

言語横断手法による日本語時間的順序関係推定

吉川 克正

浅原 正幸

日本アイ・ビー・エム (株) 人間文化研究機構 国立国語研究所

東京基礎研究所

言語資源研究系

katsuy@jp.ibm.com

masayu-a@ninjal.ac.jp

1 はじめに

時間情報解析は、近年様々な言語において研究が進められ、質問応答をはじめ、数多くの応用が期待される研究分野である。その解析は、(1) 事象・時間表現抽出 (2) 事象・時間属性同定 (3) 時間情報正規化 (4) 時間的順序関係推定、という4つの部分問題へと分割することができる。時間的順序関係推定の研究は主に、英語、スペイン語、中国語といった主要な言語を中心に研究されており、日本語を含めた話者が比較的少ない言語での研究事例は非常に少なかった。ごく近年になって、浅原らがBCCWJ-TimeBank [1]を整備したことで、日本語での時間情報解析研究も行われ始めた [13, 14]。しかし、BCCWJ-TimeBankのような時間情報タグ付きコーパスは、一般にその質と規模の両面において問題を多く抱えており、それを利用した時間情報解析も難易度の高いタスクとなっている。

本研究では日本語時間的順序関係推定に焦点を当て、これまでは研究されていなかった言語横断手法を提案する。本研究で提案する言語横断手法は、

1. 推移律をベースとした言語非依存の素性を利用する
2. 他言語のコーパスを自動翻訳して、対象言語(日本語)の学習データを生成する

、という2つのアイデアに基づいている。図1では、英語のコーパスを日本語に自動変換した上で、事象・時間表現情報を英 *rightarrow* 日へとマップした事例を示している。時間的順序関係推定は、e1-DCT, e1-e3間のように、表現間における時間的な順序を推定するタスクである。従って、この図のように各表現のマッピングさえできれば、他言語からの自動翻訳データを学習データに利用することもできる。自動翻訳データは、“会談する予定しています”のように、一定の誤りを含むが、推移律を考慮することにより、エラーの伝播を抑えることができる。ここでいう推移律とは、

$$(e1 \text{ BEFORE DCT}) \wedge (\text{DCT OVERLAP } e3) \\ \Rightarrow (e1 \text{ BEFORE } e3) \quad (1)$$

表 2: 17 ラベルから 6 ラベルへの変換

BCCWJ-TimeBank	TempEval-2(略称)
after	AFTER (A)
met-by	AFTER (A)
overlapped-by	OVERLAP-OR-AFTER (OA)
finishes	OVERLAP-OR-AFTER (OA)
during, is_included	OVERLAP (O)
started-by	OVERLAP (O)
equal, identity	OVERLAP (O)
starts	BEFORE-OR-OVERLAP (BO)
contains, includes	OVERLAP (O)
finished-by	OVERLAP (O)
overlaps	BEFORE-OR-OVERLAP (BO)
meets	BEFORE (B)
before	BEFORE (B)
vague	VAGUE (V)

のような制約であり、時間的順序関係を推定する上で有力な手がかりとなる。この制約は言語に依存しない論理的制約であるから、言語非依存の素性として利用することが可能である。

さらに本研究の提案手法については、BCCWJ-TimeBankを利用して評価を行い、日本語のデータが無い、または非常に少ないという状況下で、一定の性能改善が見られることを確認した。

2 日英の時間情報タグ付きコーパス

本研究で学習・評価に利用するBCCWJ-TimeBankは、“日本語書き言葉均衡コーパス”(BCCWJ) [7]のコーパスに対して、TimeMLフォーマット [9]により、解析に必要とされる事象(EVENT)、時間表現(TIMEX3)、時間的順序関係(TLINK)をアノテーションしたコーパスである。その詳細は先行研究 [1]に譲るが、特に重要なTLINKには、順序関係ラベルとして17種類、また対象とする表現の種類(事象・時間表現)により、表1に示す4つの関係推定タスクが定義されている。

