

Stativity 判定に基づいた時制誤り検出

永田 亮[†] Vera Sheinman^{††}

[†] 甲南大学知能情報学部 ^{††} 株式会社教育測定研究所

E-mail: [†]rnagata@konan-u.ac.jp, ^{††}sheinman@jiem.co.jp

1. はじめに

本稿では、英文中の時制誤りを検出する手法を提案する。時制誤りは、学習者の英文に頻出する [7]。そのため、英文の自動添削 [2], [9] や CALL システムでは、時制誤りを検出できることが望ましい。しかしながら、我々が知る限り、時制誤り検出に関する研究はほとんどなされていない。数少ない例外として、Izumi ら [6] の誤り検出手法がある。この手法は、対象となる誤りの種類を選ばない一般的な誤り検出手法であり、時制誤り検出にも応用可能である。また、最大エントロピー法に基づいた分類器で、正しい冠詞や前置詞を推定する手法 [3], [4] も時制誤り検出に応用可能である（すなわち、同様な手法で正しい時制を推定する）。しかしながら、4. で示すように、これらの手法は時制誤りに用いてデザインされておらず、その性能には改善の余地が大いにある。

更なる議論に進む前に、学習者の英文を観察することにする。図 1 は、Konan-JIEM Learner Corpus (KJ コーパス^(注1)) から抜粋した英文である。図 1 からわかるように、学習者の英文には、時制誤り (<v.tns> でタグ付けされた部分) を含め、多様な文法誤り (冠詞誤り: <at>, 前置詞誤り <prp>) が頻出する。また、綴り誤り (例: Seasun, skky. 正しくは, Season と ski) や不自然な表現 (例: play skky) も出現する。このことは、学習者の英文はノイズを多く含むことを示唆する。従って、学習者の英文を対象とした言語処理では、このノイズを十分に考慮することが重要である。

Lee ら [8] が主張しているように、時制誤り検出は難しいタスクである。例えば、

I like the BMW Z3^(注2).

と、

I liked the BMW Z3.

において、どちらの時制が正しいかを決定することを考える。書いている時点で BMW Z3 が好きである場合は、現在が正しい時制である。一方、過去に好きであった場合は、過去が正しい。言い換えると、時制の決定は、書き手の意図と文脈に大きく依存する。したがって、計算機で正しい時制を決定することは不可能のように思われる。

(注 1) : <http://www.gsk.or.jp/> より入手可能である。

(注 2) : BMW Z3 とは車の一種である。

```
<at crr="The"></at> Seasun <v tns crr="was">is</v tns> winter.
It <v tns crr="was">is</v tns> very cold.
I <v tns crr="played">play</v tns> skky and
<v tns crr="played">play</v tns> <prp crr="with"></prp> friends.
It <v tns crr="was">is</v tns> very interesting.
I <v tns crr="have">had</v tns> a good memory.
```

図 1: 学習者の英文例

このような時制誤り検出の難しさにもかかわらず、本稿では、*stativity* と呼ばれる動詞の性質を利用することで、学習者の英文に頻出する時制の誤りを効果的に検出できることを示す。我々は、*stativity* が動詞における時制の選択に制限を与えることに着目する。この制限を時制誤りの検出に利用するため、提案手法では、まず、動詞の *stativity* を予測する。次に、予測結果に基づいて時制誤り検出を行う。評価実験で示すように、提案手法は F 値で 0.571 を達成し、比較のために実装した手法の性能を大きく上まわる。

2. 基本アイデア

提案手法の基本アイデアを導くために次の例文を考える:

I read the book.

ここで、上の例文を発音してみる。驚くべきことに、動詞 “read” は不規則動詞であるため現在形と過去形が同一であるにもかかわらず、大抵の読者は /red/ と (すなわち、過去として) 発音するであろう。すなわち、表層情報には明示されないにもかかわらず過去として認識する。

この現象は、*stativity* により説明される。*stativity* とは、動詞の分類方法の一種である。*stativity* に基づくと、状態を表す動詞は状態動詞、動作を表す動詞は動作動詞に分類される。例えば、動詞 like は、

I like the BMW Z3.

では状態動詞であり、動詞 drove は、

I drove the BMW Z3.

