

アノテーション時の作業者の振舞いの収集とその分析

飯田 龍^{1,a)} 光田 航^{1,b)} 徳永 健伸^{1,c)}

概要: テキストアノテーション時の作業者の振舞いは作業者が文章を理解するための認知的な処理を反映していると考えられるため、この振舞いを分析することで深い理解が必要となる言語処理に役立つ情報を明らかにできる可能性がある。そこで本研究では、この分析の一例として、日本語の述語項構造アノテーション時の作業者の振舞いを収集したデータをいくつかの観点から分析する。この結果、人間のアノテーションの振舞いを分析することで得られる知見が機械学習に基づく手法で用いる素性を検討する際に役立つことを示す。

1. はじめに

アノテーションを行い、その結果に従い学習を行うというアプローチは、自然言語処理の浅い処理から深い言語理解が必要となる処理までさまざまな自然言語処理で広く適用されている。このため、テキストアノテーションは教師あり手法を用いる上で必須となる課題である。品詞付与や名詞句チャンキングのような浅い処理では、単語や品詞などの表層的な情報が分類に有効な素性として利用できるため、教師あり手法では高い性能を得ている。一方、共参照解析や談話関係解析などの深い解析では、表層的な手がかりに加えて、より深い言語的な知識が必要となるが、それらを素性として利用する方法は自明ではない。既存研究では、WordNet[14]のような人手で構築された言語資源やセンタリング理論[7]などの言語理論に基づいた深い言語知識を解析モデルに統合することで、各問題の解析精度は向上しているが、応用処理で利用可能なほどの精度が得られているわけでない。このような問題に対して、言語資源の利用方法や参照すべき手がかりの検討など、さまざまな観点から改善を行う余地がある[3], [12], [15], [16]。

そこで、本研究では、経験的に素性を設計する既存の方法論とは異なり、解析モデルの性能の向上に直結する認知科学的なアプローチを採用し、既存の言語資源や言語理論を用いる代わりに、アノテーション時の作業者の振舞いに着目し、深い言語処理に有益な情報を明らかにすることを目的とする。本研究では、特にアノテーション時

の作業者の視線の動きとアノテーションツールの操作に着目する。近年、視線追跡に関する技術が発展したことにより、視線データが心理言語学や問題解決などのさまざまな研究分野で広く利用されている[2]。また、視線の動きと言語理解・生成の関係に関しても多くの研究が存在する[6], [17]。一方、それらの研究と比較して、一般的な問題解決の問題設定における視線の役割を調査する研究は多くはない[1], [18], [23]。我々の興味の対象であるテキストアノテーションは言語理解の問題であるとともに問題解決の課題ともみなすことができるため、我々の問題を定義する際に両方の研究を参照する必要がある。このようなアプローチでアノテーション作業者の視線データを分析し、問題解決に役立つ有益な情報を発見することを目指す。

本稿では、2節で日本語述語項構造アノテーション時の作業者の振舞いに関するデータを収集するための実験と、その結果得られたデータの詳細について紹介する。次に、3節で1事例に関するアノテーションの処理の構造を説明し、4節で、人間のアノテーションの処理に関する詳細な分析を示し、その分析結果の利用方法について提案する。5節で関連研究を概観し、6節でまとめと今後の課題について議論する。

2. 視線データの収集

2.1 アノテーション課題の設計

アノテーション時の作業者の操作とその際の視線情報を収集するために、日本語書き言葉を対象とした述語項構造のアノテーション作業を行う。この作業では作業者のアノテーション作業を容易にするために、文章中の述語と項候補の範囲（セグメント）は事前にアノテーションされており、作業者は述語-項関係（ガ格、ヲ格、ニ格）をアノテ

