
|---|---|
| 全体の経常損益は千六百七十二億円の赤字で、 | - |
| 累積欠損金は | acc-inf |
| 二兆三千四百五十四億円に | new |
| 上っている。 | - |

#### 3.2.2 アノテーションタグ集合

Götze[2] の情報状態アノテーションの拡張アノテーションスキームに基づいてタグ集合を規定した。表 7 に一覧を示す。

以下、日本語向けに解釈したタグについて説明する。

- giv-active:
  直前に**明示的に**言及されている対象にのみ用いる。日本語では直前に言及されている要素は代名詞などで繰り返さず省略することが多いので、このタグはあまり使わない可能性がある。
- giv-inactive[*5]:
  二つ前の文に明示的に言及されている対象に用いる。
- acc-sit:
  目の前にある事物などに言及する場合に用いる。(例：「砂糖 取って」など。)
  書き言葉の場合、書き手や読み手に対する外界照応などがこれにあたる。
- acc-aggr:
  [2] で挙げられているのは次のような例である：
  – Peter went shopping with Maria. **They** bought many flowers.
  acc-inf の set-rel(集合関係) との区別を行う必要がある。いくつかの先行詞をまとめて複数形代名詞で参照するような、事実上 giv の亜種である場合に限定して用いて、それ以外の場合は acc-inf にする。
- acc-inf:
  [2] で挙げられているのは次のような例である：
  – part-whole: The garden beautiful. **Its entrance** is just across this river.
  – set-rel: The flowers in the garden blossom. **The flowers near the gate** blossom violet.
  – set-rel: The children swam in the lake. **The family** experienced a beautiful day.
  – entity-attribute: The flowers enchanted Peter. **Their scent** was wonderful.
  全体-部分、集合-要素、上位集合-下位集合、同一集合に属する要素、実体-属性、所有者-所有物など具体的な関係を決めておいて、その関係に該当する referent が先行文脈中に明示的に出てきている (given) かで判

[*5] 【用語】inactive は discourse-new + hearer-new の意味で使う場合が多く、semi-active とか textually accessible などというほうが一般的。

**表 7** 日本語に対する情報状態アノテーション

| タグ (Coarse) | タグ (Fine) | Description |
|---|---|---|
| giv | (underspecified) | given: 旧情報 |
| | giv-active | active: 同一文、一つ前の文に談話要素がある |
| | giv-inactive | inactive: 二つ前の文以前に 談話要素 がある |
| acc | (underspecified) | accessible: 談話要素 はないが認定可能なもの |
| | acc-sit | situationally accessiblly accessible: 談話の状況から認定可能 |
| | acc-aggr | aggregation: 集約したものが 談話要素 になるもの |
| | acc-inf | inferable: 推論により認定可能 (part-whole, set-rel(subset/superset),entity-attribute など) |
| | acc-gen | general: 世界知識で補完可能なもの |
| new | | new: 新情報 |
| nil | | non-referential: 談話要素ではない |
| (cat) | | cataphor: 後方照応 |

定する。これらは、聞き手が既出談話要素と関連するものだと認識できるかどうかの問題なので、判断に揺れが生じるものと考えられる。

- acc-gen:
  「ライオン」(Type) や「ペレ (サッカー選手)」(Token) など、誰もが共通知識として知っているものを acc-gen とする。
  日本語の場合、聞き手にとって未知の対象は「ライオンという動物」「ペレという人」のように「という」を使うので、名前で直接言及している場合は acc-gen である可能性が高い。
  「ライオンが歩いてきた」(不定名詞) のように、既知のクラスではなくインスタンスを表す場合、本稿では new とする。
  「若いライオン」(限定修飾) のように、既知のクラスではなくその下位クラスを表す場合、本稿では new とする。
- new:
  聞き手にとって未知で、談話に新たに導入されるもの。

なお、談話要素にあたらないものに対しては nil を付与する。cat は後方照応 (catphora)[*6] のためのタグだが、あまり出現しない現象のため作業者間でタグの存在を共有するにとどめた。

### 3.3 評価

今回は情報状態のアノテーションの可能性について検討を進めるために、MAMA サイクル [25] に基づくアノテーションを行った。最初に 1 文書 [2] に基づいてアノテーションを行い、その後 3 文書に対するアノテーションを 3 回繰り返した。

1 回目については、完全に独立して行うのではなく、議論しながらアノテーションを進めた。2 回目については、主節に対する係り受け関係について、{ 主語、修飾 (連用)、修飾 (連体)、補語、接続 } などの情報を明示したうえでアノテーションを行った。3 回目については、3 人の作業者の 2 回目以前のアノテーションを対比したものを確認したうえでアノテーションを行った。4 回目については、図 4 のようなフローチャートを作成したうえでアノテーションを行った。

- 談話に既出 (discourse-old)
  - 直接照応 (同一文または直前の文) “giv-active”
  - 直接照応 (二つ以上前の文) “giv-inactive”
    厳密に言えば直前の文でゼロ照応している場合は “giv-active” にすべき
  - 集合化 (太郎と次郎がいた。彼ら は〜) “acc-aggr”
- 談話に未出 (discourse-new)
  - 現場指示 “acc-sit”
  - 推測可能 “acc-inf”
    談話に明示的に既出してはしていないが、それに準じて考えられるような、活性化 (activeness) 状態の高い要素
    既出要素と何らかの意味関係があるもの
    * 全体-部分関係
    * 集合関係 (上位集合、下位集合、姉妹要素)
    * 実体と属性

    間接照応や意味関係があっても、活性化状態が高い場合と低い場合がありうるが、判別困難なので意味関係だけで判定
  - 既知 (hearer-old) (≃ 記憶指示) “acc-gen”
    活性化状態が高いとはいえないが、何であるか既知であるもの
    総称名詞 (猫)、有名な固有名詞 (シャーロック・ホームズ)
  - 未知 (hearer-new) “new”
    既知とは言えないが、何であるかは分かるもの
    * 限定修飾された総称名詞 (「3 年間熟成させた納豆」)
    * 有名でない固有名詞 (札幌・南北線)、数詞、日付
    * 不定名詞 (おじいさん と おばあさん がいました)
    * 「〜という人」など (桃太郎という男の子 がいました)

    何であるかわからないもの
    * 「〜というの」

**図 4** 情報状態判定フローチャート

表 8 に情報状態アノテーションの試行の一致度を示す。アノテーションは最も粒度が細かい拡張アノテーションスキーム (タグ (Fine)) で実施し、評価は拡張アノテーションスキーム (タグ Fine) とコアアノテーションスキーム (タ

[*6] 例：「これ は 2 年前の話である。」

**表 9** 情報状態アノテーションの試行 (4 回目作業者 B-C 間)

| B-C | a-g | a-i | g-a | g-i | new | nil | 計 |
|---|---|---|---|---|---|---|---|
| acc-gen | 3 | | | | | | 3 |
| acc-inf | 1 | 3 | | | | 2 | 6 |
| giv-active | | | | 1 | | | 1 |
| giv-inactive | | 2 | | 4 | | | 6 |
| new | 2 | 2 | 1 | 2 | 10 | | 17 |
| nil | | 4 | | 1 | 6 | 20 | 17 |
| 計 | 6 | 11 | 1 | 8 | 16 | 22 | 64 |

グ Coarse) の 2 レベルで行った。表中 A, B, C はそれぞれ作業者を表し、2 作業者間の (Cohen's)$\kappa$ 値と 3 作業者間の (Fleiss's)$\kappa$ 値を示す。1 回目においては B-C 間で議論しながら実施したために一致度が高い傾向にある。4 回目において作業指針であるフローチャートを確立したが、それでも 0.50 前後の $\kappa$ 値であった。

表 9 に 4 回目の作業者 B-C 間の分割表を示す。一致に関して、何を談話要素とするか (nil の認定) という問題と、何を認定可能とするか (acc の認定) という問題の二つがあると考える。前者の問題については、アノテーション単位を考えなおすことで対処したいと考える。後者の問題については、作業者間の世界知識は一致しないことから、基本的には揺れを許容したうえでアノテーションをつづけたい。今後、多人数によるアノテーションを行うことで、揺れも含めた評価を進める。

### 3.4 今後の課題

今回は主節にかかる要素のみについて情報状態を認定しており、従属節内の要素については検討を行わなかった。また要素の単位については文節より大きな単位 (係り元を根とする部分木全体) を対象としている。

旧情報 ("giv") は照応関係、認定可能 ("acc") は外界照応と関連が深く、照応関係アノテーションの精緻化とも言える。現在 BCCWJ に対して、NAIST テキストコーパス [26] 互換の述語項構造と照応関係のアノテーションを進めており、2014 年度末に完成する予定である。これらの述語項構造・照応関係アノテーションを元に、今回試行的に行って得られた知見に基づき、述語-項単位に情報構造のアノテーションを進めていきたい。その際に、acc-aggr のように集合化による照応など、タグの再設計についても検討し、日本語のとりたて詞のふるまい [27] など日本語記述文法の研究の知見を取り入れながら主題・焦点についても付与することを検討する。

述語項構造アノテーションの作業においては、国立国語研究所基幹型共同研究プロジェクト「コーパスアノテーションの基礎研究」の枠組を用いて、東工大・奈良先端大・外注業者・国語研の各所で作業者の育成を行った。育成した多数の作業者による情報構造アノテーションの揺れを評価することを行いたい。