動作動詞である。

stativity から時制は制限を受ける。具体的には、動作動詞の現在時制が制限を受ける [5]。動作動詞と現在時制の組み合わせは、特殊な場合を除き、発話時点で起こっている動作を表現するには用いられない (例: He reads the book.)。したがって、上述の例では、表層情報には明示されないにもかかわらず、過去形/red/と発音される。

学習者の英文中では、この組み合わせは特殊な用法ではなく、時制誤りであるとするのが提案手法の基本アイデアである。この基本アイデアに従うと、stativity さえわかれば、動作動詞の現在形を誤りとして検出できることになる。そこで、提案手法では stativity の予測を行う。まず、stativity を予測するための訓練データを作成する。訓練データとは、コーパス中の動詞に対して stativity のラベルを付与したデータのことである。次に、訓練データから、stativity を予測するための分類器を学習する。誤り検出の際には、分類器を用いて、誤り検出の対象となる動詞の stativity を予測する。最後に、予測結果に基づいて時制誤りの検出を行う。予測結果が動作動詞かつ対象動詞が現在形の場合、時制誤りとして検出する。

3. 提案手法

3.1 stativity の定義

動作動詞でも、習慣的動作を表す場合は現在形で使用される。上述の基本アイデアでは、この用法を誤りとして誤検出してしまう。そこで、提案手法では、一般的な定義を拡張して、次のように stativity を定義する：

[定義 1] (stativity) 動作動詞：動作を表す動詞。ただし、習慣的動作を除く。その他の動詞：それ以外の動詞

この定義のもと、提案手法では stativity の予測を行う。以下では、表記を簡単にするため、特に断わらない限り、動作動詞を D、その他の動詞を O と省略して表記する。

3.2 訓練データ

訓練データを得るため、コーパスから検出対象の動詞を含む文を収集する。各動詞に対して D/O のラベルを付与する。例えば、

I had hot coffee.

では、

I had/D hot coffee.

とラベル付けする。また、

I had a BMW Z3.

では、

I had/O a BMW Z3.

とラベル付けする。

提案手法では、ラベル付けは人手により行う。人手によるラベル付けは時間と労力を要するので、将来的には、半自動化/自動化することを検討している (詳細は、5. で議論する)。

3.3 stativity の予測と時制誤り検出

訓練データが得られると、機械学習に基づく分類器などにより stativity の予測が可能となる。提案手法では、誤り検出手法でよく使用される最大エントロピー法に基づく分類器を使用する (例えば、文献 [3], [4], [6] の誤り検出手法でも使用されている)。

素性として、次の三種類を定義する：(i) 副詞、(ii) 動詞句内の単語 (ただし、検出対象の動詞は除く)、(iii) 動詞句の左右 k 単語。副詞は、動詞を直接修飾することが可能であるので、stativity の予測に有益な情報であると考え選択した。また、動詞の周辺の単語 (ii) と (iii) も同様に予測に有益な情報として選択した。(ii) と (iii) の単語を素性とする際には副詞は無視し、左右 k 単語には含めない。また、“will” など時制を直接表す語も除外する。各単語は、原形、小文字に変換する。また、特定の単語は、その意味に対応するコードに変換して素性とする (例えば、“tea” や “coffee” を DRINK とする)。

具体例として、次の例文を考える：

The man quickly [VP had/D] hot coffee.

$k = 2$ のとき、この例文からは、次の素性が得られる：

副詞=quickly, 左=the, 左=man, 右=hot, 右=DRINK

動詞句の左の素性を抽出する際に、副詞 “quickly” は、左 k 単語に含まれていないことに注意する必要がある。また、DRINK は、coffee に対応する意味コードである。

このような素性を基に、訓練データを素性ベクトルに変換する。得られた素性ベクトルから分類器を学習し、stativity の予測に用いる。

誤り検出では、上述と同様の手順で検出対象の文章中の動詞を素性ベクトルに変換する。その素性ベクトルを分類器に与え、stativity を予測する。予測結果に基づき誤りを検出する。予測結果が D かつ対象動詞が単純現在の場合、誤りとして検出する。

誤り検出精度を高めるため、提案手法では次のような例外を設ける。命令形では、動作動詞も現在形 (原形) で使用される。また、「時」「条件」を表す副詞節では、未来時制の代わりに現在形が使用される。これらの例外的用法に対しては、常に正しいと判定する。

4. 評価実験

評価実験のために、10 種類の動詞^(注 3)を選択した。英語教

(注 3) : buy, come, find, get, go, make, play, see, take, use.

材などからなるコーパス（約 18 万語）から、これらの動詞を含む文を抽出し、訓練データを人手で作成した。その結果、2563 の訓練事例が得られた。

素性の k の値は leave-one-out クロスバリデーションを用いて決定した。 $1 \leq k \leq 10$ の各 k を試し、予測精度が最大となる $k = 3$ を選択した。その際の予測精度は 0.789 であった^(注 4)。この結果得られた分類器を用いて、誤り検出の評価実験を行った。

検出対象の文章は、KJ コーパスに収録されている 170 のエッセイとした。同エッセイには、155 の時制誤りが含まれていた。155 の誤りのうち、対象とした 10 種類の動詞における時制誤りは 41 であった。検出性能は、recall, precision, F -measure で評価した。