¹ 東京工業大学
Tokyo Institute of Technology

^{a)} ryu-i@cl.cs.titech.ac.jp

^{b)} mitsuda@cl.cs.titech.ac.jp

^{c)} take@cl.cs.titech.ac.jp

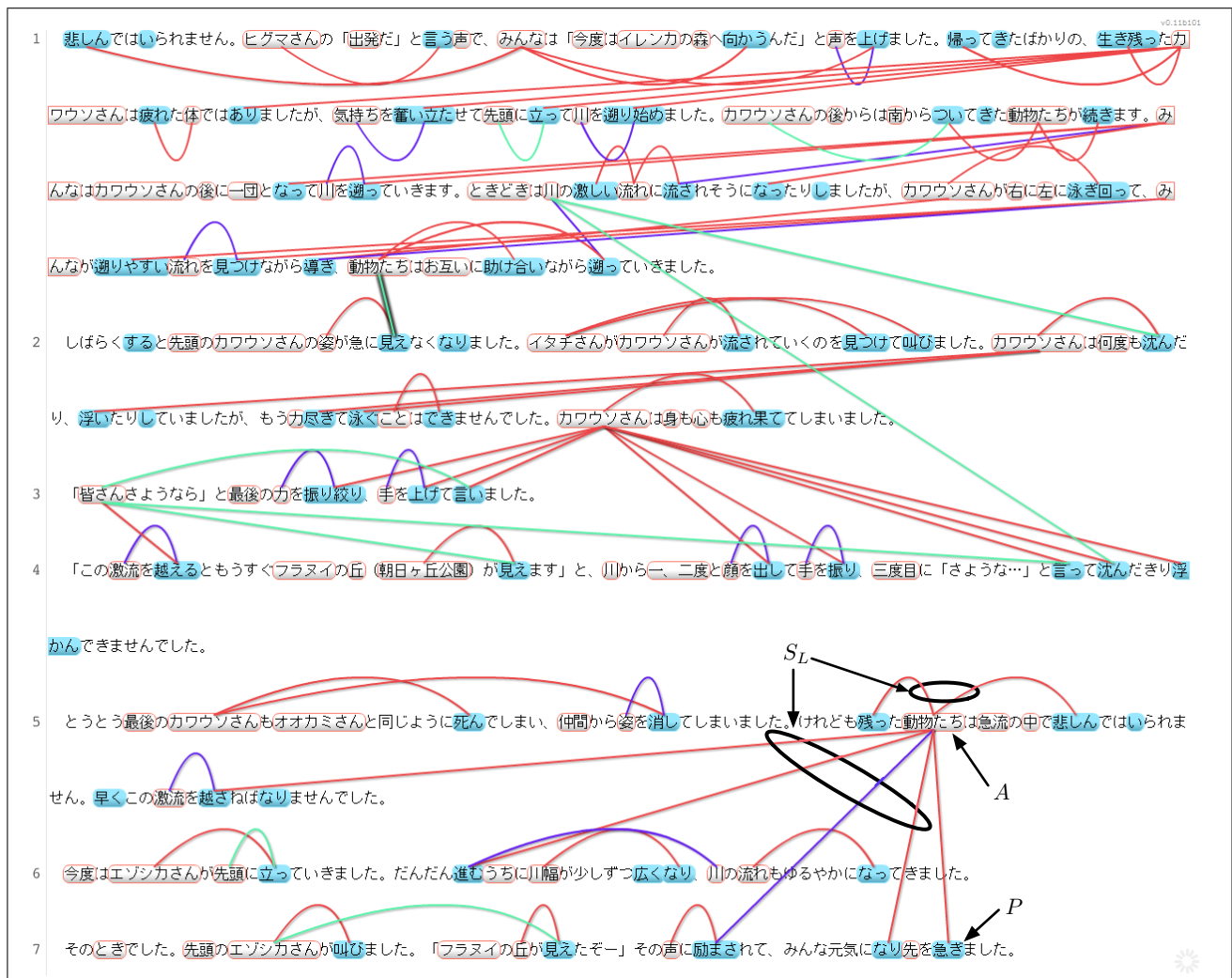


図1 アノテーションツール Slate のスナップショット

シオンするために、述語セグメントと項候補セグメント間に関係（リンク）を付与する。日本語では述語の項が頻繁に省略されるため、その省略された項（ゼロ代名詞）の先行詞を付与する問題もこのアノテーションに含まれる。また、述語や項候補のセグメントは係り受け解析器の出力に基づいてアノテーションされているため、述語（項候補）に対して対応する項候補（述語）が存在しない場合もあり、その場合に誤った関係をアノテーションしないことも必要になる。

アノテーション作業者の操作履歴を収集するために、アノテーションツール Slate[10] を使用した。Slate では簡単なマウスとキーボードの操作で述語と項のリンクを付与できる。ただし、現在公開されている版の Slate^{*1} には操作履歴を記録する機能がないため、表1に示すアノテーション操作に関するイベントを記録できるようにツールの修正を行った。また、Slate にはタグ選択やアノテーションされた結果を表示する領域があるが、これらを表示することでこの領域にも視線が動き、有益な視線の動きを収集すること

表1 記録するアノテーションのイベント

ラベル	説明
create_link_start	リンク付与を開始（マウスドラッグ開始）
create_link_end	リンク付与を終了（マウスドラッグ終了）
select_link	リンクを選択
delete_link	リンクを削除
select_segment	セグメントを選択
select_tag	タグ（付与するリンクの種類）を選択
annotation_start	アノテーションを開始
annotation_end	アノテーションを終了

が困難になる。そこで、図1のようにアノテーション対象の文章のみを画面上に表示するよう修正した。図1はアノテーションされた結果を表している。図1では、述語のセグメントは薄青色の長方形で、項候補のセグメントは赤枠の長方形で表示され、また、アノテーションされたタグ、二格のリンクはそれぞれ赤、青、緑の線で表示されている。

アノテーション作業者の視線は視線計測装置 Tobii T60 を用いて記録した。Tobii T60 のディスプレイサイズは17インチ、解像度は1,280×1,024であり、ディスプレイと作

^{*1} <https://bitbucket.org/dainkaplan/slate/>

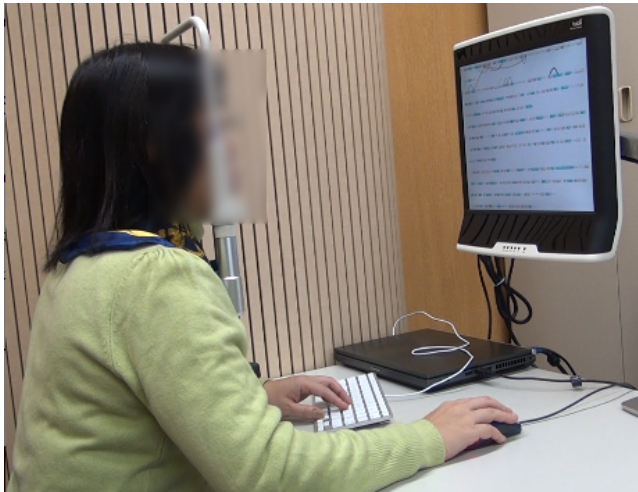


図2 Tobii T60 を用いたアノテーション作業のスナップショット

業者の間の距離は約 50cm になるよう調整した。アノテーションを行う前にはキャリブレーションを行い、また、アノテーション作業時には視線検出のエラーを抑制するために、図 2 のように顎台を用い、作業者の頭の動きを固定した状態でデータ収集を行った。