さらに、被験者実験により得られた要約が、それぞれどのタグが付与されている部分が選択される傾向にあるかを分析することで、情報状態に基づいた単一文書要約の内容評価方法を確立したい。

## 4. 読み時間評価を用いた読みやすさの評価に向けて

自動要約結果の評価軸として、内容評価の観点以外に読みやすさの観点の評価が必要となる。本稿では、現象論的なリーダビリティ評価ではなく、テキストのベクトル空間上の距離に基づく定量的な評価でもなく、被験者実験に基づく定量的な評価を提案する。

心理言語学の分野では、人間の文処理過程を解明するために、理論や仮説の構築だけでなく、実験に基づいて仮説を検証することが行われ、さまざまな実験手法が提案されている。本稿では、テキストの読み時間をアノテーションする方法を検討する。具体的には、心理言語学における様々な実験手法のうち次の 2 種類の実験を行う。一つは自己ペース読文法による読文速度の計測である。やや人工的な読み方だが、普通の計算機のみで読文速度を取得することが可能である。もう一つは視線操作法 (視線走査装置による読文速度の計測) である。高価な機器を必要とするが、読文速度を直接的に取得することが可能である。

以下、読み時間の取得指針について説明する。まず、テキストに対する読み時間付与に関する先行研究について示す。次に自己ペース読文課題法、視線走査装置による読文速度計測について説明する。次に被験者から得る被験者の特性について示し、最後に読み時間の構造化手法について示す。

### 4.1 先行研究

数少ない先行研究して Kennedy ら [29] の Dundee Eye-Tracking Corpus がある。対象言語を英語とフランス語、それぞれ 10 人の母語話者を被験者とし視線走査装置を利用して 20 ファイルの新聞社説に対する視線走査情報を記録している。各ファイルは 5 行毎からなる 40 画面により構成されており、研究用途に一次情報が公開されている。Dundee Eye-Tracking Corpus は特定の言語現象分析を目的とせずに作成されたコーパスに基づいており、心理言語学におけるさまざまな仮説の客観的な検証に用いられている。

例えば、Demberg ら [30] は、Gibson による Dependency Locality Theory(DLT) の integration cost[31] と Hale による suprisal[32] とを Dundee Corpus 上の読文時間を用いて検証を行った。公開され共有されるデータを用いることにより先行研究の結果を再検証するとりくみも行われている。例えば、Roland らは [33]、Demberg ら [34] の Dundee

**表 8** 情報状態アノテーションの試行

| 試行 | 評価単位 | 文書数 | 対象数 | 一致度 (Cohen's $\kappa$) | | | 一致度 |
|---|---|---|---|---|---|---|---|
| | | | | (A-B) | (B-C) | (C-A) | Fleiss's $\kappa$ |
| 1 回目 | (Fine) | 1 | 54 | 0.54 | 0.67 | 0.40 | 0.53 |
| | (Coarse) | | | 0.53 | 0.71 | 0.38 | 0.54 |
| 2 回目 | (Fine) | 3 | 77 + 69 + 37 = 183 | 0.43 | 0.41 | 0.42 | 0.42 |
| | (Coarse) | | | 0.54 | 0.44 | 0.47 | 0.48 |
| 3 回目 | (Fine) | 3 | 37 + 51 + 42 = 130 | 0.38 | 0.40 | 0.36 | 0.37 |
| | (Coarse) | | | 0.41 | 0.49 | 0.43 | 0.44 |
| 4 回目 | (Fine) | 3 | 16 + 33 + 15 | 0.50 | 0.49 | 0.49 | 0.49 |
| | (Coarse) | | | 0.49 | 0.52 | 0.51 | 0.50 |

Corpus の分析が、限られたデータポイントに基づいて歪められていることを再検証により証明している。

### 4.2 読み時間の評価方法

#### 4.2.1 移動窓方式自己ペース読文法

自己ペース読文課題は、キーボード入力などに基づき逐次的に文字列を表示し、被験者のペースで文を読ませる課題である。英語においては視線走査情報と高い相関があることが知られており [35]、安価な機器で読文速度を取得することが可能である。刺激の呈示方法として最初に文字数がわかる移動窓方式を用いる。ソフトウェアとして Linger [*7] を用いる。テキストは横書き、等幅フォントを用い、5 行単位で呈示する。呈示する単位として、国語研文節単位を用いる。きちんと文を読んでいるかを確認するために、1 文書毎に内容を問う Yes/No Question に回答させる。

#### 4.2.2 視線操作法

視線走査装置は、被験者がディスプレイ画面上のどの文字を注視しているのかを取得することができる機材である。この視線走査装置を用いて、視線停留箇所と停留時間を計測することにより、読文速度を取得することができる。視線走査装置として、SRResearch 社の EyeLinkCL シリーズを用いる。テキストは横書き、等幅フォントを用い、5 行単位で呈示する。自己ペース読文法と同様に、1 文書毎に内容を問う Yes/No Question に回答させる。

図 5 に視線走査実験結果を示す。呈示する各文字の 1/2 幅毎に interest area (図中黄色の grid で表示) と呼ばれる領域を設定する。各 interest area 毎に視線停留箇所と停留時間、サッケード眼球運動の通過などが付与される。この interest area が設定されている半文字単位の情報に BCCWJ に付与されている短単位形態論情報、長単位形態論情報、文節境界情報を重ね合わせることにより、それぞれの単位での分析を行う。この実験法で得られるデータは読み戻しができ、かつ周辺視野により隣接する形態素・文節が読まれることもあり、全ての文節が必ず一度は読まれるわけではない。この読み時間の扱いについては 4.4 で議論する。

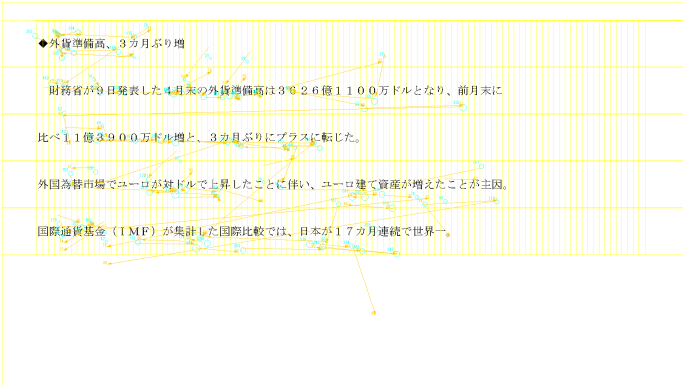


**図 5** 視線走査実験結果

### 4.3 被験者の特性の取得

読み時間を評価するうえで被験者の特性を得ることが重要である。被験者の特性の取得のために、個人情報アンケート・語彙数テスト・記憶力テストを実施する。

個人情報アンケートでは被験者の以下の情報を得る。個人情報の提供は任意であるため、情報が欠損している場合もある。

- 年齢: 5 歳刻み
- 性別
- 最終学歴: { 中学, 高等学校, 短大・専門学校, 大学, 大学院 } の 5 択とともに専門分野について自由記述
- 言語形成地: 本人の 15 歳までに住んだ地域を都道府県単位で編年体で自由記述。父親と母親の出身地も記述する。

次に、被験者の語彙知識が読み時間に与える影響を分析するために、語彙数テストを行う。語彙数テストは「日本語の語彙特性」[15] に含まれる『百羅漢』[*8][37] の結果を記録する。

さらに、被験者の記憶力が読み時間に与える影響を分析するために、記憶力テストを行う。記憶力テストは「リーディングスパンテスト」[38] を行い結果を記録する。

[*7] http://tedlab.mit.edu/~dr/Linger/

[*8] http://www.kecl.ntt.co.jp/icl/lirg/resources/goitokusei/goi-test.html

これらの被験者の特性に基づき、被験者毎の文書理解の差を捉えることを目標とする。

### 4.4 読み時間の扱い

自己ペース読文法で得られるデータは読み戻しも読み飛ばしもできないため、全ての文節が必ず 1 回読まれることになる。得られた結果は元テキストの文節に対して、文節順に 1 対 1 対応した読み時間情報が得られる。

一方、視線走査実験の場合には、被験者は自由に読み戻し・読み飛ばしが可能である。元文書の語順に沿って分析するために次の指標が用いられる。

- First pass time
  最初に「分析単位」に視線が停留してから、他の「分析単位」に出るまでの間の視線停留時間の合計
- Total time
  「分析単位」内の視線停留時間の合計
- Regression path time
  最初に「分析単位」に視線が停留してから、より右側 (もしくは下側) の「分析単位」に出るまでの間の視線停留時間の合計 (左側 (もしくは上側) へ停留している停留時間は累計される)

付録の表 A·3 に視線停留順の視線走査実験結果例を示す。表 A·4,A·5, に文書出現順に集計した視線走査実験結果例 (文節単位集計・単語単位集計) を示す。

### 4.5 今後の課題

2014 年 12 月 10 日現在、要約元文書に対してのべ 98 人分の自己ペース読文法による実験結果とのべ 43 人分の視線走査実験結果を収集している。本来、自動要約の読みやすさの評価としては機械生成された要約文を用いて被験者実験を行わなければならない。これについては自動要約器を作成している他機関の研究者からデータをいただいて、上に示した実験を行い分析を進めたいと考えている。