英語の時間的順序関係推定は数多くの先行研究 [2, 8, 3, 6, 4, 5]があるが、その多くはTempEval-1,2,3 [11, 12, 10]のデータを利用している。本研究では、順序関係推定タスクの対応付けが比較的容易なTempEval-2のデータを利用している。しかし、TempEval-2の関係ラベルは6種

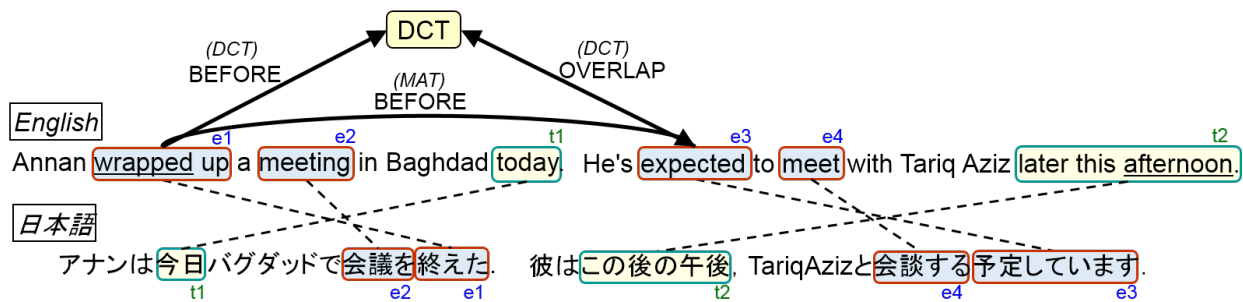


図 1: 英語から日本語に自動翻訳された時間情報タグ付きコーパスの例

表 1: BCCWJ-TimeBank で定義される順序関係推定タスク

タスク	TempEval-2	タスク説明
T2E	(C)	時間表現と事象の順序関係推定
DCT	D	文書作成日時 (DCT) と事象の順序関係推定
E2E	(F)	隣接する 2 事象間の順序関係推定
MAT	E	隣接 2 文内にある主事象間の順序関係推定

類であり, BCCWJ-TimeBank のそれよりも粗いため, 表 2 に示した変換表に基づき, 17 ラベルから 6 ラベルへの変換を行う. さらに, タスクの対応付けは, 表 1 のように Task D をタスク DCT に, Task E をタスク MAT へとマップすることになる. 残りの 2 タスクについては, BCCWJ-TimeBank と TempEval-2 でアノテーション基準が無視できないほど大きく異なるため, 本研究では扱わないものとする.

3 言語横断手法による日本語時間的順序関係推定

本研究では, 先行研究を元に時間的順序推定のための分類学習モデルを構築し, その上で言語横断手法へと拡張する. 分類学習器, 素性等は全て吉川ら [14] に倣い, ここでは言語横断拡張についてのみ解説する.

3.1 推移律素性

推移律素性 (*Transition Features*) とは, 論理的に成立する推移律を元に構築された言語に非依存の素性である. まず日・英それぞれの言語で, 推移律がどの程度保持されているかを検証する. ここでいう推移律とは図 3 のように $e1, e2, DCT$ の間で成立する関係である. この “ $r1 \wedge r2 \Rightarrow R$ ” の中で, $r1$ と $r2$ が取るラベルの組み合わせに対して R にどのラベルがアノテーションされているかを調べる. この結果を日英それぞれのコーパスごとに示したのが, 表 4 と表 5 である. この 2 つの表を比べて見ると, ほぼ全ての $(r1, r2)$ の組み合わせに対し, R の最頻ラベルが日本語・英語で一致していることが分かる. 例えば, $(r1, r2) = (AFTER, BEFORE)$ のとき, 日本語・英語ともに 100% の比率で R のラベルは AFTER となり, 完全一致である. 即ち, 推移律は言語に依存することなく利用できると考えられる. 本研究ではこの推移律について, 表 4 と表 5 に示す “比率” をそのまま実