このデータ収集実験では、述語項構造アノテーションの経験のある 3 人のアノテーション作業者を雇用した。各アノテーション作業者は現代日本語書き言葉均衡コーパス (BCCWJ) の書籍コーパス (PB) から選択した 43 記事²を対象にアノテーション作業を行う。また、作業中に画面のスクロールが起こった場合、収集した視線情報と画面中の文字との対応付けが困難になるため、画面に表示される文字数を約 1,000 文字に制限し、画面のスクロールが起こらないようにした。アノテーション作業者は記事単位で作業を行い、記事全体への作業が終了した後は必要に応じていつでも休憩をとることができる。休憩後の作業開始時には毎回キャリブレーションを行った。

2.2 データ収集の結果

述語と項の間のリンク付与の結果、各作業者 (A_0 , A_1 , A_2) が付与したリンク数はそれぞれ 3,353 (A_0), 3,764 (A_1), 3,462 (A_2) となった。ただし、ある述語に対して複数の同じ関係が付けられるなど、以降の分析に不適合な事例が含まれたため、それらを除外したリンク付与数は 3,054 (A_0), 3,251 (A_1), 2,996 (A_2) となる。さらに、我々が行う分析では述語とその項の両方が注視されている必要があるため、最終的に以降の分析で扱うリンク付与数は 1,776 (A_0), 1,430 (A_1), 1,795 (A_2) となり、それぞれの作業者の作業結果の約半分となった。各作業者が付与した格関係

表 2 各作業者による作業結果 (述語、項ともに注視された場合のみ)

格	A_0	A_1	A_2	合計
ガ格	1,170	904	1,105	3,179
ヲ格	383	298	421	1,102
ニ格	223	228	269	720
合計	1,776	1,430	1,795	5,001

の数の詳細を表 2 にまとめる。この表にまとめた全リンク付与数 5,001 事例を以降の分析で利用する。

3. アノテーション作業者の振舞いの分析

収集したデータ内のアノテーション作業者の振舞いを分析した結果、述語-項関係をアノテーションする際のアノテーション過程は以下の 3 つの段階に分解できることがわかった。

- (1) アノテーション作業者が与えられた文章を読み、文脈を理解する。
- (2) 作業者はアノテーション対象となる述語を注視した後、述語との関係 (例えば、ガ格) を考えながら、述語より前方に出現した項候補の中から項を探索する。
- (3) 作業者は正しい対応関係があると考えられる項をみつけた後、述語-項関係を確認するために周辺文脈を見回す。この確認作業は述語とその項の間にリンクを作成することで終了する。

上述の (1)~(3) の述語を注視した後に項を探索するアノテーションの戦略は、述語が項を下位範疇化する言語的な知識を反映していると考えられる。加えて、日本語が主辞終端型の言語であるため、述語は基本的にはその項の後に出現する。そのため、この戦略を採用する場合、文内に出現する項を探索するときには、述語の位置から文頭のほうへ探索することになる。

ここで示したアノテーション過程の分割のように、ある認知過程をいくつかの機能的な段階に分割するという考え方は認知科学の分野で一般的に用いられている。例えば、Just ら [9] は問題解決の過程を探索段階、比較段階、確認段階の 3 つの段階に分割している。彼らの問題設定では、各面に 1 文字が書かれた 2 つの立方体が描かれた絵が与えられ、被験者はそれに対して 2 つの立方体が同じものであるか否かを決定しなければならない。2 つのうち 1 つの立方体はもとの向きに対してある角度で回転されているため、被験者は頭の中でその立方体を回転させて、2 つが一致するかを判断する必要がある。Russo ら [19] は、与えられた絵などの可視化された情報に基づいて意思決定を行う過程を適応段階、評価段階、確認段階の 3 つに分割している。彼らの実験では、被験者は可視化して表示されたいくつかの日常的な食品の 1 つを選択することを指示される。上述の 3 つの段階の境界は被験者の視線の動きや実験終了を

² 記事の選別では、記事の先頭から 1,000 文字程度抽出する際に、節見出しなどの文章の断絶を引き起す要因が含まれない記事のみを抽出し、さらにその記事集合の一部に対して第一著者もしくは第二著者があらかじめアノテーションを行い、局所文脈だけを参照することでアノテーション可能な事例が頻出する記事を除外したものをランダムに 43 記事を選択した。

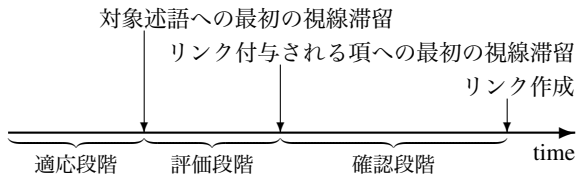


図3 アノテーションの過程の分割

宣言する発話に基づいて決定される。Malcolm ら [13] は、可視化された内容を探索する過程を初期化段階、走査段階、確認段階に分割するという考えを採用している。また、Gidlöf ら [5] は意思決定と可視化された内容の探索過程の2つを対象に、内容の探索過程を分割するという観点から議論している。