現在のところは読み時間の収集と集計方法の確立にとどまっているが、要約元文書に対する読み時間評価は、利用者毎の自動要約に向けたものであり、一部の被験者は要約元文書に対する読み時間を評価したあとに要約文作成を依頼している。5 節では、利用者毎の自動要約をどのようにして実現するのかについての構想について示す。

## 5. 読み時間を用いた利用者毎の自動要約に向けて

2 節では、既存の評価指標を用いた要約文の評価を行い、被験者間だけでなく、1 人の被験者の複数回間でも要約文作成が揺れることを示した。3 節では、要約文の内容評価のために情報状態アノテーションを提案し、そのなかで accessible タグにおいては被験者間で揺れを許すことを論じた。4 節では、読み時間の評価方法について論じた。

以下では、まず本稿の着想に至った情報構造と読み時間の関係に示す。文の読み時間については、ぎなた読み・ガーデンパス・多義性などさまざまである。形態論的・統語意味論的曖昧性とは別に情報の新旧についても読み時間に影響があることが報告されている [39]。次に読み時間を手がかりとした利用者毎の自動要約の可能性について論じる。

### 5.1 情報構造と読み時間の関係

#### 5.1.1 関係節と情報構造と読み時間

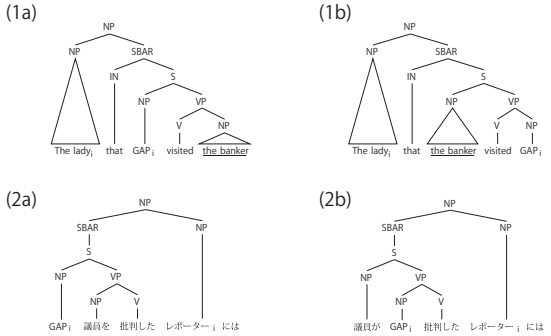


**図 6** 主語関係節と目的語関係節

心理言語学では構文毎の読み時間の差を手がかりとして、人の文処理機構を研究する。その中で英語において目的語関係節 (1b) が主語関係節 (1a) より処理が難しいことが報告されている [40]。

(1) a. 主語関係節
The lady$_i$ [ that GAP$_i$ visited the banker ] ...

b. 目的語関係節
The lady$_i$ [ that the banker visited GAP$_i$ ] ...

これを説明するために構造的距離仮説 (構文上の距離)[41][42] と線形的距離仮説 (文字列上の距離) [31] により説明されるが、英語の場合どちらも目的語関係節の方が遠い (図 6 上) ためにどちらかをサポートすることはない。

日本語の場合はこの仮説に差があり、構造的距離では主語関係節の方が目的語関係節より埋語と空所の距離が近く、線形的距離では目的語関係節の方が主語関係節より埋語と空所の距離が近くなる (図 6 下)。

(2) a. 主語関係節
GAP$_i$ 議員を 批判した レポーター$_i$ には ...

b. 目的語関係節
議員が GAP$_i$ 批判した レポーター$_i$ には ...

読み時間の評価の結果日本語でも目的語関係節の方が負荷が高いことが報告されている。[43]。

Roland ら [44] は談話機能仮説として関係節内の名詞が

新情報か旧情報かで読み時間が変わることを報告している。コーパス調査において、(1a) のような主語関係節の場合、新情報が 57%・旧情報 43% であるのに対し、(1b) のような目的語関係節の場合、新情報が 2%・旧情報 98% であった。目的語関係節は旧情報文脈で利用される傾向がある。

彼らの被験者実験において、(4a) のような旧情報文脈をトピック条件として与えた場合と、(4b) のような新情報文脈を中立条件として与えた場合、中立条件だと目的語関係節の方が主語関係節より読み時間がかかるが、トピック条件においては読み時間に差がないことが報告されている。このことから関係節内の名詞が旧情報であれば目的語関係節の負荷が軽減されることがわかる (例文は Roland [44] の abstract より)。

(3) a. トピック条件 (旧情報文脈)
The banker was very friendly.

b. 中立条件 (新情報文脈)
There was a dinner party on Saturday night.

(4) a. 主語関係節
The lady$_i$ [ that GAP$_i$ visited the banker ] enjoyed the dinner very much.

b. 目的語関係節
The lady$_i$ [ that the banker visited the GAP$_i$ ] enjoyed the dinner

日本語においても同様の研究が佐藤 [45](第 4 章) により実施されている。

BCCWJ によるコーパス調査においては、前文脈 400 字以内に談話要素があるかどうかを分析したところ、主語関係節の場合は新情報 70%・旧情報 30% であり、目的語関係節の場合は新情報 20% ・旧情報 80% であった。英語ほど明確な違いはなかった。また以下の例文 (5) を用いた被験者実験による読み時間評価においては英語と同様の現象が確認できなかったことが報告されている (例文は佐藤 [45](p.63) より)。

(5)1-a. トピック文脈 (1 文目)
特捜部の 刑事が 事件の 捜査に あたった。

1-b. 中立文脈 (1 文目)
陰惨な 殺人事件の 現場で 捜査が 行われた。

2-a. 主語関係節 (2 文目)
(その／特捜部の) 刑事を 呼び止めた 担当者は 手短に 現場を 案内した。

2-b. 目的語関係節 (2 文目)
(その／特捜部の) 刑事が 呼び止めた 担当者は 手短に 現場を 案内した。

3. ラップアップ文 (3 文目)
その後 刑事は 無事に 犯人を 逮捕した。

残念ながら日本語においては読み時間が関係節内の新旧と相関があることは明確には示されていない。

#### 5.1.2 分裂文と情報構造と読み時間

英語においては、関係節と同様に、目的語分裂文が主語分裂文より処理が難しいことが報告されている [46][47][48]。Gordon ら [46] は以下の分裂文中の動詞の読み時間評価において、主語分裂文に比べて目的語分裂文の処理に時間がかかることを示している (例文 (6) は Gordon ら [46] より [*9])。

(6) a. 主語分裂文
It was the barber/John that saw the lawyer/Bill in the parking lot.

b. 目的語分裂文
It was the barber/John that the lawyer/Bill saw in the parking lot.

これに対し、カフラマン [49][50][51] は日本語分裂文の処理負荷についての研究を行っている。

カフラマン [50](第 5 章) は日本語主語分裂文・目的語分裂文間の読み時間の違いの評価を行い、分裂文動詞出現時に目的語分裂条件 (7a) の方が主語分裂条件 (7b) よりも早く読まれることが示されている (例文はカフラマン [50](第 5 章) より)。

(7) a. 主語分裂条件
去年 祖母を 田舎で 介抱したのは 遠い 親戚だと 母が 言った。

b. 目的語分裂条件
去年 祖母が 田舎で 介抱したのは 遠い 親戚だと 母が 言った。

カフラマン [49] は 300 万語からなるコーパスから「のは」を含む 2085 文を抽出し、1756 文 (84%) が分裂文、329 文 (16%) が非分裂文であった。このうち 656 文にタグ付けをしたところ、主語分裂文 (6a) が 31%、目的語分裂文 (6b) が 8%、他のタイプの分裂文 (6c) が 45%であった (例文はカフラマン [49] より)。

(8) a. 主語分裂文
GAP$_i$ 敵を 倒したのは サラマンカ $_i$ だった

b. 目的語分裂文
俺が GAP$_i$ 相手にしていたのは 犯罪者 $_i$ だった

[*9] Gordon らは名詞句がクラス相当の普通名詞 (description: “the barber”, “the lawyer”) である場合とインスタンス相当の固有名 (name: “John”, “Bill”) である場合の違いにも言及している。

c. 分裂文 (その他)
GAP$_i$ 改札口を 出たのは 9 時 15 分 $_i$ だった

d. 非分裂文
体力が あったのは いうまでもない

カフラマンは他動詞のみについて限定して、「のは」を含む文型の分裂文生成確率を求めている。分裂文生成確率とは「名詞句-を/が + 他動詞 のは」の文型のうち、主語/目的語分裂文であった割合を意味する。表 10 に分裂文生成確率を示す。

**表 10** 分裂文の分布 (カフラマン [49] より)

| 文型 | 分裂文構造 | 分裂文生成確率 |
|---|---|---|
| 名詞句-を + 他動詞 のは | 主語分裂文 | 0.57 (205/357) |
| 名詞句-が + 他動詞 のは | 目的語分裂文 | 0.75 (86/114) |

カッコ内、左の数値は分裂文の数、右の数値は当該文型の総数。

このことから、「のは」の文型において主語分裂文が多いが、他動詞に限定して分析すると分裂文生成確率としては目的語分裂文の方が確率が高いことがわかる。

その後、矢野ら [52] はこのコーパス中の構造的な偏りが分裂文の処理に影響を及ぼしているとし、以下のような文脈を統制した実験を行った (例文は矢野ら [52] より)。

