

アノテーション作業者の内省を顕在化するためのデータ収集

飯田龍

徳永健伸

東京工業大学 大学院情報理工学研究科

{ryu-i,take}@cl.cs.titech.ac.jp

1 はじめに

近年、自然言語処理の問題の多くは、まずその問題の出力としてどのような結果を出力すべきかという情報(タグ)をあらかじめ対象となるコーパスにアノテーションし、そのコーパスを訓練・評価用のデータとして利用するという方法に基づき研究が進められている。この研究パラダイムは約20年続いており、形態素・構文解析や語義曖昧性解消、照応解析や談話構造解析など、さまざまな問題に関してこの考え方が適用され、成果をあげてきた。

このアノテーションを行い、その結果を機械学習に利用するという枠組みを採用することで、研究者間で共通の研究基盤となるデータを共有し、個別の解析手法を共通のデータを対象に比較可能となる。さらに、コーパス全体の現象を対象とすることで人手分析が難しい現象が顕在化し、その現象について研究者が分析するきっかけを与えるなど、さまざまな利点がある一方、考えなければならない問題も多い¹。

まず、アノテーション作業者がアノテーションの判断の基準とした言語的な手がかりと解析モデル側で導入される手がかりの齟齬があげられる。これは、作業者が基本的に出力結果として出力したい情報にしかアノテーションを行わないため、なぜそのアノテーションを行ったかという作業者の意図を解析モデルを構築する際に反映できないために起こる。

また、コーパスの構築にはあらかじめ設計したアノテーションの仕様にしたがい予備的な作業を行い、その結果、仕様に記述できていない問題点を洗い出した後に、大規模なデータに対して作業を行うことが多い。また、作業対象の一部に対して複数人でアノテーションの作業を行い、カップ値などに基づく作業の一致率が高いことで、アノテーション作業の品質の良さを示している。ただし、アノテーション作業者の言語能力の高さのためにアノテーションの際に起こる問題が顕

在化していない場合もあり、アノテーションの仕様と作業手順に関する方法論だけでは成果物であるコーパスの良さを客観的に見積ることができない。

また、アノテーション作業者についてもどのような人材を雇用したかについては「大学の学部生」や「言語学者」などの荒いカテゴリでしか言及されておらず、作業個人言語的な資質や問題に関する理解度が把握できないため、同種の人材を揃えても必ずしも同質の作業結果が得られるとは限らないという問題がある。例えば、OntoNotes プロジェクト [3] では9割のアノテーション作業の一致率を目標としており、これが達成されているが、これがアノテーションの仕様・方法論の良さのためなのか作業者の言語的な分析能力が高いためなのかは判断できない。

アノテーション作業者は提示されたデータからさまざまな言語的な特徴を読み取り、その特徴に基づきどのようにアノテーションすべきか(もしくはアノテーションすべきでないか)を考え、最終的にその考えた結果を反映してアノテーションの作業(例えば、あるタグを付けるなど)を行う。ただし、一般的には作業者の最終的な判断に基づく作業の結果しか記録しないため、作業者がどの言語的特徴に着目したか、どの問題に判断のための時間を割いたか、なぜその問題に時間をかけなければならなかったのかといったアノテーションの過程に関する情報は記録されず、この結果作業者のアノテーション作業過程に関する分析ができないという問題が起こる。そこで、本研究ではこの問題を解決するために、アノテーション作業者が行った作業結果に加え、作業者の内省を反映したデータの収集も試みる。具体的には、作業者のアノテーション作業中の視線や操作の履歴を収集し、それを分析することで、作業者がどのような点に着目して(もしくは着目せずに)作業しているかを明らかにする。また、視線や作業の履歴などの客観的な情報は作業者が何を見て判断したか、その判断にどのくらい時間をかけたかといった分析には有効であると考えられるが、必ずしも作業者の判断の全貌を明らかにするものではない。そ

¹この問題について本稿で述べる議論と関係が無い点については言及しないため、捕捉情報として文献 [8] も参照されたい。

こで、より作業者の内省が顕在化するような仕組みとして作業者にどうしてそのアノテーションをしたのかを発話してもらうことによって作業者の作業意図を記録する。さらに、各データ収集方法で集めた情報の利用法として、作業者の質の客観的な評価や素人の作業者の学習などについて検討する。