このように、上述の研究はそれぞれ異なる認知過程を扱っているが、最初の段階で問題の概観を捉え、次に一時的な解を求め、最後にその解を確認するという部分は共通している。そこで、我々が対象とするアノテーションの過程に関してもこの考え方に従う。特に、我々が対象とする課題は Russo ら [19] が対象とした意思決定の課題に類似しているため、基本的には彼らの考え方を採用する。ただし、彼らが適応段階と呼ぶ問題解決の最初の段階と比較した場合、一つの述語に対するアノテーションの概観を捉える開始位置が不明瞭である。これは、2節で述べたように、我々のデータ収集の実験設定では一つの記事をアノテーションする間に複数の述語に対する意思決定（リンク作成作業）が含まれるため、最初の適応段階の後に複数の評価段階と確認段階が存在するためである。さらに、意思決定の既存研究では一つの対象を選択する課題であるのに対し、本研究で対象にする述語項構造では述語とその項という2つの対象の間の関係を扱うという違いも存在する。

これらの違いや既存手法 [5], [19] の考え方を反映し、次のようにアノテーションの過程を3つの段階に分割した。まず、前述したように、適応段階の始点を決定することは困難であるため、アノテーション対象となる述語への最初の視線滞留^{*3}の開始点が起こった時刻を適応段階の終点とする。次の評価段階の始点はアノテーション対象の述語への最初の視線滞留の開始点とし、終点は最終的にリンク関係がアノテーションされる項への最初の視線滞留の開始点とする。最後の確認段階はリンク付与された項への最初の視線滞留の開始点に始まり、リンクが付与された時点で終了する。これらの定義と各段階の関係を図3にまとめる。

各段階の境界の時刻は記録した視線データとアノテーションツールで記録した操作履歴に基づいて決定する。また、Dispersion-Threshold Identification (I-DT) アルゴリズム [20] に従って収集した視線情報から注視となる視線の動きを決定する。I-DT アルゴリズムでは、視線の速度が

^{*3} 本研究では、セグメントへの連続した一つ以上の注視を一つの視線滞留と定義する。

注視の近傍では遅くなるという根拠に基づいて、特定の分散の範囲内で連続して出現する視線の集合を注視として同定する。この注視を決定するために視線間の最大の距離と注視の最小の時間間隔という2つの閾値が存在するが、(i) ディスプレイサイズとその解像度、(ii) アノテーション作業者とディスプレイとの距離、(iii) 視線計測装置の解像度という3つの実験設定に基づいて、分散の閾値を16ピクセル、Richardson ら [17] の研究に従って時間間隔の閾値を100 msec に設定した。この注視に基づいて、2つの連続する注視が100 msec 未満で隣接するような同一のセグメントへの注視の集合を一つの視線滞留とする。また、セグメントへの視線滞留を同定する際には、水平方向への視線の誤りを考慮し16ピクセル（一文字分）のずれを許容する。また、リンク作成の時刻を決定する際には表1の“create_link_start”が記録された時間を参照する。

図3に示した3つの段階のうち、評価段階でアノテーション作業者がアノテーション対象となる述語の項を同定するため、この段階の分析を行うことが述語項構造解析に役立つ素性を検討する上で有益だと考えられる。しかし、リンク付与される項は適応段階で読み進める際にも考慮されるため、評価段階で読み返すことで述語とその項の関係を判断する場合もあれば、作業者の短期記憶に基づいて判断される可能性もあるため、評価段階を視線情報に基づいて分析してもどのように項を探索しているかを調査することは難しい。また、前述したように、適応段階については厳密な時間間隔を定義することが困難であるため、適応段階も含めた分析を行うことも困難であると考えられる。

一方、確認段階ではリンク付与すべき項が考慮された状態で競合する項候補を調査することで、最終的にアノテーションするリンク関係を確認する。つまり、この段階では考慮に値する競合する項候補が明示的に注視されることが期待される。また、この確認段階は操作記録と視線情報からその期間を定義できるため、収集したデータ中のアノテーション作業者の確認段階における振舞いが分析できる。このような理由により、今後の分析対象を確認段階の分析に限定し、この段階における作業者の振舞いをいくつかの観点から分析する。

4. 確認段階の振舞いの分析

アノテーションされた述語-項関係は、その関係をアノテーションする際のアノテーション作業者の振舞いや述語と項の出現位置に応じて明示的に分類できると考えられる。そこで、アノテーションされた事例（述語、項、格の3つ組）をいくつかの観点であらかじめ分類し、その分類に基づいて分析を行う。事例を分類するために、まず確認段階で起こる視線の動きに着目する。確認段階で起こる視線の動きは典型的に大きく2通りに分類できる。一つは、確認段階で述語-項関係を決定する際に項となるセグメントを注

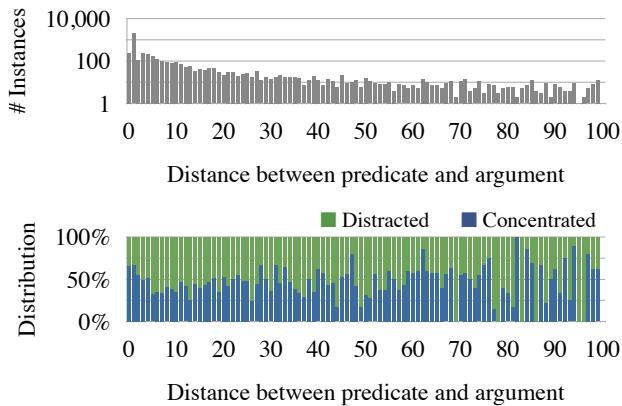


図4 述語と項の距離

視してそのセグメントを認識し、他の対立する項候補を注視することなくリンク作成を行う場合ある。もう一つはアノテーションする述語-項関係を確認するために、他の対立する項候補や文脈を見回すという視線の動きである。本研究では、前者を Concentrated, 後者を Distracted というラベルで区別し、またこの分類を Concentrated/Distracted 分類と呼ぶ。