(9) 0. 【“竹内さん” と “小西さん” と示される 2 人の画像を提示】

1. 文脈
この 2 人の うち

2-a. 主語分裂文
去年 一郎を 手厚く 介抱したのは 竹内さんだ

2-b. 目的語分裂文
去年 一郎が 手厚く 介抱したのは 竹内さんだ

統制の結果、自己ペース読文法において動詞の読み時間に差がないことが報告されている。また矢野らは同じ例文を刺激文として与え得た際の事象関連電位についても調査結果を示しており、主語関係節の波形に比べて目的語関係節の波形が陽性に偏位していることが観察されたことを報告している。このことから、文脈を統制した場合に空所と埋語の構築の処理負荷については主語分裂文よりも目的語分裂文の方が高いことが示されている。

分裂文は新情報である焦点 (focus) を明示する構文 (焦点の右方移動) としても知られている。文脈により読み時間の実験結果に異なる結果が得られていることから、談話において新情報 (未出) なのか旧情報 (既出) なのかとは別に、受容者にとって新情報 (未知) なのか旧情報 (既知) なのか既知情報との齟齬があるのかなどの観点で読み時間が変化することは十分ありうると考える。

### 5.2 読み時間を用いた利用者毎の自動要約

2 節では、再話の生産過程の揺れを考慮したとしても、受容過程としての要約に揺れがあることを示した。利用者毎に重要な情報が異なることを明らかにするために 3 節に示したとおり、情報構造による要約作業の認知的な分析を進めている。特に accessible に相当する情報状態については、言語受容者毎に背景知識が異なるために揺れを許容したレベルでアノテーションを行い、どの要素において accessible-new 間の曖昧性が言語受容者間で起きうるのかの分析を進める。4 節で読み時間の評価方法を示したが、5.1 節で示したとおり情報構造と読み時間について、英語においては特定の構造について情報の新旧が読み時間の差をもたらすことを示している。

自動要約を含めて多くの自然言語処理は「全言語受容者に共通で不変の最適解」を教師として「最適化手法」を用いて解かれる傾向にある。この「全言語受容者に共通で不変の最適解」は言語分析の熟練作業者によるアノテーションによって行われるが、「最適化手法」で解く都合上、kappa 値などが高いアノテーションが言語処理研究者には好まれる。教師なし学習においても言語生産者側の成果物のみに基づいて統計処理を行う物が多い。

本稿では言語分析の熟練作業者を含めて言語受容過程の曖昧性を認めて、言語理解を実現するための言語受容者毎の最適化問題を定義することを試みる。視線走査可能なタブレット PC が販売されていることを鑑み、手がかりとして読み時間を用いた自動文書要約の問題を設定する。そのために必要な言語資源として、BCCWJ に対する多様な被験者による要約文、情報構造アノテーション、読み時間付与を進めている。BCCWJ に付与されている多様なアノテーション [53] と重ねあわせることで、言語処理における新しい問題設定が可能になると考える。

## 6. おわりに

本稿では現在進めている単一文書要約のための言語資源整備について、背景にある考え方を交えながら解説した。2 節では、受容過程としての要約に揺れがあることを様々なスコアを用いて示した。3 節では、要約文で提示すべき情報は何かを検討するために付与する情報構造アノテーションの方針について示した。4 節では機械生成された文の読みやすさを定性的に評価するために読み時間の評価方法について検討した。

また 2 節においては、既存の自動要約・機械翻訳の評価指標について、カーネル・距離・相関係数などとの関係を整理して数理的な構造を明らかにするとともに、多様な言語資源を用いてそれぞれの指標の特性を明らかにしようと試みた。

さらに 5.1 節では、心理言語学で研究されている情報構造と読み時間に触れて、読み時間に基づく利用者毎の文書

要約の可能性について論じた。言語資源整備が進み次第、5.2 節に述べた利用者毎の文書要約器開発を進める。

本稿では自動翻訳について議論してきたが、一部の考え方については機械翻訳にも用いられると考える。近年の英日・日英翻訳では計算量を減らすために語順の並び替えが多く行われている。この語順の並び替えにより何が失われているのかを今後明らかにしていきたい。なお、機械翻訳のために「現代日本語書き言葉均衡コーパス」のコアデータの一部を翻訳したデータ BCCWJ-Trans を作成した。表 11 に BCCWJ-Trans の摘要について示す。利用したい方は第一著者まで。

# 付　　録

## A.1　2 節で用いる用語定義

以下 2 節で用いる用語を定義する：

- 記号集合: 本稿では記号の集合を $\sigma$ で表す。
- 記号列: 何らかの全順序が付与されている記号集合。本稿では記号列ベクトル $s = \langle s_1, \ldots, s_m \rangle, t = \langle t_1, \ldots, t_m \rangle$ などで表現する
- 文字 (character), 文字ベース (character-based): 記号集合 $\sigma$ の要素の記号 $s_i \in \sigma$ としての文字。記号集合 $\sigma$ の要素が文字であること。
- 形態素 (morpheme), 形態素ベース (morpheme-based): 記号集合 $\sigma$ の要素の記号 $s_i \in \sigma$ としての形態素。記号集合 $\sigma$ の要素が形態素であること。
- 文字列 (string): 評価する記号列上の連続列。記号列の要素が文字 (character) である場合を「文字ベースの文字列 (character-based string)」、記号列の要素が形態素 (morpheme) である場合を「形態素ベースの文字列 (morpheme-based)」と呼ぶこととする
- 部分文字列 (substring): 記号列に対して隣接性と順序を保持した部分的記号列。長さ $n$ の部分文字列を特に n-gram 部分文字列と呼ぶ。
  記号列 $s$ の $i$ 番目の要素からはじまる n-gram 部分文字列を $s_{i,\ldots,i-n+1}$ で表現する。
- 部分列 (subsequence): 記号列に対して順序を保持した部分的記号列。隣接性は保持しなくてよい。長さ $p$ の部分列を特に p-mer 部分列と呼ぶ。記号列 $s$ の p-mer 部分列を、インデックスベクトル $\vec{i} = \langle i_1, \ldots, i_p \rangle (1 \leq i_1 < i_2 < \cdots < i_p \leq |s|)$ を用いて、$s[\vec{i}]$ と表す。
- 参照要約/翻訳 (reference):人間が作成した正解要約/翻訳。本稿では記号列 $R$ で表す。
- システム要約/翻訳 (candidate):要約作成器/機械翻訳器が出力した要約/翻訳。本稿では記号列 $C$ で表す。
- 距離 (distance): 集合 $X$ 上で定義された 2 変数の実数値関数で、本稿では $d : X \times X \rightarrow R$ などの記号を使う。正定値性 ($d(x,y) \geq 0$), 非退化性 ($x = y \Leftrightarrow d(x,y) \geq 0$), 対称性 ($d(x,y) = d(y,x)$), 三角不等式 ($d(x,y) + d(y,z) \geq d(x,z)$) を満たす。
- 絶対値 (absolute value):大きさの一般化概念。実数については 0 からの距離、集合については要素数を表すのに用い $|x|$ で表す。
- $\theta$-ノルム (norm):ベクトル空間上に距離を規定する長さの一般化概念。ベクトル $x = \langle , x_1, \ldots, x_n \rangle$ の $\theta$-ノルムを $||x||_\theta = (\sum_{i=1}^{n} |x_i|^p)^{1/p}$ により定義する。特に $\theta$ を定義しない場合 ($||x||$) は 2-ノルムを用いる。
- 内積記号 $\cdot$: 文字列に対しては連結、整数・実数については積、ベクトルなどについては内積、対称群については写像の合成 (積) を扱うために用いる
- 類似度 (similarity):二つの元の距離は遠さを表すのに対し、類似度は近さを表す。距離の逆数は類似度として扱える。
- 相関係数 (correlation): 二つの確率変数の間の相関を表す指標で、類似度として扱える。[-1,1] 区間の値をとり、1 に近い場合は正の相関があると呼び、-1 に近い場合には負の相関があると呼ぶ。0 に近い場合には相関が弱いという意味がある。
- カーネル関数 (kernel function): 特徴空間中の座標の明示的な計算を経由せずに特徴量空間における内積 (正定値性と非退化性をもち、実数ベクトル空間では対称性ももつ) を定義するもの。本稿では $K(s,t)$ と表記する。内積を正規化することにより cosine 類似度 ($\frac{K(s,t)}{||K(s,s)|| \cdot ||K(t,t)||}$) を定義することができる。
- スコア (score): 類似度を [0,1] 区間に正規化したもの。本稿では score などの記号で示す。
- 接頭辞 (prefix):記号列の先頭要素を含む連続文字列
- 接尾辞 (suffix):記号列の末尾要素を含む連続文字列
- 部分集合 (subset): 記号列を集合とみなした場合の部分集合。隣接性と順序は保持しなくてよい。要素数 $k$ の部分集合を特に $k$-element 部分集合と呼ぶ。
- 順位ベクトル (rank vector): インデックス $i$ 要素が対象 $i$ の順位を表すベクトル。本稿では $m$ 次元の順位ベクトル空間を $S_m$ で表し、順位ベクトル空間の要素である順位ベクトルを $\mu = \langle \mu(1), \ldots, \mu(m) \rangle$ で表す。$\mu(i)$ には対象 $i$ の順位を表す自然数が入る。
- 順序ベクトル (order vector): 順位が $i$ 番目である要素がインデックス $i$ の位置に格納されているベクトル。本稿では $m$ 次元の順序ベクトル空間を $T_m$ で表し、順位ベクトル $\mu(i)$ に対応する順序ベクトルを $\mu^{-1} = \langle \mu^{-1}(1), \ldots, \mu^{-1}(m) \rangle$ で表す。$\mu^{-1}(i)$ には順位が $i$ である要素 (の順位ベクトル上でのインデックス) が入る。
- 同順 (concordant): 二つの順位ベクトル中で対象対 $i$ と $j$ が以下を満たすとき、その対象対が同順であると