本稿では、まず2節でそれぞれのデータ収集の方法について説明し、3節で収集したデータの利用法についてまとめる。4節で具体的なデータ収集の計画を述べ、5節で関連研究を概観する。最後に6節でまとめる。

2 アノテーション作業者の内省を顕在化するデータ収集

1節で述べたように、本研究では作業者のアノテーション作業時の判断の根拠を収集することを考える。この際、視線や操作履歴は視線追跡装置などを導入することにより、作業者に比較的負荷をかけることなく収集できると考えられる。この視線や操作履歴などは、それをあらためて人手で分析しなければ作業者の内省が理解できないため、アノテーション作業結果とともにこれらの情報収集する方法を**非明示的なデータ収集**と呼ぶ。これに対し、内観法を用いて作業者がなぜそのアノテーションを行ったかを口頭で実験者に伝える方法が考えられる。この方法では、作業の意図が明確になる代わりに作業者のアノテーションの際の負荷が増大することが考えられる。そこで、本研究で提案するデータ収集の手法では、素人の作業者に玄人の作業者が教示するという状況でアノテーションを行うことで、作業者の負荷を軽減することを考える。このデータ収集の方法を非明示的なデータ収集と対比して**明示的なデータ収集**と呼ぶ。以降で、それぞれの詳細について説明する。

2.1 非明示的なデータ収集

客観的なアノテーション作業の指標として、アノテーション作業における作業者の作業履歴や作業時の視線を収集することを考える。作業履歴はアノテーションツールで作業者が行った行為をその行為が行われた時刻とともに記録する。また、作業者が画面のどこを見ているかについては視線追跡装置を用い、作業者の視線を細かい時間単位で記録する。また、作業者の振舞いに関する分析のために作業風景も記録することで、

作業者がマウスを持たずに考え込んでいるために作業を行っていないなどの外見からわかる情報を採取する。

このように、ツールや器具を利用することで作業者の行為のログは容易に記録できる。ただし、単純にマウスの軌跡やキーストロークを記録するだけでは、作業者がどの行為を行ったのかは理解できない。このため、アノテーション作業に関して意味のある行為に関してのみ操作履歴を収集することを考える。例えば、述語項構造のアノテーションを考えると、項や述語となるセグメントを認定するためにマウスで範囲指定する、セグメントにラベルを付与する、述語と項を選択し、その間に関係のラベルを付与するといった行為が重要だと考えられる。このような粒度で作業者の行為を記録することで、これらの行為がどういう順序で、またどのくらいの時間間隔で達成されるかを分析するのに役立つと考えられる。

また、視線の情報に関しても単純に画面のどの位置を見ているかという位置情報として扱うのではなく、例えば、アノテーション対象と同じ文を見ている、作業に関して重要な箇所を見ているなど、収集した視線情報と画面の表示を照らし合わせて解釈することが必要となる。この際、画面上の情報をあまり細かく分類すると、視線追跡装置の抽出精度によっては対応付けることができないといった問題が起こる。このため、作業するタスクと関連するオブジェクトを注視している、もしくはその近傍を注視しているといった近似が必要になる。この点をどのように設計するかについては、これまでに類似する研究が無いため、タスクごとに経験的に決定する必要がある。

2.2 明示的なデータ収集

2.1に示したように、非明示的なデータ収集を行った場合、収集した視線や作業行為の情報は問題の性質に従って解釈する必要があるが、その解釈が常に可能だとは限らない。このため、折角さまざまな器具を活用しデータを収集したとしても、その結果から何も考察できないということがあり得る。そこで、内観法を用い、アノテーション作業者が各アノテーションを行った場合、なぜそのアノテーションを行ったのかも口頭でコメントを残すという別の方法が考えられるが、単純にコメントさせるだけではこちらが意図しない内容をコメントされたり、そもそもアノテーション作業に加えてさらに作業に負荷をかけるため、自然な作業とはならないといった問題が起こる。