表 6: 翻訳データの概要

	Event	Time	C	D	E	F
Train	5639	1037	959	640	1587	1721
Test	962	155	138	365	255	304
Total	6601	1192	1097	1005	1842	2025

数値素性の形式で利用している. 例えば, 英語のデータにおいて, $(r1, r2) = (BEFORE, AFTER)$ であるとき, 推移律素性は “(BEFORE:0.80, OVERLAP:0.20)” の形式になる. この英語から抽出した推移律素性を, 日本語の時間的順序関係推定に適用することにより, 言語横断手法を実現できる.

3.2 機械翻訳による学習データ生成

本研究における言語横断拡張の 2 つ目は, 他言語のデータを機械翻訳することで, 解析対象となる言語の学習データを生成する手法である. ここでは英日翻訳を利用し, TempEval-2 の英語データを日本語へと変換する. この機械翻訳には Microsoft Translation API¹ を利用している. この API の出力には, 生成された翻訳文だけでなく, 原言語から目標言語に対する文字ベースのアラインメント情報が含まれる. 従って, このアラインメント情報を元にして, 英語データに含まれる TimeML タグ (EVENT, TIMEX3) を日本語翻訳文へとマッピングする. 図 1 にはその一例が示されており, 例えば事象 $e1$ は *wrapped up* から “*終えた*” へ, 時間表現 $t1$ は *today* から “*今日*” へとマッピングされることが分かる. このマッピングは 1 単語が必ずしも 1 単語に紐づくわけではなく, n 対 n 単語のマッピングになる. つまり, 英語での 2 単語が日本語での 1 単語にマップされたり, 逆に英語 1 単語が日本語 2 単語にマップされることもある. 本研究では出力される文字ベースの範囲を最大限に被覆するよう

¹<http://msdn.microsoft.com/en-us/library/dd576287.aspx>

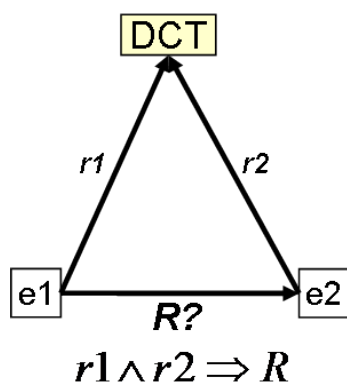


表 3: 推移律

表 4: 日本語 (BCCWJ-TimeBank)

$r1 \wedge r2 \Rightarrow$	R	Ratio
$B \wedge A \Rightarrow$	B	91.7%
	V	5.56%
	A	2.78%
$A \wedge B \Rightarrow$	A	100%
$B \wedge O \Rightarrow$	O	57.1%
	B	42.9%
$O \wedge B \Rightarrow$	O	68.2%
	A	27.3%
	V	4.5%
$A \wedge O \Rightarrow$	A	66.7%
	O	16.7%
	V	16.7%
$O \wedge A \Rightarrow$	O	68.2%
	B	33.3%

表 5: 英語 (TempEval-2)

$r1 \wedge r2 \Rightarrow$	R	Ratio
$B \wedge A \Rightarrow$	B	80.0%
	O	20.0%
$A \wedge B \Rightarrow$	A	100%
$B \wedge O \Rightarrow$	O	72.0%
	B	28.0%
$O \wedge B \Rightarrow$	O	76.0%
	A	12.0%
	V	8.0%
	OA	4.0%
$A \wedge O \Rightarrow$	A	33.3%
	O	33.3%
	OA	33.3%
$O \wedge A \Rightarrow$	B	100.0%

な n 単語列を決定し、それを事象または時間表現の範囲としている。しかし、中には翻訳された事象や時間表現がアラインメント先を持たない場合も生じたため、マップできないタグは利用しないこととした。このような翻訳とマッピング処理を経て得られたデータの統計を表 6 に示した。