比較のため、次の 3 種類の手法を実装した：

- **Izumi らの手法**: Izumi ら [6] により提案された手法である。誤り情報が付与された学習者コーパスから、正誤を判定するための最大エントロピー法に基づく分類器を学習し誤り検出を行う。本評価実験では、時制誤りのみを対象とするように実装した。また、実装には、文法誤り情報が付与された学習者コーパスが必要であるため、leave-one-out クロスバリデーションにより性能を評価した（一度に一エッセイを取り除いた）。

- **3-way**: この手法では、最大エントロピー法に基づいた分類器を用いて、現在、過去、未来の 3 値分類を行い誤りの検出を行う。分類結果と実際の時制が異なるときに誤りと判定する。この方法は、冠詞誤り検出手法 [4] や前置詞誤り検出手法 [3] の時制誤り検出への適用と見なすことができる。素性は、提案手法と同じものを使用した。

- **ベースライン**: 現在形を常に誤りと判定した。

表 1 に実験結果を示す。表の上段は、10 種類の動詞を対象にした検出性能を示す。下段は、全ての動詞を対象にした検出性能を示す^(注 5)。

5. 考 察

表 1 から、提案手法の検出性能は、他の手法に比べて高いことがわかる。この理由として、提案手法では、stativity の判定に基づいて時制誤りの検出を行うことが挙げられる。実際、提案手法で使用している文脈情報は、stativity の判定に有効であることが実験結果より確かめられた。例えば、動詞 “play” に対して、“But I don’t play/O well it” や “My

(注 4) : 比較のために、ベースライン（多数決による判定）精度を求めたところ 0.657 であった。

(注 5) : 提案手法では 10 種類の動詞のみを対象にして訓練データを作成したため、他の動詞に対しては誤りの検出が行えない。従って、10 種類の動詞を対象としたときの検出結果をそのまま用いて、全動詞に対する性能を求めた。その結果、precision は同一、recall が低下する結果となっている。一方、従来手法では、全動詞を対象にして検出を行った。

表 1: 検出性能

Method	Recall	Precision	F -measure
提案手法	0.585	0.558	0.571
	0.155	0.558	0.242
Izumi らの手法	0.171	0.538	0.259
	0.058	0.360	0.100
3-way	0.585	0.113	0.189
	0.503	0.065	0.115
ベースライン	0.854	0.213	0.341
	0.832	0.071	0.131

上段: 10 種類の動詞に対する性能

下段: 全動詞に対する性能

brother and I played/D the video game” のように、文脈に応じて正しく stativity を判定した。前者では、副詞 “well” が判定の手掛りとなっていることが確認できた（学習された分類器において、D と O に対する重みは、それぞれ -9.52 と 2.36 であった）。同様な例を重みが高い単語に下線を引いて示す：“It makes/O me happy.” vs “I make/D pies last year.”; “I went/D to Hawaii with my family.” vs “I go/O to school everyday.”

一方で、従来手法の性能は低い。“3-way” では、特に precision が低い。多くの場合、時制（現在、過去、未来）の予測に失敗し、実際に使用されている時制と一致せず誤りと判定してしまう。そのため、precision が低く、recall が高い結果となった。このことは、1. で述べたように、時制の決定が書き手の意図と文脈に大きく依存することを考慮すると、容易に予想できるものである。逆に、Izumi らの手法は recall が低い。同手法では、誤り情報が付与された学習者コーパスを訓練データとして、正誤を判定する分類器を学習する。そのため、訓練データに出現しない事例に対しては上手く検出が行えない傾向にある。訓練データの規模を大きくすることで検出性能が向上することが予想できる。しかしながら、誤り情報が付与された学習者コーパスが不足している現状を考慮すると、容易には実現できない。

提案手法は、現在形に関する誤りしか検出しかできないので recall が低いことが予想される。しかしながら、実際には、他の手法に比べても遜色ない recall を達成している。この理由は、学習者の時制誤りの傾向により説明できる。実験結果を分析したところ、155 の時制誤りのうち、89.7% が現在形に関する誤りであることが明らかになった^(注 6)。また、NICT JLE コーパスでも同様な傾向が確認できる。学習者は、英文を書く際に、語彙や構文など時制以外にも多くのことを考え

(注 6) : 表 1 において、現在形を常に誤りとするベースラインの recall は 0.832 であり、0.897 と一致しない。これは、解析ミスにより抽出できない動詞が存在することに基因する。

なければならない。そのため、時制を明示するのを忘れ（動詞を原形のまま使用し）、このような傾向が現れたと予想できる。この傾向により、提案手法では現在形に関する誤りのみ対象にしているにもかかわらず、高い recall を達成したと考察できる。