さらに、我々は述語と項の距離によってアノテーション作業に違いが生じると考えた。つまり、項が述語の近くに出現する場合はその項がその文脈で表す意味や項の出現位置を把握できているため、アノテーションが容易であると考えられるのに対し、項と述語の距離が遠い場合は必ずしもその項が文脈中で表す意味や出現位置を把握できているとは限らないため、その項を特定するためになんらかの視線の動きが起こることが期待できる。そこで、この違いを調査するために、述語と項の距離が近い場合を Near, 遠い場合を Far というラベルで区別し、それぞれの振舞いに関して調査を行う。この分類を以後、Near/Far 分類と呼ぶ。

まず、Near/Far 分類を決定するために、アノテーションされた述語とその項の距離の分布を調査した。結果を図4の上部のグラフに示す。このグラフでは、横軸が文字単位の距離、縦軸が各距離で出現したアノテーション事例の頻度を表している。図4より、事例は距離が1の位置に集中していることがわかる。これは項の後に1文字の格助詞が出現し、その直後に述語が出現するという事例が頻出することを反映している。図4の下部のグラフは各距離における Distracted と Concentrated にそれぞれ分類される事例の割合を表している。この分布から、距離と Distracted/Concentrated 分類の間に明確な関係が無いことがわかる。これらの間の相関係数を調査したところ、 -0.26 という結果を得た。この結果より、述語と項の距離は Distracted/Concentrated 分類に影響を及ぼさないことがわかる。そのため、図4に示された距離の傾向を考慮し、Near/Far 分類に関する閾値を全事例の平均距離である 22

表3 格ごとの Near と Far の分布

格	Near	Far	合計
ガ格	2,201 (0.44)	978 (0.90)	3,179 (0.64)
ヲ格	1,042 (0.34)	60 (0.05)	1,102 (0.22)
ニ格	662 (0.22)	58 (0.05)	720 (0.14)

表4 4つのカテゴリに関する項の分布

格	NC	ND	FC	FD
ガ格	1,172 (0.53)	1,029 (0.60)	504 (0.91)	474 (0.88)
ヲ格・ニ格	1,024 (0.47)	680 (0.40)	52 (0.09)	66 (0.12)
合計	2,196 (1.00)	1,709 (1.00)	556 (1.00)	540 (1.00)

に定めた。つまり、ある事例の項と述語の距離が 22 未満なら Near に分類し、それ以外なら Far に分類する。

表2に示した 5,001 事例を上記の 2 分類を組み合わせた 4 分類 (Near-Concentrated (NC), Near-Distracted (ND), Far-Concentrated (FC), Far-Distracted (FD)) に分け、次節に示す 3 つの観点から分析を行う。

4.1 述語-項の距離と格の関係

日本語では、典型的にガ格が省略されやすいなど、格ごとに異なった傾向が存在するため、アノテーション作業が項の格に応じて異なるアノテーションの振舞いをしていると仮定した。この調査のために、まず Near/Far 分類と格の関係を分析する。Near もしくは Far の事例がどのような格をとまって出現しているかを調査した結果を表3にまとめる。表3に示した結果から、Near と Far で異なった事例の出現傾向があることがわかる。まず、述語と項が近くに出現する場合、アノテーションされた事例は 3 つの格に分散していることがわかる。一方、述語と項が離れている場合には事例がガ格に集中していることがわかる。これは、ガ格が文頭に出現する、もしくは格要素が頻繁に省略されるという傾向を反映していると考えられる。我々のアノテーションの仕様では、省略された格要素 (ゼロ代名詞) に対して述語-項関係をアノテーションする場合、前方文脈に出現する先行詞にリンクを付与するため、省略される場合はより距離が離れた位置に出現する項にアノテーションされることになる。一方、ヲ格やニ格の項は、ガ格と比較して省略されない傾向にあるため、述語と項が離れた位置に出現する割合が低くなる。この出現傾向の違いから、述語項構造解析を行う際にそれぞれの格ごとに個別の扱いが必要になることがわかる。例えば、ヲ格やニ格を解析する場合には述語の近傍のみを探索して項を決定すればよいのに対し、ガ格に対しては前方の広い文脈から項を探索する必要がある。

次に、Concentrated/Distracted 分類が格ごとにどう影響するかを調査するために、表3の各頻度を Concentrated/Distracted 分類に関して細分化した結果を表4に示す。表4の NC と ND を比較、もしくは FC と FD を比較

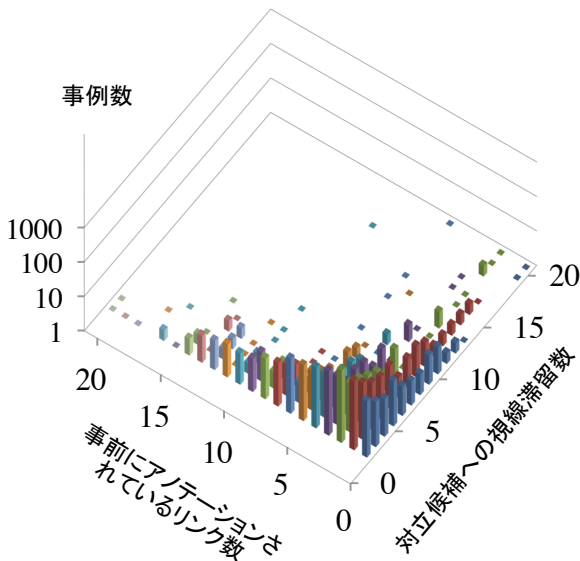


図5 対立候補への視線滞留数と事前アノテーションされたリンク数の関係

してわかるように、Concentrated/Distracted の分類が項の格に関する分布にほとんど影響していないことがわかる。直観的にはガ格が近傍に出現する Near の場合には、その項を作業者が記憶している可能性があるため、評価段階で容易に述語-項関係を認定できると考えられるのに対し、述語と項が離れている Far の場合は、述語-項関係を同定するために確認作業が必要となるため、項の近傍文脈の確認のため Distracted の割合が増えることが期待される。しかし、表4からわかるように、ガ格の NC と ND、FC と FD の割合がほとんど同じであり、項が離れた位置に出現している場合でも近い場合と同程度に確認が容易に完了していることがわかる。この点について次節で別の観点から分析する。