機械翻訳によって得られた翻訳文は一定の解析誤りを含んでいる。例えば、図 1 の中で、“会談する (e3)” と “予定しています (e4)” の 2 つの事象は、文法的に適切な接続となっていない。このような誤りは、順序関係推定に重大な影響を及ぼすことも多くあるが、3.1 節で述べた推移律素性と併用することで、誤りが軽減され、効果的な学習が可能になる。

4 実験と結果

4.1 実験設定

本研究では、主に日本語のデータが非常に少量であり、その状態では時間情報解析が難しい場合を想定して実験を行う。次の 5 つのデータセットと実験設定を用意する。

1. 日本語のデータが全くなく、英語の翻訳データが利用できる (En)
2. 少量の日本語データが利用できる ($Ja-Small$)
3. 少量の日本語データと、英語の翻訳データの両方が利用できる ($En+Ja-Small$)
4. 全ての日本語データが利用できる ($Ja-Large$)
5. 全ての日本語データと、英語の翻訳データの両方が利用できる ($En+Ja-Large$)

尚、ここでいう少量の日本語データとは、BCCWJ-TimeBank 全体 (54 文書) に対して、1/10 のサイズ (5 文書) をいう。実験の評価には吉川ら [14] と同様に 5 分割交差検定を行い、各分割セットに含まれる学習データからランダムに選ばれた 5 文書だけを利用して学習デー

タとする。TempEval-2 の翻訳データは常に全てを利用する。

さらにタスク MAT については、推移律素性の利用によって 3 種類の実験設定を用意する。MAT *w/o TF* は推移律素性を利用しない場合、MAT *with TF(Sys)* と MAT *with TF(Gold)* は共に推移律素性を利用するが、推移律の DCT ラベルにシステム出力を利用するのが (Sys)、コーパス中の正解ラベルを利用するのが (Gold) である。

4.2 実験結果

実験結果は表 7 にまとめて示す。まず、タスク DCT の結果を見ると、英語翻訳データのみでの結果 (En) では 67.7% になり、小規模であっても日本語データを利用した場合 ($Ja-Small$) の方が性能が良い (69.8%)。しかし、翻訳データと小規模日本語データを組み合わせた場合 ($En+Ja-Small$) はさらに結果が良くなり (71.5%)、 En と $Ja-Small$ から学習できる情報は、互いに補い合う関係にあると考えられる。

次にタスク MAT の結果だが、推移律素性の効果を見ると、 En 、 $Ja-Small$ 、 $En+Ja-Small$ 、 $Ja-Large$ 、 $En+Ja-Large$ 、いずれの設定でも大きな性能改善が見られる。特に性能差が大きいのは $Ja-Small$ で、MAT *w/o TF* (41.0%) と MAT *with TF(Gold)* (51.4%) の間では 10.4% の開きがあった。さらに、推移律素性を使わずにデータセットでの違いに着目すると、英語翻訳データのみを利用した場合 (En) であっても 46.1% となり、小規模日本語データのみを利用した場合 ($Ja-Small$, 41.0%) よりも明らかに性能が良い。ただし、全ての日本語データを利用できる環境下では、英語翻訳データを利用しても効果がなく、日本語データのみを利用するよりも性能を落とす結果となった (53.6% \rightarrow 51.9%)。

以上から得られた考察として、推移律素性は日本語データ、英語翻訳データ、そしてその両方、いずれを利

表 7: 実験結果 (F1)

Model	DCT	MAT w/o TF	MAT with TF(Sys)	MAT with TF(Gold)
<i>En</i>	67.7	46.1	47.2	49.4
<i>Ja-Small</i>	69.8	41.0	45.6	51.4
<i>En+Ja-Small</i>	71.5	48.5	50.3	53.5
<i>Ja-Large</i>	75.4	53.6	53.9	61.3
<i>En+Ja-Large</i>	75.6	51.9	53.6	60.2