この問題を解決するために、作業者を二人用意して

対話をしながら作業を行ってもらふことで、一人でコメントを発言するという不自然な状況を回避することを考える。ただし、この二人の作業者にどのような人材を割り当てるかという作業の環境のデザインが重要な要素となる。例えば、他の心理実験と同様に同性の場合と異性の場合や二人の親しさによって会話の内容が異なることが想定される。加えて、与えられたアノテーション作業の習熟度によってもこれが変わり得る。例えば、作業を行う二人がアノテーション作業に不慣れな場合は相談しながら作業を行うかもしれないが、どちらもその作業に関して慣れている場合は一方が作業を行いながら、他方がその作業を確認するだけでほとんど有益なコメントが得られない可能性がある。さらに、作業者の言語学などの専門的な背景によってもコメントされる粒度が異なることが考えられる。

このように、作業環境の設計が収集されるデータの質に影響するため、これを適切に設計する必要がある。本研究では、その一例として、作業者二名にアノテーションに習熟し、ある程度の言語学的知識を持った人材（以下、*玄人*）とアノテーション作業に不慣れな人材（以下、*素人*）を採用し、対話しながら作業を行ってもらふというアプローチを試みる。さらに、*素人*がアノテーションの作業を指示し、それを*玄人*が操作するという制約をかけることで、*素人*のアノテーションに関する言語的直感を言語化することができる。また、*玄人*が*素人*の作業に疑問を持った点について質問をすることで、*素人*のアノテーションの仕様に関する認識のずれや*玄人*の問題に関する具体的な説明も言語化して記録することができる。このような一方が他方を教示するというデザインにすることで、対象とするアノテーションに関してさまざまな内容が積極的に発話され、効率的に収集できると考えられる。

3 収集したデータの利用法

2節で提案した（非）明示的なデータ収集の過程やその収集結果は、現在のアノテーションに基づく言語処理の課題でさまざまな活用ができる。以下で、この活用例について説明する。

3.1 データに基づく問題分析

まず、提案手法によって蓄積されたデータを対象となる問題の分析に利用することを考える。非明示的なデータ収集の結果にはアノテーション作業者の視線情報を含むが、このうちアノテーションの行為を行った

時刻の近傍ではその判断を行うための手がかりとなる箇所を見ている可能性がある。例えば、述語項構造の関係のアノテーションについては項とならない表現はそもそも注視していない、共参照関係を考えるとき作業者はどのような表現を先行詞の候補として見るのか、2つの文の含意関係を判断する際にどのような特徴に着目するなど、解きたい問題に関連する手がかりに注視している、などがあげられる。このような特徴を作業者のアノテーション時の視線から検出することで、実際に解析モデルを構築する際にどのような言語現象や言語的特徴を考える有益な手がかりとなると考えられる。

また、明示的なデータ収集を行った場合でも、*素人*が*玄人*に出すアノテーションの手順や、*玄人*の質問への*素人*の回答などに解析のための手がかりが含まれている可能性がある。どのような収集の方法を採用すればこれらの有益な情報が得られるかは自明ではなく、今後提案するアノテーション作業を行うことによって経験的に明らかにしていく必要がある。

3.2 アノテーション作業者の能力の客観的評価

現在の自然言語処理におけるアノテーションでは、アノテーションの仕様や方法論の評価は複数人が同一記事に同じ基準で作業を行い、その結果の一致率をカッパ値 [2] などに基づいた一致率で評価を行い、一致率が高いことによってアノテーションの良さを間接的に見積っている。しかし、アノテーション作業の品質はアノテーションの仕様だけでなく、アノテーション作業者の言語分析能力にも依存するため、一致率だけではアノテーションの仕様や方法論の良さを議論できない。さらに、作業者の言語能力については論文中で「*素人*」や「大学の学部生」、「言語学者」など荒い粒度で言及されるのが一般的で、各カテゴリに属する作業者の言語能力を見積ることが難しいため、同じカテゴリに属する人材を雇用しても、雇用された作業者の質に幅があるため、同程度の品質のアノテーション結果が得られるとは限らない。