4.2 事前アノテーションされたリンクの影響

Concentrated に分類される事例をアノテーションする際、作業者は対立する項候補を調べることなく着目した項の正しさを判定することができているが、この要因の一つに、我々が使用したアノテーションツールの表示方法が影響していると考えられる。図1に示したアノテーション時の画面では、すでにアノテーションしたリンクが継続して表示されている。このように、リンクが可視化されていることで、作業者がこれまでの作業で頻繁にアノテーションした項は作業者の短期記憶の中で視覚的もしくは認知的に顕現性が高くなっており、その結果、その項を以降の作業でアノテーションするには確認作業が容易に完了すると考えられる。例えば、図1に示した状況で述語 *P* に対して項 *A* をアノテーションする状況を考えると、この状況では項 *A* はすでに S_L に示される6つのリンクがアノテーションされており、この結果、リンクがまったくアノテーショ

ンされていない場合と比較して、容易に *P* と *A* の述語-項関係を把握できると考えられる。

事前にリンクがアノテーションされていることで確認作業が容易に行われることになるという仮説について定量的に調査するために、アノテーション対象となる項に事前にアノテーションされているリンクの数と対立候補に対する視線滞留数の関係を調査した。結果を図5に示す。ただし、この分析では、Near に分類された事例は事前にアノテーションされたリンク数が少ないと考えられるため、Far の事例のみを対象に分析を行った。図5に示されたグラフは、対立候補への滞留数とリンク数が少ない箇所に頻度最大の点が出現している3次元の勾配を表している。このグラフは対立候補の滞留数が少ない場合とリンク数が少ない場合でほぼ対称になっており、事前にアノテーションされているリンク数が多いほど、対立候補への視線滞留数が減り、また逆に、視線滞留数が増えたときにはアノテーションされているリンクの数が少ないことがわかる。この結果から、可視化されたリンクの影響による視覚的もしくは認知的な顕現性がアノテーション時の作業者の認知負荷を下げ、その結果、効率的に正しい項を確認する作業に貢献していることがわかる。

この結果は事前にアノテーションされた項候補へのリンク数とその顕現性を反映していることを示しているため、この傾向を反映して、例えば、より多くの照応もしくは共参照のリンクを持つ項候補を優先的に項とすべきか否か判断するといった特定の処理を考えることができる。また、本研究では確認段階における分析しか行っていないが、評価段階、つまり項を探索する段階においても同様の影響があると考えられる。このため、このリンク数の情報を学習ベースの手法の素性として導入することで述語項構造解析モデルの性能が向上することが期待できる。

4.3 項の特定性と視線の分散の関係

日本語は主辞終端型の言語であるため、統語的な主辞となる名詞は常に名詞句の最右に出現する。このため、日本語を対象に述語項構造をアノテーションしたコーパス[8],[11]では、述語-項関係をアノテーションする際に項となる名詞句の開始位置を明示的に定めず、その名詞句の主辞を含む文節内の名詞句（もしくは主辞となる形態素）に対し述語-項関係をアノテーションしている。また、この特徴は日本語を対象に格フレーム辞書を自動構築する場合にも採用されており、コーパス中の〈（名詞句の主辞となる）語、格、動詞〉の3つ組を利用して辞書構築が行われている[21]。このため、我々が行った視線データ収集のためのアノテーション実験でも項のセグメントに関して同様の基準を採用した。

しかし、項の主辞となる名詞が必ずしも十分な情報を提示するとは限らない。例えば、主辞の位置に出現する「こ

表5 項の修飾要素と視線の拡散の関係

	修飾要素なし	修飾要素あり		合計
		NP のみ注視	NP の外も注視	
C	1,562	1190	—	2,752
D	1,168	242	839	2,249

C, D はそれぞれ Concentrated, Distracted を表す。

と」や「の」などの非自立名詞はその名詞句の意味を特定するための情報を持たない。例えば、「勉強すること」という名詞句における「こと」は動詞句を名詞化させるだけで、その表現だけでは文脈中で意味をなさない。そのような場合、確認段階で項とすべきか否かを決定するために、その名詞句の修飾要素を含めた名詞句全体を吟味する必要があると考えられる。そこで、実際にアノテーション作業者がそのように振る舞っているのかを調査するために、収集したデータに基づいて分析を行った。

この分析では、形態素・係り受け解析の結果に基づき、アノテーションされた項が名詞句の一部であり、かつ、他に修飾要素がある場合とそれ以外に分類する。この分類はそれぞれ表5中の「修飾要素あり」と「修飾要素なし」に対応する。このうち、「修飾要素あり」に分類された事例をさらにアノテーション作業者の視線の動きに基づいて「NPのみ注視」と「NPの外も注視」の2つのクラスに分類する。前者の「NPのみ注視」は確認段階でアノテーションされた項のセグメントを含む名詞句のみを注視する場合であり、後者の「NPの外も注視」は名詞句の外側の他のセグメントも注視する場合である。ただし、前述したように、我々のアノテーションの仕様では、述語とその項の主辞の間にリンクを作成する。このため、Distractedの事例の注視は名詞句の中にのみ出現する可能性があるため、必ずしも「NPの外も注視」に分類されるとは限らない点に注意されたい。表5に各分類の分布を示す。