表 11 BCCWJ-Trans の概要

| 言語 | 文書数 | 文数 | 下訳 | 摘要 |
|---|---|---|---|---|
| 英語 | 6 | 319 | 有 | OY 1, OC 1, PN 1, PB 1, PM 1, OW 1 |
| 中国語 (簡) | 6 | 319 | 有 | OY 1, OC 1, PN 1, PB 1, PM 1, OW 1 |
| イタリア語 | 16 | 436 | 無 | OY 6, OC 6, PN 1, PB 1, PM 1, OW 1 |
| インドネシア語 | 10 | 337 | 無 | OY 3, OC 3, PN 1, PB 1, PM 1, OW 1 |

文数は日本語側のもの。文書はアノテーションの優先順位順に選択。
摘要で用いている記号の意味: OY “ブログ”, OC “知恵袋”, PN “新聞”, PB “書籍”, PM “雑誌”, OW “白書”。

いう。
$(\mu(i)-\mu(j))(\nu(i)-\nu(j)) \geq 0$

- 逆順 (discordant): 二つの順位ベクトル中で対象対が同順でないことを逆順という。
- 文字列上の編集:挿入 (insertion)、削除 (deletion)、代入 (substitution) の三つを規定する。
- 順序ベクトル上の編集:順序ベクトルを対称群 (symmetric group) と考えて編集する際の操作を規定する。:並び替えの編集操作 (置換:permutation) を元とする群。順序ベクトル $\mu^{-1} = \langle \mu^{-1}(1), \ldots, \mu^{-1}(m) \rangle$ のうち、$\mu^{-1}(k_1), \mu^{-1}(k_2), \ldots, \mu^{-1}(k_r)$ 以外は動かさず、$\mu^{-1}(k_1) \to \mu^{-1}(k_2), \mu^{-1}(k_2) \to \mu^{-1}(k_3), \ldots$ のように順にずらす置換
$$\begin{pmatrix} \mu^{-1}(k_1) & \mu^{-1}(k_2) & \ldots & \mu^{-1}(k_r) \\ \mu^{-1}(k_2) & \mu^{-1}(k_3) & \ldots & \mu^{-1}(k_1) \end{pmatrix}$$
のことを巡回置換と呼び、$\pi_r = (k_1, k_2, \ldots, k_r)$ で表す。二つの元のみを入れ替えて他の元は変えないもの (2 元の巡回置換) を互換 (transposition) と呼び、$\pi_2 = (i, j)$ で表す。隣接する二つの元のみを入れ替えて他の元は変えないものを隣接互換 (adjacent transposition) と呼び、$\pi_2 = (i, i+1)$ で表す。
- クロネッカーのデルタ $\delta$ : $\delta(i,j) = \begin{cases} 1 & (i=j) \\ 0 & (i \neq j) \end{cases}$

## A.2 要約課題とスコア

表 A·1 に要約課題・語釈課題と各種スコアの比較を表 A·2 に再話課題と各種スコアの比較を示す。標準偏差は BCCWJ-SUMM_L(T) が最も大きい。これは繰り返し要約する際に全く同じ要約文を再生産する被験者と全く異なる要約文を再生産する被験者とが存在するからだと考えられる。しかし、他の再話 (RETELLING_K(T),RETELLING_M(T)) でも被験者間の標準偏差と比して高いことから要約文特有の現象ではないと考える。

## A.3 視線走査実験結果例

表 A·3 に視線停留順の視線走査実験結果例を示す。オフセット値は半文字単位 (0-origin) で横方向に 0-107(54 文字分) 区間定義している。時間・時刻の単位はミリ秒である。文節は国語研文節単位、形態素は国語研短単位による。表中 “□” は全角空白を意味する。文節、形態素の空白は全角空白以外で文字がない部分の視線停留を意味する。表 A·4,A·5 に文書出現順に集計した視線走査実験結果例 (文節単位集計・形態素単位集計) を示す。時間・時刻の単位にはミリ秒の実時間 (左) と文字数で回帰分析した標準化残差 (右) を示す。標準化残差は約 95% が [-2.0, 2.0] 区間に入るため、2.0 以上を外れ値とする (表中 2.0 以上を *、2.5 以上を**付記)。残差分析においては、語順などの他に読み時間に影響を与えるものを今後考慮して行う必要がある。

## 謝辞

要約文の作成および評価については NTT CS 研の平尾努氏の助言を受けました。情報構造と読み時間の関係および読み時間の評価方法については津田塾大学の小野創先生の助言を受けました。本研究の一部は、国立国語研究所基幹型共同研究プロジェクト「コーパスアノテーションの基礎研究」および国立国語研究所「超大規模コーパス構築プロジェクト」によるものです。本研究は JSPS 科研費 基盤研究 (B) 25284083, 若手研究 (B) 26770156 の助成を受けたものです。若手研究 (B)26770167 の助成は受けておりません。

### 参考文献

[1] Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M. and Den, Y.: Balanced corpus of contemporary written Japanese, *Language Resources and Evaluation*, Vol. 48, pp. 345–371 (2014).

[2] Götze, M., Weskott, T., Endriss, C., Fiedler, I., Hinterwimmer, S., Petrova, S., Schwarz, A., Skopeteas, S. and Stoel, R.: Information structure, *Interdisciplinary Studies on Information Structure* (Dipper, S., Göetze, M., Skopeteas, S. and Stoel, R., eds.), Working Papers of the SFB 632, Vol. 7, Universiätsverlag, Potsdam, 2nd edition (updated version 2014) edition, chapter Information Structure, pp. 147–187 (2007).

[3] Lin, C.-Y.: ROUGE: A Package for Automatic Evaluation of Summaries, *Proc. of Workshop on Summarization Branches Out, Post Conference Workshop of ACL 2004*, pp. 74–81 (2004).

[4] Lin, C.-Y. and Hovy, E.: Automatic Evaluation of Summaries Using N-gram Co-occurrence Statistics, *Proc. of the 4th Meeting of the North American Chapter of the Association for Computational Linguistics and Human Language Technology*, pp. 150–157 (2003).

[5] Lin, C.-Y. and Och, F. J.: Automatic Evluation of Machine Translation Quality Using Longest Common Subsequence and Skip-Bigram Statistics, *Proceedings of*