つまり、アノテーションの評価にはアノテーションの一致率や仕様の質的な良さだけではなく、アノテーション作業者の能力もできるだけ客観的な指標で見積る必要がある。作業者の能力は言語学に関する知識に関する問について回答してもらふことで評価できるが、どのタスクに関してどの知識が必要であるかは自明ではないため、そもそも作業者評価用の問を作成するこ

とが難しいと考えられる。そこで、本研究では 2.1 で提案した非明示的なデータ収集を利用し、ある言語処理のアノテーションの課題において、作業に習熟し、言語的な知識が豊富な人材（玄人）とアノテーションの作業はできるが言語的知識が乏しい人材（素人）でどのようにアノテーション作業に違いがでるか視線やツールの操作履歴の観点から比較する。玄人と素人に関する十分な量の非明示的なデータを収集することで、玄人（もしくは素人）が典型的にどのような行動を行うかを分析でき、その行為に近い作業を行っている人材は玄人（もしくは素人）であると推定できる。ただし、どのような粒度で行為の類似度を取るかは自明ではなく、さらにアノテーションごとに扱うべき行動の粒度が異なると考えられるため、この方法でアノテーション作業の質の評価ができるか否かはいくつかのアノテーションの課題に関して調査が必要となる。

3.3 素人の作業者のアノテーション作業の学習支援

アノテーションの結果を機械学習の訓練事例として利用する際、大規模でかつ質の良い（つまり、アノテーションの誤りを含まない）訓練事例をいかに効率的に作成するかが重要な課題となる。この実現のためには、アノテーション作業者の育成が必須の課題であり、アノテーションに不慣れな、また言語的知識の乏しい人材に対してどのように指導を行うかが問題となる。

アノテーションの作業の初期段階では、アノテーションの仕様作成者がアノテーションの作業者に対して作業の基本的な概要を説明し、次に小規模なデータに対して作業を行わせ、作業結果が安定した時点で作業対象に対して作業を行わせることが一般的であるが、この方法で訓練された作業者は必ずしも作業の内容とそれに必要となる言語的な知識を十分に把握できていないことがある。これは、作業者がバッチで行った作業内容を仕様作成者が作業後に確認するために、作業者が実際の作業時にどのような点で悩み、どのような箇所に時間をかけ、またどのような問題が起こったかを把握できないために起こる。

この問題を解決するために、2.2 で提案した明示的なデータ収集の方法を利用する。この方法を採用し、玄人が素人の指示に従って作業を行っていく過程で、素人の誤ったアノテーションや不明瞭な発言に対して玄人が適切にアドバイスをすることで効率的に素人を訓練できると考えられる。

素人がどの程度訓練されたかを評価する際は、あらか

じめ玄人によって正解がアノテーションされた問題集合に対し、訓練前と訓練後にそれぞれ作業させることで正解率がどう変化するかを調査する。また、素人がその問題集合に対して作業する際のどのくらいの速度で作業ができるようになったかも調査する。この結果、訓練前と訓練後で作業効率が上がり、かつ作業の誤りを減らすことができれば、明示的なデータ収集が作業者の訓練に役立つという証拠となる。

4 具体的なデータ収集の計画

2 節では明示的と非明示的なデータ収集について提案し、3 節ではそれらの具体的な利用法について紹介したが、そもそもアノテーション作業を行いデータ収集を行うには金銭的なコストがともなうため、予備実験としてはデータ収集の方法を限定することが望ましい。そこで、まずは非明示的なデータ収集のみを採用し、コーパスへのアノテーションを試みる。

非明示的なデータ収集を行う場合、言語処理のどのような課題についてアノテーションを行うか、またどのような種類の作業者を雇用するかという 2 軸を考える必要がある。しかし、一度に両方の軸を変動させた場合、その分金銭的なコストやアノテーションの作業時間がかかるため、まずはアノテーションの作業を 1 つに固定し、作業者の種類としてはアノテーションの作業を行ったことがない素人と専門的にアノテーションを仕事として行っている玄人という 2 種類の作業者のタイプでそれぞれ作業を行わせ、作業者間で作業内容に違いが出るかを調査する。