表5より、ConcentratedとDistractedがほぼ同程度の頻度で出現していることがわかる ($2,752 : 2,249 = 0.55 : 0.45$)。この観点から見れば、ConcentratedとDistractedの事例は述語項構造解析で同等に扱うべきであるが、このうち「修飾要素あり」に着目して見た場合、22%のDistractedの事例が「NPのみ注視」に分類されていることがわかる。つまり、確認段階で修飾要素を持つ名詞句を項として認定する場合にはその修飾要素のみを参照することで確認が完了する事例が修飾要素を持つ項の約22%存在することになる。この数値は無視できない値であり、今後は項候補が修飾要素を持つ場合とそれ以外で異なった処理を考えるといった方向性も考えられる。例えば、主辞の情報に加えて、その修飾要素の情報を項候補をより特徴付けるための情報として導入することなどが考えられる。

5. 関連研究

近年の視線追跡に関する技術が発展したことによって、

視線データはさまざまな研究領域で利用されている [2]。例えば、Bednarikら [1] はプログラマがデバッグを行う際の視線データを収集している。彼らはプログラムの統合開発環境 (IDE) の領域に基づき、プログラマの3つの関心領域 (ソースコードの領域、クラス関係が可視化された領域、プログラムの出力の領域) を定義し、初心者プログラマと玄人のプログラマが関心領域に関してどのように視線を遷移させるかを比較している。ただし、彼らが定義した関心領域は画面を3分割するだけの荒い分割であるため、プログラマの技能の推定には利用可能かもしれないが、得られた玄人の遷移パターンに基づき、ある特定の遷移パターンとプログラミングの技能があることの関連性を説明することは難しいと考えられる。同様に、言語に関する視線の動きから分析を行う際も、より細かい関心領域に関する調査が必要となるため、本研究では文字レベルのより細かい関心領域を採用し、それによってどのセグメントを注視したかの調査を行った。

Rosengrant [18] は、被験者の視線データとプロトコル分析 [4] で導入されている被験者の口述説明を統合する gaze scribing と呼ばれる新しい分析手法を提案している。この研究では、素人と玄人の問題解決時の戦略の違いを発見するために、事例研究としてディスプレイ上に表示された物理の電気回路の問題を解く際の被験者の振舞いを分析している。彼は他の問題へ gaze scribing を適用し分析することの重要性を強調しているが、プロトコル分析を単純に用いることは被験者の認知負荷が高まり、想定していた作業への影響が懸念されるため、もともと達成すべき目標を阻害しないよう課題を設計することが重要となる。

Tomanekら [23] は能動学習の効率的な事例選択のために、固有名のアノテーションの難易度を推定するために視線データを利用している。彼らは固有名の特徴を制御することでさまざまな設定におけるアノテーション時の視線データを収集している。彼らの固有名に関する視線データの収集では、アノテーション対象となる表現を見ているか、もしくはその周りの文脈を見ているかという荒い関心領域を扱っているが、本研究で扱う述語項構造アノテーションの場合、正解となる項に対して競合する項の候補が出現しているか否かという分析が必要になるため、より細かい関心領域を扱う必要があるという違いが存在する。

我々の先行研究では2節で紹介したデータを利用して述語項構造アノテーションの難易度推定について調査を行った [22], [24]。人手による分析結果に基づき、経験的に各インスタンスに関する作業時間を定義し、その作業時間が短いほど複数人が作業した結果の一致率が高くなることを示した。

6. おわりに

本稿では、自然言語処理の問題、特に述語項構造解析に

必要となる知見を明らかにするために、アノテーション作業者の振舞いを分析した結果について報告した。本研究では、述語-項関係の各事例に対するアノテーション作業（リンク付与など）の履歴とその際の視線の動きを収集したデータを対象に、(i) 述語と項の距離と格の関係、(ii) 既出リンクのアノテーション作業への影響、(iii) 項の特定性と視線の分散に関する関係という3つの観点から人手による分析を行った。それぞれの分析の結果より、人手によるアノテーション行為の分析が述語項構造解析の問題を解く上で必要となる特徴を明らかにするために役立つことを示した。

今後の課題として、まず本研究で収集した作業者の振舞いのデータをさらに分析することが考えられる。本研究では、分析者の内省に基づき、分析する観点を定めたが、これに対し、テキストマイニング技術などを用いて作業者の振舞いからアノテーションに顕著に出現する振舞いを発見するという方向性もありうる。このため、今後はマイニング技術を用いて言語処理の問題に有益な視線のパターンなどを発見することを考えている。

また、本研究で調査した結果得られた知見を実際に機械学習に基づく述語項構造解析のモデルに導入し、その有効性を定量的に評価することも重要だと考えている。

本研究では、述語とその項を注視した後の確認段階における作業者の振舞いを分析したが、その前の段階に相当する評価段階を分析することで、項を探索するための手がかりをより直接的に分析できる可能性がある。ただし、この研究で収集したデータはその分析に適用することが困難であるため、この特徴を捉えることに特化した実験設定を考える必要がある。例えば、画面に表示した一つの述語に対して項を探索する際の視線を収集するといった実験設定が考えられる。今後はこのような異なる実験設定においてデータを収集するとともに、プロトコル分析[4]なども併用することで、より直接的にアノテーション作業者の認知過程を顕在化させることを行う予定である。