用する設定においても効果が高いということ、一方、翻訳データの利用が効果を発揮できるのは、日本語データが少ない場合に限定され、日本語データを十分に利用できる場合には、翻訳誤りによる悪影響の方が強くなるということである。

5 おわりに

本稿では日本語時間的順序関係推定について言語横断手法を提案した。日英間の言語横断手法を実現するために、本研究では2つのアイデアを提案した。

1. 推移律を元にした言語非依存な素性を利用する
2. 機械翻訳により英語のコーパスを日本語へ翻訳し、それを学習データとして利用する

いずれのアイデアも一定の効果が見られたが、特に日本語データを少量しか利用できない場合において、本研究の提案手法は有効であることが確認できた。

今後の方向性として、日本語・英語の組み合わせ以外の言語対においても、本研究の言語横断手法が有効に働くかどうかを調べるとともに、機械翻訳の精度が時間情報解析にどのような影響を与えるかなど、より詳細な分析を行うことを検討している。

参考文献

- [1] Masayuki Asahara, Sachi Yasuda, Hikari Konishi, Mizuho Imada, and Kikuo Maekawa. Bccwj-timebank: Temporal and event information annotation on japanese text. In *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation*, November 2013.
- [2] Branimir Boguraev and Rie Kubota Ando. Timeml-compliant text analysis for temporal reasoning. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pp. 997–1003, 2005.
- [3] Nathanael Chambers and Daniel Jurafsky. Jointly combining implicit constraints improves temporal ordering. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 698–706, Honolulu, Hawaii, October 2008. Association for Computational Linguistics.
- [4] Pascal Denis and Philippe Muller. Predicting globally-coherent temporal structures from texts via endpoint inference and graph decomposition. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume Three, IJCAI'11*, pp. 1788–1793. AAAI Press, 2011.
- [5] Quang Do, Wei Lu, and Dan Roth. Joint inference for event timeline construction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 677–687, Jeju Island, Korea, July 2012. Association for Computational Linguistics.
- [6] Xiao Ling and Daniel S. Weld. Temporal information extraction. In *Proceedings of the Twenty Fifth National Conference on Artificial Intelligence*, 2010.
- [7] Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den. Balanced corpus of contemporary written japanese. *Language Resources and Evaluation*, Vol. 48, pp. 345–371, 2014.
- [8] Inderjeet Mani, Marc Verhagen, Ben Wellner, Chong Min Lee, and James Pustejovsky. Machine learning of temporal relations. In *ACL-44: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pp. 753–760, Morristown, NJ, USA, 2006. Association for Computational Linguistics.
- [9] James Pustejovsky, Jose Castano, Robert Ingria, Reser Sauri, Robert Gaizauskas, Andrea Setzer, and Graham Katz. Timeml: Robust specification of event and temporal expression in text. *IWCS-5, Fifth International Workshop on Computational Semantics.*, 2003.
- [10] Naushad UzZaman, Hector Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pp. 1–9, Atlanta, Georgia, USA, June 2013. Association for Computational Linguistics.
- [11] Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Graham Katz, and James Pustejovsky. Semeval-2007 task 15: Tempeval temporal relation identification. In *Proceedings of the 4th International Workshop on SemEval-2007.*, pp. 75–80, 2007.
- [12] Marc Verhagen, Roser Sauri, Tommaso Caselli, and James Pustejovsky. Semeval-2010 task 13: Tempeval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 57–62, Uppsala, Sweden, July 2010. Association for Computational Linguistics.
- [13] 稲田和明, 松林優一郎, 乾健太郎. 同一文内の表現対を対象とした日本語における時間関係認識. 言語処理学会 第 20 回年次大会 発表論文集, pp. 348–351, 2014.
- [14] 吉川克正, 浅原正幸, 飯田龍. Bccwj-timebank を対象とした時間的順序関係の推定. 言語処理学会 第 20 回年次大会 発表論文集, pp. 1103–1106, 2014.