**表 A·1** 要約課題・語釈課題と各種スコア

| score | BCCWJ-SUMM_C | | BCCWJ-SUMM_L(P) | | BCCWJ-SUMM_L(T) | | GROSS_C | | GROSS_L(P) | | GROSS_L(T) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| n-gram(1)_c | 0.63 | 0.12 | 0.64 | 0.11 | 0.68 | 0.27 | 0.68 | 0.09 | 0.65 | 0.06 | 0.83 | 0.07 |
| n-gram(2)_c | 0.36 | 0.17 | 0.33 | 0.15 | 0.48 | 0.32 | 0.24 | 0.10 | 0.22 | 0.09 | 0.56 | 0.16 |
| n-gram(3)_c | 0.26 | 0.17 | 0.22 | 0.15 | 0.40 | 0.34 | 0.11 | 0.07 | 0.09 | 0.06 | 0.41 | 0.18 |
| n-gram(4)_c | 0.20 | 0.16 | 0.16 | 0.14 | 0.35 | 0.34 | 0.05 | 0.06 | 0.04 | 0.04 | 0.31 | 0.18 |
| n-gram(≤2)_c | 0.52 | 0.13 | 0.53 | 0.12 | 0.61 | 0.28 | 0.55 | 0.09 | 0.51 | 0.07 | 0.75 | 0.10 |
| n-gram(≤3)_c | 0.45 | 0.14 | 0.45 | 0.12 | 0.55 | 0.29 | 0.47 | 0.08 | 0.42 | 0.06 | 0.68 | 0.11 |
| n-gram(≤4)_c | 0.40 | 0.14 | 0.39 | 0.12 | 0.51 | 0.30 | 0.41 | 0.08 | 0.36 | 0.06 | 0.62 | 0.12 |
| p-mer(2)_c | 0.34 | 0.14 | 0.35 | 0.13 | 0.49 | 0.33 | 0.39 | 0.10 | 0.34 | 0.07 | 0.63 | 0.14 |
| p-mer(3)_c | 0.18 | 0.13 | 0.18 | 0.12 | 0.37 | 0.35 | 0.20 | 0.08 | 0.15 | 0.05 | 0.45 | 0.17 |
| p-mer(4)_c | 0.09 | 0.10 | 0.09 | 0.10 | 0.30 | 0.36 | 0.09 | 0.06 | 0.06 | 0.03 | 0.31 | 0.17 |
| p-mer(≤2)_c | 0.34 | 0.14 | 0.35 | 0.13 | 0.49 | 0.33 | 0.39 | 0.10 | 0.34 | 0.07 | 0.63 | 0.14 |
| p-mer(≤3)_c | 0.18 | 0.13 | 0.18 | 0.12 | 0.37 | 0.35 | 0.20 | 0.08 | 0.16 | 0.05 | 0.45 | 0.17 |
| p-mer(≤4)_c | 0.10 | 0.11 | 0.09 | 0.10 | 0.30 | 0.36 | 0.10 | 0.06 | 0.06 | 0.03 | 0.31 | 0.17 |
| Footrule_c | 0.50 | 0.15 | 0.50 | 0.14 | 0.59 | 0.29 | 0.48 | 0.10 | 0.45 | 0.08 | 0.69 | 0.14 |
| Kendall_c | 0.48 | 0.14 | 0.47 | 0.13 | 0.57 | 0.29 | 0.44 | 0.08 | 0.41 | 0.06 | 0.64 | 0.13 |
| n-gram(1)_m | 0.60 | 0.12 | 0.62 | 0.11 | 0.67 | 0.27 | 0.62 | 0.10 | 0.58 | 0.07 | 0.79 | 0.09 |
| n-gram(2)_m | 0.25 | 0.16 | 0.24 | 0.15 | 0.41 | 0.34 | 0.13 | 0.08 | 0.11 | 0.07 | 0.42 | 0.18 |
| n-gram(3)_m | 0.15 | 0.15 | 0.14 | 0.14 | 0.34 | 0.35 | 0.04 | 0.06 | 0.03 | 0.04 | 0.27 | 0.18 |
| n-gram(4)_m | 0.10 | 0.13 | 0.09 | 0.12 | 0.29 | 0.35 | 0.02 | 0.05 | 0.01 | 0.02 | 0.19 | 0.16 |
| n-gram(≤2)_m | 0.46 | 0.13 | 0.48 | 0.12 | 0.57 | 0.29 | 0.48 | 0.09 | 0.43 | 0.07 | 0.67 | 0.11 |
| n-gram(≤3)_m | 0.38 | 0.13 | 0.39 | 0.13 | 0.51 | 0.30 | 0.39 | 0.08 | 0.34 | 0.06 | 0.58 | 0.12 |
| n-gram(≤4)_m | 0.32 | 0.13 | 0.33 | 0.12 | 0.47 | 0.31 | 0.33 | 0.07 | 0.28 | 0.05 | 0.52 | 0.13 |
| p-mer(2)_m | 0.30 | 0.14 | 0.32 | 0.14 | 0.47 | 0.33 | 0.32 | 0.10 | 0.27 | 0.07 | 0.55 | 0.15 |
| p-mer(3)_m | 0.15 | 0.12 | 0.15 | 0.12 | 0.35 | 0.35 | 0.14 | 0.07 | 0.11 | 0.04 | 0.35 | 0.17 |
| p-mer(4)_m | 0.07 | 0.09 | 0.07 | 0.10 | 0.28 | 0.36 | 0.06 | 0.05 | 0.03 | 0.02 | 0.22 | 0.16 |
| p-mer(≤2)_m | 0.31 | 0.14 | 0.33 | 0.14 | 0.47 | 0.33 | 0.32 | 0.10 | 0.28 | 0.07 | 0.55 | 0.15 |
| p-mer(≤3)_m | 0.16 | 0.12 | 0.16 | 0.13 | 0.36 | 0.35 | 0.14 | 0.07 | 0.11 | 0.04 | 0.36 | 0.17 |
| p-mer(≤4)_m | 0.08 | 0.10 | 0.08 | 0.10 | 0.29 | 0.36 | 0.06 | 0.05 | 0.04 | 0.02 | 0.23 | 0.17 |
| Footrule_m | 0.48 | 0.15 | 0.48 | 0.14 | 0.59 | 0.30 | 0.44 | 0.10 | 0.42 | 0.10 | 0.66 | 0.14 |
| Kendall_m | 0.46 | 0.15 | 0.46 | 0.14 | 0.56 | 0.29 | 0.41 | 0.09 | 0.39 | 0.08 | 0.61 | 0.13 |

*the 42nd Annual Meeting on Association for Computational Linguistics*, pp. 311–318 (2004).

[6] 平尾　努，奥村　学，磯崎秀樹：拡張ストリングカーネルを用いた要約システムの自動評価法，情報処理学会論文誌，Vol. 47, No. 6, pp. 1753–1765 (2006).

[7] Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J.: Bleu: a Method for Automatic Evaluation of Machine Translation, Technical report, IBM Research Report RC22176 (W0109-022) (2001).

[8] Echizen-ya, H. and Araki, K.: Automatic Evaluation of Machine Translation based on Recursive Acquisition of an Intuitive Common Parts Continuum, *Proceedings of the MT Summit XII Workshop on Patent Translation*, pp. 151–158 (2007).

[9] 平尾　努，磯崎秀樹，須藤克仁，Kevin, D.，塚田　元，永田昌明：語順の相関に基づく機械翻訳の自動評価法，自然言語処理，Vol. 21, No. 3, pp. 411–444 (2014).

[10] Birch, A. and Osborne, M.: LRscore for Evaluation Lexical and Reordering Quality in MT, *Proceedings of the Joint 5th Workshop on Statistical Machine Translation and MetricsMATR*, pp. 327–332 (2010).

[11] Shawe-Taylor, J.，Cristianini, N.，大北剛 (訳)：カーネル法によるパターン解析 (Kernel Methods for Pattern Analysis)，chapter 第 11 章構造化データに対するカーネル: 文字列、木など，共立出版 (2010).

[12] 神嶌敏弘：順序の距離と確率モデル，人工知能学会研究会資料 SIG-DMSM-A902-07 (2009).

[13] Nivre, J.: Non-Projective Dependency Parsing in Expected Linear Time, *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pp. 351–359 (2009).

[14] 平尾　努，奥村　学，安田宣仁，磯崎秀樹：投票型回帰モデルによる要約自動評価法，人工知能学会論文誌，Vol. 22, No. 2, pp. 115–126 (2007).

[15] 天野成昭，近藤公久：日本語の語彙特性 第 1 期 CD-ROM 版，三省堂 (1999).

[16] 保田　祥，田中弥生，荒牧英治：繰り返しにおける独話の変化，社会言語科学会第 31 回大会発表論文集，pp. 190–193 (2013).

[17] 保田　祥，田中弥生，荒牧英治：同じ話であるとはどういうことか，社会言語科学会第 32 回大会発表論文集 (2013).

[18] 宮部真衣，四方朱子，久保　圭，荒牧英治：音声認識による認知症・発達障害スクリーニングは可能か？–言語能力測定システム”言秤”の提案–，グループウェアとネットワークサービスワークショップ 2014 (2014).

[19] 保田　祥，荒牧英治：人が同じ話を何度もするとどうなるか？：繰り返しによって生じる物語独話の変化，日本認知科学会第 29 回 (2012).

[20] 保田　祥：同じ話を成立させる語 – 「桃太郎」を「桃太郎」として成立させる語彙–，社会言語科学会第 33 回大会発表論文集 (2014).

[21] Lambrecht, K.: *Information Structure and Sentence Form*, Cambredge University Press (1994).