ここまでの考察で、提案手法は時制誤りの検出に有効であることが確認できたが、依然、検出漏れと誤検出が存在する。以下では、検出漏れと誤検出について考察を行う。

検出漏れは、17 件存在した（10 種類の検出対象の動詞における 41 の誤り中 17 件）。検出漏れの大きな要因は、文法誤りと綴誤りであった（53%が該当）。これらの誤りは、stativity の予測に悪影響を及ぼす。訓練データは母語話者のコーパスを基にしているため、誤りがある単語は素性として活用できない。別の大きな要因として、stativity の判定に有効な単語が素性ベクトルの文脈窓の外に存在する場合が挙げられる（29%が該当）。単純に、文脈窓の大きさ（すなわち k の値）を大きくすると無関係な情報や不必要な情報も素性として含まれるため精度低下につながる。構文解析の導入は有効であると期待されるが、解析ミスという新たなノイズも生じることには注意する必要がある。

検出漏れと同様に、文法誤りと綴誤りが誤検出の大きな要因である（19 の誤検出の 37%に相当）。複数の文にまたがる文脈情報も誤検出の要因である（16%に相当）。例えば、“She makes interesting stories.” も “She made interesting stories.” も一文としては時制に誤りはない^(注 7)。時制を正しく決定するためには前後の文を考慮する必要があるが、提案手法でも従来手法でも考慮していない。

検出性能を向上させることに加えて、提案手法には別の改善の余地がある。提案手法では、人手で訓練データを作成する。具体的には、本評価実験の場合、2563 の動詞にラベル付けを行うのに約 20 時間要した（1 日あたり平均 2 時間、計 10 日間）。この作業にかかる時間と労力を極力減らすことが好ましい。一つの実現方法として、bootstrapping などの半教師付き学習が効果的であると予想している。すなわち、表層情報により少数の動詞に D/O のラベルを付与し、半教師付き学習により分類器を学習する。例えば、進行形の動詞は基本的に動作動詞である。また、ある種の副詞（例えば、“deliberately” や “carefully”）に修飾された動詞は動作動詞である可能性が高い。このような表層情報を基に、半教師付き学習により分類器を学習できる。

6. おわりに

本稿では、学習者の英文に頻出する時制誤りを効果的に検出する手法を提案した。評価誤検出実験の結果、次の 3 点を

明らかにした：(i) 時制を直接推定することは現状の技術では難しい、(ii) 一方で、stativity は高い精度で判定可能である、(iii) stativity の判定に基づいた提案手法の検出性能は高い。また、提案手法では、誤り情報付きの学習者コーパスを必要としないという利点もある。

今後は stativity 判定のための訓練データを半自動的に作成する方法の考案に取り組む予定である。また、より高度な時制誤り検出を実現するために、時間的順序関係推定（例えば、文献 [1], [10] など）と提案手法を組み合わせることに取り組む予定である。

謝 辞

本研究の一部は科研費 (19700637) の助成により実施した。

参考文献

- [1] B. Boguraev and R.K. Ando, “TimeML-compliant text analysis for temporal reasoning,” Proc. of 19th International Joint Conference on Artificial Intelligence, pp.997–1003, 2005.
- [2] J. Burstein, M. Chodorow, and C. Leacock, “Automated essay evaluation: The *Criterion* online writing service,” AI Magazine, vol.25, no.3, pp.27–36, Sep. 2004.
- [3] M. Chodorow, J.R. Tetreault, and N.R. Han, “Detection of grammatical errors involving prepositions,” Proc. of 4th ACL-SIGSEM Workshop on Prepositions, pp.25–30, 2007.
- [4] N.R. Han, M. Chodorow, and C. Leacock, “Detecting errors in English article usage by non-native speakers,” Natural Language Engineering, vol.12, no.2, pp.115–129, 2006.
- [5] R. Huddleston and G.K. Pullum, The Cambridge Grammar of the English Language, Cambridge University Press, Cambridge, 2006.
- [6] E. Izumi, K. Uchimoto, T. Saiga, T. Supnithi, and H. Isahara, “Automatic error detection in the Japanese learners’ English spoken data,” Proc. of 41st Annual Meeting of ACL, pp.145–148, 2003.
- [7] J. Lee and S. Seneff, “Automatic grammar correction for second-language learners,” Proc. of 9th International Conference on Spoken Language Processing, pp.1978–1981, 2006.
- [8] J. Lee and S. Seneff, “Correcting misuse of verb forms,” Proc. of 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technology Conference, pp.174–182, 2008.
- [9] E.B. Page, J.P. Poggio, and T.Z. Keith, “Computer analysis of student essays: Finding trait differences in student profile,” Proc. of 1997 Annual Meeting of the American Educational Research Association, 1997.
- [10] K. Yoshikawa, S. Riedel, M. Asahara, and Y. Matsumoto, “Jointly identifying temporal relations with Markov logic,” Proc. of 47th Annual Meeting of the Association for Computational Linguistics and 4th International Joint Conference on Natural Language Processing, pp.405–413, 2009.

(注 7) : これらの例は、実際に学習者が書いた文 “She make very interesting story.” を基にしている。