また、アノテーションの課題が簡単すぎる場合には素人と玄人で作業内容に差が出ない可能性があるため、適度に複雑なものであるほうが望ましい。そこで、本研究では文章中の述語項構造と共参照関係のアノテーションを同時に行う作業を採用する。このタスクでは、共参照関係のアノテーションにおける先行詞の探索や、述語項構造関係のアノテーションで必要となるゼロ照応関係の検出において、照応詞（ゼロ代名詞）が出現した場合に作業者がこれまでにどのように前方文脈を読み進めてきたかが大きく影響するため、この読み方や先行詞の探索方法に関して素人と玄人で作業過程に関して大きな違いが出ると考えられる。そこで、素人と玄人の作業者をそれぞれ複数人雇用し、同一の作業データに作業してもらうことで作業性能の違いを調査する。

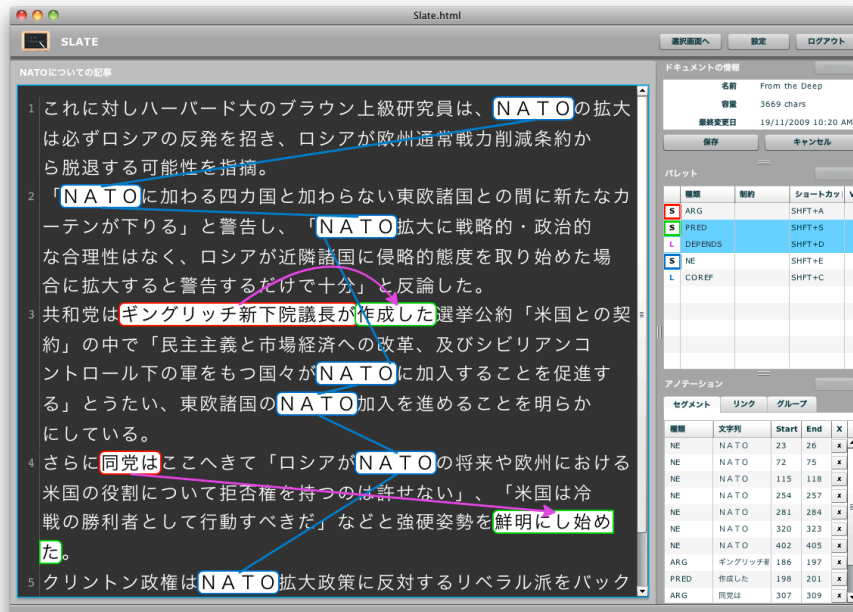


図 1: Slate のアノテーション作業画面の例

4.1 述語項構造・共参照関係アノテーション

述語項構造と共参照関係のアノテーションにおいて多様な言語現象の中で述語と項や共参照の関係が出現したほうが、より素人と玄人で違いが顕著になると考えられる。そこで、本研究ではさまざまな種類の記事が含まれる現代日本語書き言葉コーパス（以下、BCCWJ）に対してアノテーションを行う。述語項構造と共参照関係のアノテーションの仕様としては NAIST テキストコーパス [9] の仕様に準拠して作業を行ってもらう。このため、動詞や形容詞などの述語だけではなく、文章中で事態を表すサ変名詞にも項構造をアノテーションする必要がある、比較的複雑な作業となる²。

4.2 アノテーションツール

アノテーション作業ではアノテーションツールとして Slate[4] を利用することで、名詞句や述語に関するセグメンテーション、またそれらの間の述語と項の関係を直感的な操作でアノテーション可能である。Slate を用いた作業の例を図 1 に示す。この例では、文章中で「NATO」は青色の線分で関連付けられており、これらが共参照関係にあることを表す。また、「作成する」と「ギングリッチ新下院議長」、「鮮明にし始める」

と「同党」はそれぞれ述語とそのガ格の項であることを矢印で関連付けられている。ツールの利用方法については実際に行う述語項構造・共参照解析以外の作業において操作に慣れてもらい、ツールの操作による負荷だけは軽減した上で、評価対象となる作業を行ってもらう。

4.3 作業結果の評価

素人と玄人の作業結果をいくつかの観点で評価する。まず、作業者のアノテーションの正確さを見積もるため、すでにアノテーション済みの結果と比較することで正解率を求める。BCCWJ のコアデータへの述語項構造・共参照関係のアノテーションはすでに奈良先端科学技術大学院大学松本研究室で作業が進められているため、この作業結果を正解データとみなすことで作業者のアノテーション作業の正解率を求めることができる。この正解率に基づき、素人と玄人の作業の正確さがどの程度異なるのかを調査する。