[22] Prince, E. F.: *Discourse description: Diverse linguistic*

**表 A·2** 再話課題と各種スコア

| score | RETELLING_I(T) | | RETELLING_K(P) | | RETELLING_K(T) | | RETELLING_M(P) | | RETELLING_M(T) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| n-gram(1)_c | 0.96 | 0.09 | 0.86 | 0.05 | 0.91 | 0.19 | 0.87 | 0.08 | 0.92 | 0.13 |
| n-gram(2)_c | 0.85 | 0.12 | 0.59 | 0.10 | 0.78 | 0.21 | 0.58 | 0.20 | 0.75 | 0.16 |
| n-gram(3)_c | 0.73 | 0.17 | 0.39 | 0.13 | 0.65 | 0.22 | 0.44 | 0.21 | 0.64 | 0.18 |
| n-gram(4)_c | 0.61 | 0.19 | 0.23 | 0.13 | 0.50 | 0.21 | 0.33 | 0.18 | 0.52 | 0.18 |
| n-gram(≤2)_c | 0.94 | 0.09 | 0.80 | 0.06 | 0.89 | 0.19 | 0.80 | 0.11 | 0.88 | 0.13 |
| n-gram(≤3)_c | 0.91 | 0.09 | 0.76 | 0.07 | 0.86 | 0.19 | 0.74 | 0.12 | 0.84 | 0.13 |
| n-gram(≤4)_c | 0.89 | 0.10 | 0.73 | 0.07 | 0.84 | 0.18 | 0.70 | 0.13 | 0.81 | 0.14 |
| p-mer(2)_c | 0.91 | 0.13 | 0.72 | 0.09 | 0.85 | 0.19 | 0.72 | 0.13 | 0.83 | 0.14 |
| p-mer(3)_c | 0.86 | 0.16 | 0.60 | 0.11 | 0.79 | 0.21 | 0.58 | 0.16 | 0.74 | 0.16 |
| p-mer(4)_c | 0.81 | 0.18 | 0.49 | 0.13 | 0.73 | 0.22 | 0.45 | 0.16 | 0.64 | 0.19 |
| p-mer(≤2)_c | 0.91 | 0.13 | 0.72 | 0.09 | 0.85 | 0.19 | 0.72 | 0.13 | 0.83 | 0.14 |
| p-mer(≤3)_c | 0.86 | 0.16 | 0.60 | 0.11 | 0.79 | 0.21 | 0.58 | 0.16 | 0.74 | 0.16 |
| p-mer(≤4)_c | 0.81 | 0.18 | 0.49 | 0.13 | 0.73 | 0.22 | 0.45 | 0.16 | 0.64 | 0.19 |
| Footrule_c | 0.88 | 0.11 | 0.73 | 0.07 | 0.83 | 0.18 | 0.75 | 0.13 | 0.85 | 0.14 |
| Kendall_c | 0.81 | 0.11 | 0.66 | 0.06 | 0.77 | 0.17 | 0.68 | 0.12 | 0.78 | 0.13 |
| n-gram(1)_m | 0.94 | 0.15 | 0.85 | 0.06 | 0.92 | 0.18 | 0.83 | 0.07 | 0.90 | 0.13 |
| n-gram(2)_m | 0.75 | 0.18 | 0.48 | 0.10 | 0.72 | 0.21 | 0.45 | 0.18 | 0.64 | 0.17 |
| n-gram(3)_m | 0.58 | 0.21 | 0.21 | 0.13 | 0.48 | 0.19 | 0.28 | 0.15 | 0.47 | 0.16 |
| n-gram(4)_m | 0.39 | 0.21 | 0.09 | 0.07 | 0.30 | 0.15 | 0.10 | 0.08 | 0.27 | 0.15 |
| n-gram(≤2)_m | 0.92 | 0.15 | 0.78 | 0.08 | 0.88 | 0.18 | 0.73 | 0.10 | 0.84 | 0.13 |
| n-gram(≤3)_m | 0.89 | 0.15 | 0.73 | 0.08 | 0.85 | 0.18 | 0.66 | 0.11 | 0.78 | 0.13 |
| n-gram(≤4)_m | 0.86 | 0.15 | 0.69 | 0.09 | 0.82 | 0.18 | 0.60 | 0.11 | 0.73 | 0.13 |
| p-mer(2)_m | 0.90 | 0.18 | 0.70 | 0.11 | 0.86 | 0.19 | 0.64 | 0.11 | 0.78 | 0.14 |
| p-mer(3)_m | 0.85 | 0.19 | 0.56 | 0.13 | 0.79 | 0.19 | 0.46 | 0.13 | 0.66 | 0.17 |
| p-mer(4)_m | 0.79 | 0.20 | 0.44 | 0.14 | 0.73 | 0.20 | 0.32 | 0.12 | 0.54 | 0.18 |
| p-mer(≤2)_m | 0.90 | 0.18 | 0.70 | 0.11 | 0.86 | 0.19 | 0.64 | 0.11 | 0.78 | 0.14 |
| p-mer(≤3)_m | 0.85 | 0.19 | 0.56 | 0.13 | 0.79 | 0.19 | 0.46 | 0.13 | 0.66 | 0.17 |
| p-mer(≤4)_m | 0.79 | 0.20 | 0.44 | 0.14 | 0.73 | 0.20 | 0.33 | 0.12 | 0.54 | 0.18 |
| Footrule_m | 0.86 | 0.16 | 0.71 | 0.07 | 0.83 | 0.18 | 0.72 | 0.10 | 0.83 | 0.14 |
| Kendall_m | 0.80 | 0.16 | 0.63 | 0.06 | 0.76 | 0.17 | 0.66 | 0.10 | 0.77 | 0.13 |

**表 A·3** 視線走査実験結果例 (視線停留順)

| 画面 ID | 行番号 | オフセット値 | 停留順 | 停留開始時刻 | 停留終了時刻 | 停留時間 | 文節 | 形態素 |
|---|---|---|---|---|---|---|---|---|
| 14 | 1 | 0 | 1 | 7 | 209 | 203 | ◆ | ◆ |
| 14 | 1 | 7 | 2 | 231 | 434 | 204 | 外貨準備高、 | 準備 |
| 14 | 1 | 15 | 3 | 459 | 623 | 165 | 3カ月ぶり増 | 3 |
| 14 | 1 | 18 | 4 | 743 | 772 | 30 | 3カ月ぶり増 | カ月 |
| 14 | 1 | 5 | 5 | 825 | 1026 | 202 | 外貨準備高、 | 外貨 |
| 14 | 1 | 5 | 6 | 1047 | 1072 | 26 | 外貨準備高、 | 外貨 |
| 14 | 1 | 4 | 7 | 1164 | 1223 | 60 | 外貨準備高、 | 外貨 |
| 14 | 1 | 5 | 8 | 1245 | 1298 | 54 | 外貨準備高、 | 外貨 |
| 14 | 1 | 8 | 9 | 1317 | 1600 | 284 | 外貨準備高、 | 準備 |
| 14 | 1 | 4 | 10 | 1614 | 1689 | 76 | 外貨準備高、 | 外貨 |
| 14 | 1 | 18 | 11 | 1722 | 1836 | 115 | 3カ月ぶり増 | カ月 |
| 14 | 1 | 20 | 12 | 2127 | 2164 | 38 | 3カ月ぶり増 | カ月 |
| 14 | 2 | 5 | 13 | 2236 | 2388 | 153 | □財務省が | 財務 |
| 14 | 2 | 3 | 14 | 2399 | 2860 | 462 | □財務省が | 財務 |
| 14 | 2 | 11 | 15 | 3170 | 3239 | 70 | 9日 | 9 |
| 14 | 2 | 12 | 16 | 3258 | 3490 | 233 | 9日 | 9 |
| 14 | 2 | 16 | 17 | 3507 | 3738 | 232 | 発表した | 発表 |
| 14 | 2 | 26 | 18 | 3766 | 3993 | 228 | 4月末の | 月 |
| 14 | 2 | 24 | 19 | 4004 | 4089 | 86 | 4月末の | 4 |
| 14 | 1 | 31 | 20 | 4371 | 4396 | 26 | | |

**表 A·4** 視線走査実験結果例 (元文書出現順:文節単位集計)

| 画面 ID | 行番号 | 文節 ID | First Pass | | Total | | Reg. Path | | 文節 |
|---|---|---|---|---|---|---|---|---|---|
| 14 | 1 | 0 | 203 | 0.69 | 203 | 0.27 | 203 | 0.09 | ◆ |
| 14 | 1 | 1 | 204 | -0.17 | 906 | 0.92 | 204 | -0.38 | 外貨準備高、 |
| 14 | 1 | 2 | 195 | -0.21 | 348 | -0.36 | 1203 | 0.52 | 3カ月ぶり増 |
| 14 | 2 | 0 | 615 | 1.76 | 672 | 0.58 | 615 | 0.09 | □財務省が |
| 14 | 2 | 1 | 303 | 0.95 | 303 | 0.31 | 303 | 0.09 | 9日 |
| 14 | 2 | 2 | 232 | 0.30 | 232 | -0.24 | 232 | -0.17 | 発表した |
| 14 | 2 | 3 | 314 | 0.65 | 1048 | 1.63 | 314 | -0.09 | 4月末の |
| 14 | 2 | 4 | 144 | -0.42 | 907 | 0.92 | 144 | -0.43 | 外貨準備高は |
| 14 | 2 | 5 | 612 | 0.19 | 612 | -1.30 | 612 | -0.76 | 三千六百二十六億千百万ドルと |
| 14 | 2 | 6 | 322 | 0.86 | 322 | 0.16 | 422 | 0.10 | なり、 |
| 14 | 2 | 7 | 0 | -0.69 | 0 | -0.77 | 0 | -0.38 | 前月末に |
| 14 | 3 | 0 | 0 | -0.35 | 0 | -0.39 | 0 | -0.19 | 比べ |
| 14 | 3 | 1 | 553 | 0.11 | 1299 | 0.47 | 553 | -0.72 | 十一億三千九百万ドル増と、 |
| 14 | 3 | 2 | 260 | 0.07 | 481 | -0.06 | 260 | -0.33 | 3カ月ぶりに |
| 14 | 3 | 3 | 353 | 0.82 | 603 | 0.61 | 1348 | 0.84 | プラスに |
| 14 | 3 | 4 | 0 | -0.69 | 0 | -0.77 | 0 | -0.38 | 転じた。 |

*analyses of a fund-raising text*, chapter The ZPG letter: Subjects, definiteness, and information status, pp. 295–325, Benjamins (1992).

[23] Jacobs, J.: The dimensions of topic-comment, *Linguistics*, Vol. 39, pp. 641–681 (2001).

[24] Cook, P. and Bildhauer, F.: Identifying "aboutness topics": two annotation experiments, *Dialogue and Discourse*, Vol. 4, No. 2, pp. 118–141 (2013).

[25] Pustejovsky, J. and Stubbs, A.: *Natural Language Annotatino for Machine Learning – A Guide to Corpus-Buildling for Applications*, O'Reilly (2012).

[26] 飯田 龍，小町 守，井之上直也，乾健太郎，松本裕治：述語項構造と照応関係のアノテーション：NAIST テキストコーパス構築の経験から，自然言語処理，Vol. 17, No. 2, pp. 25–50 (2010).

[27] 日本語記述文法研究会（編）：現代日本語文法5，くろしお出版 (2009).