さらに、今後の方向性として、さまざまな言語処理の問題に関するアノテーションを通じて作業者の多様な振舞いを収集し、それを分析することが必要になる。このため、どのようにアノテーション作業者の振舞いを収集するかについての方法論を構築する必要がある。言語処理の問題を横断的に収集することで、この方法論を確立させることも今後深い自然言語処理を達成する上で重要となると考えられる。

参考文献

- [1] Bednarik, R. and Tukiainen, M.: Temporal eye-tracking data: Evolution of debugging strategies with multiple representations, *Proceedings of the 2008 symposium on Eye tracking research & applications (ETRA '08)*, pp. 99–102 (online), DOI: 10.1145/1344471.1344497 (2008).
- [2] Duchowski, A. T.: A breadth-first survey of eye-tracking applications, *Behavior Research Methods, Instruments, and Computers*, Vol. 34, No. 4, pp. 455–470 (2002).
- [3] duVerle, D. and Prendinger, H.: A Novel Discourse Parser Based on Support Vector Machine Classification, *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. 665–673 (2009).
- [4] Ericsson, K. and Simon, H. A.: *Protocol Analysis – Verbal Reports as Data* –, The MIT Press (1984).
- [5] Gidlöf, K., Wallin, A., Dewhurst, R. and Holmqvist, K.: Using eye tracking to trace a cognitive process: Gaze behaviour during decision making in a natural environment, *Journal of Eye Movement Research*, Vol. 6, No. 1, pp. 1–14 (2013).
- [6] Griffin, Z. M. and Bock, K.: What the eyes say about speaking, *Psychological Science*, Vol. 11, No. 4, pp. 274–279 (2000).
- [7] Grosz, B. J., Joshi, A. K. and Weinstein, S.: Centering: A Framework for Modeling the Local Coherence of Discourse, *Computational Linguistics*, Vol. 21, No. 2, pp. 203–225 (1995).
- [8] Iida, R., Komachi, M., Inui, K. and Matsumoto, Y.: Annotating a Japanese text corpus with predicate-argument and coreference relations, *Proceeding of the ACL Workshop ‘Linguistic Annotation Workshop’*, pp. 132–139 (2007).
- [9] Just, M. A. and Carpenter, P. A.: Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability, *Psychological Review*, Vol. 92, No. 2, pp. 137–172 (1985).
- [10] Kaplan, D., Iida, R., Nishina, K. and Tokunaga, T.: Slate – A tool for creating and maintaining annotated corpora, *Journal for Language Technology and Computational Linguistics*, Vol. 26, No. 2, pp. 89–101 (2012).
- [11] Kawahara, D., Kurohashi, S. and Hasida, K.: Construction of a Japanese relevance-tagged corpus (in Japanese), *Proceedings of the 8th Annual Meeting of the Association for Natural Language Processing*, pp. 495–498 (2002).
- [12] Lin, Z., Ng, H. and Kan, M.-Y.: A PDTB-styled end-to-end discourse parser, Technical Report TRB8/10, School of Computing, National University of Singapore (2010).
- [13] Malcolm, G. L. and Henderson, J. M.: The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements, *Journal of Vision*, Vol. 9, No. 11, pp. 8:1–13 (online), DOI: doi: 10.1167/9.11.8 (2009).
- [14] Miller, G. A.: WordNet: A Lexical Database for English, *Communications of the ACM*, Vol. 38, pp. 39–41 (1995).
- [15] Ng, V.: Supervised noun phrase coreference research: The first fifteen years, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*, pp. 1396–1411 (2010).
- [16] Pradhan, S., Moschitti, A., Xue, N., Uryupina, O. and Zhang, Y.: CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes, *Joint Conference on EMNLP and CoNLL – Shared Task*, pp. 1–40 (2012).
- [17] Richardson, D. C., Dale, R. and Spivey, M. J.: Eye movements in language and cognition: A brief introduction, *Methods in Cognitive Linguistics* (Gonzalez-Marquez, M., Mittelberg, I., Coulson, S. and Spivey, M. J., eds.), John Benjamins., pp. 323–344 (2007).
- [18] Rosengrant, D.: Gaze scribing in physics problem solving, *Proceedings of the 2010 symposium on Eye tracking research & applications (ETRA '10)*, pp. 45–48 (online), DOI: 10.1145/1743666.1743676 (2010).
- [19] Russo, J. E. and Leclerc, F.: An eye-fixation analysis of choice processes for consumer nondurables, *Journal of Con-*

- sumer Research*, Vol. 21, No. 2, pp. 274–290 (1994).
- [20] Salvucci, D. D. and Goldberg, J. H.: Identifying fixations and saccades in eye-tracking protocols, *Proceedings of the 2000 symposium on Eye tracking research & applications (ETRA '00)*, pp. 71–78 (2000).
 - [21] Sasano, R., Kawahara, D. and Kurohashi, S.: The effect of corpus size on case frame acquisition for discourse analysis, *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2009)*, pp. 521–529 (2009).
 - [22] Tokunaga, T., Iida, R. and Mitsuda, K.: Annotation for annotation - Toward eliciting implicit linguistic knowledge through annotation -, *Proceedings of the 9th Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation (ISA-9)*, pp. 79–83 (2013).
 - [23] Tomanek, K., Hahn, U., Lohmann, S. and Ziegler, J.: A Cognitive cost model of annotations based on eye-tracking data, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*, pp. 1158–1167 (2010).
 - [24] 光田 航, 飯田 龍, 徳永健伸: テキストアノテーションにおける視線と操作履歴の収集と分析, 言語処理学会第 19 回年次大会, pp. 449–452 (2013).