また、非明示的なデータ収集の結果には視線情報やアノテーションツールの操作履歴が含まれるため、素人と玄人でどのような行為や注視の違いがでるかを調査する。このタスクにおける行為の差については、例えば、作業者がアノテーション対象となる照応詞を選択してから先行詞を発見しアノテーションをするまでの平均時間や、その作業後に次の照応詞の候補を発見

²具体的な作業内容については文献 [9] を参照されたい

するまでの平均時間に違いが生じるのかを調査する。また、作業者の視線に着目した場合、作業者が文章を読んでいく過程でどのような談話要素を注視し、また先行詞を探索する際にどのような先行詞候補を注視しているかの分析も興味深い。アノテーション作業に関するこのような詳細な分析はこれまでになされていないため、どの程度有効なものであるかは未知数だが、もし意味のある差分が出るのであれば、提案する非明示的な情報を使ったアノテーション作業の分析をさらに推し進める必要がある。この分析の結果、例えば、玄人は正しい関係をアノテーションしているのに対し、素人が誤った作業を行った場合、それが何に起因しているかを詳細に調査することが可能になると考えられる。この調査によって、例えば、ガ格のアノテーションを行う場合は比較的前の文脈まで見るべきであるのに対し、ヲ格を付与する場合は主題となっていない談話要素を細かく検討すべき、といった作業方針に関するフィードバックができ、最終的に構築される言語資源の品質を有意に向上できると考えられる。

5 関連研究

アノテーション作業を与えられた課題に関する問題解決という観点で考えると、本稿で提案する研究課題に関して関連する研究が存在する。例えば、Rosengrant[6]は物理の問題、特に電気回路の抵抗などを求める問題で素人と玄人がそれぞれ問題を解く際の振舞いの違いについて調査している。彼らの実験設定では、被験者に視線追跡装置を着用させ、ペイントツール上に表示された回路に関して抵抗を求めさせている。被験者が式を導出し、解を求める過程などを画面上に記述させることで、回路や式など、どの箇所に着目しているかを分析の際に利用している。彼らは、被験者がどのように問題について探索するか、また、どのような戦略で解くかについて、素人と玄人で異なる振舞いすることを報告している。例えば、抵抗を計算する問題では、素人が一度解を求めるとその導出に利用した式のみ注視し続けるのに対し、玄人の場合は視線が回路、解いた結果、問題を解くために書き直した回路の3つを行き来するという違いがあると報告している。これは素人が一度導出した式は正しいという前提に立つのに対し、玄人が式の導出そのものにも間違いが含まれないかという慎重な姿勢で問題を解いているという戦略の違いが出ていることに関連している。彼らは今後の課題として、物理だけでなく他の問題解決においても同様の分析を行うことの重要性を述べている。

また、他の問題解決の問題として、プログラミングにおける誤り検出に関して素人と玄人の振舞いの違いを視線情報から分析する試みがなされている [1, 5]。例えば、Bednarikら [1] はプログラマがデバッグするときの視線に関して分析を行っている。デバッグ画面を3分割し、ソースコードの領域、コードの内容を可視化した領域、出力結果の領域の3つの領域を素人（経験の少ないプログラマ）と玄人（経験豊富なプログラマ）がどう見るのかを調査し、また時間内にどのくらい誤りを発見できるかを調査している。この結果、玄人のほうが誤りを多く検出するという結果を得たが、これは素人がある領域を見て次に別の領域を見るといった注視のパターンを繰り返し行っており、その行為が素人の誤り検出率の低さと関係しているためだと分析している。また、玄人は作業の後半になるにつれてプログラムの出力も注視するようになり、頻繁に3つの領域を切り替えて見るという傾向が見られたと報告している。確かに、彼らが行ったように問題とする画面をいくつかの領域に分割することでなんらかの傾向は調査できると考えられる。しかし、この研究で分析された注視の特定の傾向は、なぜプログラマが間違いを検出できたかについて直接の答えを示すものではない。さらに精密に視線を分析するために意味のある粒度に画面（もしくは我々の対象の場合は文章）を細かく分割することが重要だと考えられるが、細かく分割しすぎると視線情報の取得が困難になる。このバランスを考えることも今後行うべき重要な課題だと考えられる。