[28] 浅原正幸，狩野芳伸，小野 創，植田禎子：『現代日本語書き言葉均衡コーパス』に対する読文時間・視線情報アノテーションの向けて，第1回コーパスアノテーションワークショップ (2012).

[29] Kennedy, A. and Pynte, J.: Parafoveal-on-foveal effects in normal reading, *Vision Research*, Vol. 45, pp. 153–168 (2005).

[30] Demberg, V. and Keller, F.: Data from eye-tracking corpora as evidence for theories of syntactic processing complexity, *Cognition*, Vol. 109, No. 2, pp. 193–210 (2008).

[31] Gibson, E.: Linguistic complexity: Locality of syntactic dependencies, *Cognition*, Vol. 68, pp. 1–76 (1998).

[32] Hale, J.: A probabilistic earley parser as a psycholinguistic model, *Proc. of the second conference of the North American chapter of the association for computational linguistics*, Vol. 2, pp. 159–166 (2001).

[33] Roland, D., Mauner, G., O'Maera, C. and Yun, H.: Discourse expectations and relative clause processing, *Journal of Memory and Language*, Vol. 66, No. 3, pp. 479–508 (2012).

[34] Demberg, V. and Keller, F.: Eye-tracking evidence for integration cost effects in corpus data, *Proc. of the 29th meeting of the cognitive science society (CogSci-07)* (2007).

[35] Just, M. A., Carpenter, P. A. and Woolley, J. D.: Paradigms and Processes in Reading Comprehension, *Journal of Experimental Psychology: General*, Vol. 3, pp. 228–238 (1982).

[36] 浅原正幸，池本 優，森田敏生：コーパスコンコーダンサ『ChaKi.NET』の連続値データ型 (2) – 読み時間の表示 –，第5回コーパス日本語学ワークショップ予稿集，pp. 39–48 (2014).

[37] Amano, S. and Kondo, T.: Estimation of mental lexicon size with word familiarity database, *Proceedings of International Conference on Spoken Language Processing*, Vol. 5, pp. 2119–2122 (1998).

[38] 苧阪満里子：脳のメモ帳 ワーキングメモリ，新曜社 (2002).

[39] 小野 創：実験言語学 第3回依存関係の構築 Part 2，日本言語学会夏期講座 2012 (2012).

[40] King, J. and Just, M. A.: Individual Differences in Syntactic Processing: The role of Working Memory, *Journal of Memory and Language*, Vol. 30, pp. 580–602 (1991).

[41] Hawkins, J. A.: Processing Complexity and Filler-Gap Dependencies Across Grammars, *Language*, Vol. 75, pp. 244–285 (1999).

[42] O'Grady, W.: *Syntactic Development*, The University of Chicago Press (1997).

[43] Miyamoto, E. T. and Nakamura, M.: Subject/object asymmetries in the processing of relative clauses in Japanese, *Proceedings of the 22nd West Coast Conference on Formal Linguistics* (Garding, G. and Tsujimura, M., eds.), pp. 342–355 (2003).

[44] Roland, D., O'Meara, C., Yun, H. and Mauner, G.: Processing object relative clauses: Discourse or frequency, *Poster presented at the CUNY sentence processing conference* (2007).

[45] 佐藤 淳：日本語関係節の処理負荷を決定する要因の検討：コーパスにおける使用頻度の影響を中心に，博士論文，広島大学教育学研究科 (2011).

[46] Gordon, P. C., Hendrick, R. and Johnson, M.: Memory Interference During Language Processing, *Jounal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 27, No. 6, pp. 1411–1423 (2001).

[47] Gordon, P. C., Hendrick, R. and Levine, W. H.: Memory-load -interference in Syntacitic Processing, *Psychological Science*, Vol. 13, No. 5, pp. 425–430 (2002).

**表 A·5** 視線走査実験結果例 (元文書出現順:単語単位集計)

| 画面 ID | 行番号 | 形態素 ID | First Pass | | Total | | Reg. Path | | 形態素 |
|---|---|---|---|---|---|---|---|---|---|
| 14 | 1 | 0 | 203 | 1.10 | 203 | 0.61 | 203 | 0.15 | ◆ |
| 14 | 1 | 1 | 342 | 1.71 | 418 | 1.27 | 342 | 0.18 | 外貨 |
| 14 | 1 | 2 | 204 | 0.66 | 488 | 1.63 | 204 | -0.06 | 準備 |
| 14 | 1 | 3 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | 高 |
| 14 | 1 | 4 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | 、 |
| 14 | 1 | 5 | 165 | 0.81 | 165 | 0.41 | 165 | 0.08 | 3 |
| 14 | 1 | 6 | 30 | -0.67 | 183 | 0.07 | 1038 | 1.41 | カ月 |
| 14 | 1 | 7 | 0 | -0.90 | 0 | -0.86 | 0 | -0.42 | ぶり |
| 14 | 1 | 8 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | 増 |
| 14 | 2 | 0 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | □ |
| 14 | 2 | 1 | 615 | 3.80** | 672 | 2.57** | 615 | 0.66 | 財務 |
| 14 | 2 | 2 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | 省 |
| 14 | 2 | 3 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | が |
| 14 | 2 | 4 | 303 | 1.86 | 303 | 1.12 | 303 | 0.32 | 9 |
| 14 | 2 | 5 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | 日 |
| 14 | 2 | 6 | 232 | 0.87 | 232 | 0.32 | 232 | -0.01 | 発表 |
| 14 | 2 | 7 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | し |
| 14 | 2 | 8 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | た |
| 14 | 2 | 9 | 86 | 0.21 | 86 | 0.01 | 86 | -0.06 | 4 |
| 14 | 2 | 10 | 228 | 1.29 | 534 | 2.30* | 375 | 0.45 | 月 |
| 14 | 2 | 11 | 147 | 0.67 | 147 | 0.32 | 734 | 1.08 | 末 |
| 14 | 2 | 12 | 281 | 1.70 | 281 | 1.00 | 281 | 0.29 | の |
| 14 | 2 | 13 | 144 | 0.20 | 398 | 1.17 | 144 | -0.17 | 外貨 |
| 14 | 2 | 14 | 339 | 1.69 | 339 | 0.87 | 339 | 0.18 | 準備 |
| 14 | 2 | 15 | 170 | 0.85 | 170 | 0.44 | 170 | 0.09 | 高 |
| 14 | 2 | 16 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | は |
| 14 | 2 | 17 | 116 | -0.01 | 116 | -0.27 | 116 | -0.22 | 三千 |
| 14 | 2 | 18 | 0 | -0.90 | 0 | -0.86 | 0 | -0.42 | 六百 |
| 14 | 2 | 19 | 0 | -0.90 | 0 | -0.86 | 0 | -0.42 | 二十 |
| 14 | 2 | 20 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | 六 |
| 14 | 2 | 21 | 231 | 1.31 | 231 | 0.75 | 231 | 0.20 | 億 |
| 14 | 2 | 22 | 265 | 1.57 | 265 | 0.92 | 934 | 1.44 | 千 |
| 14 | 2 | 23 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | 百 |
| 14 | 2 | 24 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | 万 |
| 14 | 2 | 25 | 0 | -0.90 | 0 | -0.86 | 0 | -0.42 | ドル |
| 14 | 2 | 26 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | と |
| 14 | 2 | 27 | 0 | -0.90 | 0 | -0.86 | 0 | -0.42 | なり |
| 14 | 2 | 28 | 322 | 2.01* | 322 | 1.21 | 422 | 0.53 | 、 |
| 14 | 2 | 29 | 0 | -0.90 | 0 | -0.86 | 0 | -0.42 | 前月 |
| 14 | 2 | 30 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | 末 |
| 14 | 2 | 31 | 0 | -0.45 | 0 | -0.43 | 0 | -0.21 | に |

[48] Waters, G., Caplan, D. and Yampolsky, S.: On-line syntactic processing under concurrent memory load, *Psychonomic Bulletin and Review*, Vol. 10, pp. 88–95 (2003).
[49] Kahraman, B., Sato, A., Ono, H. and Sakai, H.: Why object clefts are easier than subject clefts in Japanese: Frequency or expectation?, *IEICE Technical Report*, Vol. 111, No. 170, pp. 67–72 (2011).
[50] Kahraman, B.: Processing "gap-filler dependencies" in Japanese and Turkish: Regarding the incrementality of sentence processsing, PhD Thesis, 広島大学教育学研究科 (2011).
[51] Kahraman, B., Sato, A., Ono, H. and Sakai, H.: Incremental processing of gap-filler dependencies: Evidence from the processing of subject and object clefts in Japanese, *The Proceedings of the 12th Tokyo Conference on Psycholinguistics*, 東京, ひつじ書房, pp. 133–147 (2011).
[52] 矢野雅貴，立川　憂，坂本　勉：gap-filler 依存関係の処理について–文脈を用いた日本語分裂文の ERP 研究–，日本言語学会第 146 回大会予稿集，pp. 252–257 (2013).
[53] 前川喜久雄：「コーパスアノテーションの基礎研究」および「コーパス日本語学の創成」，国語研プロジェクトレビュー，Vol. 3, No. 2, pp. 63–83 (2012).
[54] 難波英嗣，平尾　努：テキスト要約の自動評価，人工知能学会誌，Vol. 23, No. 1, pp. 10–16 (2008).