さらに、自然言語処理の分野でもアノテーション作業の分析が試みられている。例えば、Tomanekら [7] は、能動学習で各事例を追加する（つまり、その事例をアノテーションする）コストが一樣であるという点に疑問を持ち、各事例へのアノテーションのコストを求める問題に取り組んでいる。彼らは注視の時間や視線の動きのパターンによって人手による言語分析の難しさの推定を行っている。文章中の固有名のアノテーション、特に人名・組織名・場所名の3種類に限定したアノテーションの課題に関して、統語・意味的な複雑さの組合せが異なる4種類の問題タイプに関して対象となる句とその文脈の注視時間を調査している。彼らの報告によると、統語的もしくは意味的な複雑さが増すとアノテーション対象となる句への注視時間が長くなり、また意味的に複雑な課題では文脈の注視時間が長くなることを報告している。ただし、視線の分類は対象となる句を注視しているかまたその文脈を注視しているかの2通りしかなく、それ以上の細かい分析はなされていない。また、この調査では問題の複雑さ

にのみ着目されており、作業者の種類に関する調査は行われていない。これに対し、本研究では述語項構造と共参照関係のアノテーションという比較的複雑なタスクに対し、素人と玄人の2種類の作業者のタイプを想定し、作業者間の振舞いの違いを分析することでアノテーション作業の質に何が影響するのかを調査することで、作業へ効果的なフィードバックを行うことができると考えている。

6 おわりに

本稿では、コーパスへのアノテーションへの新たな試みとして、作業者のアノテーションの過程を記録する非明示的なデータ収集と、二人の作業者が対話的に作業を行う明示的なデータ収集の方法を提案し、さらにそれぞれの具体的な利用法として、視線情報やアノテーション行為に関する情報を使った問題分析、アノテーション作業者の能力の客観的評価、素人の作業者のアノテーション作業の学習支援について紹介した。本稿で述べた内容は、人手によるアノテーション作業とその作業結果の利用をつなぐ重要な問題と考えることができ、2節で述べたデータ収集方法の有効性について今後詳しい調査が必要だと考えている。今後は、4節で述べたように、まず非明示的なデータ収集に着目し、述語項構造と共参照のアノテーションを例にデータを収集して、その結果を分析する予定である。

参考文献

- [1] R. Bednarik and M. Tukiainen. Temporal eye-tracking data: evolution of debugging strategies with multiple representations. In *Proceedings of the 2008 symposium on Eye tracking research and applications*, pp. 99–102, 2008.
- [2] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37–46, 1960.
- [3] E. Hovy, M. Marcus, M. Palmer, L. Ramshaw, and R. Weischedel. Ontonotes: The 90% solution. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pp. 57–60, 2006.
- [4] D. Kaplan, R. Iida, K. Nishina, and T. Tokunaga. Slate - a tool for creating and maintaining annotated corpora. *Journal for Language Technology and Computational Linguistics*, Vol. 26, No. 2, pp. 89–101, 2012.
- [5] S. Pietinen, R. Bednarik, T. Glotova, V. Tenhunen, and M. Tukiainen. A method to study visual attention aspects of collaboration: eye-tracking pair programmers simultaneously. In *Proceedings of the 2008*

symposium on Eye tracking research and applications, pp. 39–42, 2008.

- [6] D. Rosengrant. Gaze scribing in physics problem solving. In *Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications*, pp. 45–48, 2010.
- [7] K. Tomanek, U. Hahn, S. Lohmann, and J. Ziegler. A cognitive cost model of annotations based on eye-tracking data. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 1158–1167, 2010.
- [8] 飯田龍. 意味・談話処理課題の規格化とその諸問題. 人工知能学会誌 特集「ポスト経験主義の言語処理」, Vol. 27, No. 3, pp. 318–325, 2012.
- [9] 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.