

文間接続関係の自動同定のための人間による同定分析

齋藤 真実[†] 山本 和英[†] 関根 聡[‡]

[†]長岡技術科学大学電気系 〒940-2188 新潟県長岡市上富岡町 1603-1

[‡]ニューヨーク大学 715 Broadway, 7th floor, New York, NY, 10003 USA

E-mail: [†]{saito, ykz}@nlp.nagaokaut.ac.jp, [‡]sekine@cs.nyu.edu

あらまし 我々は文間の接続関係を自動的に推定するシステムの構築を目指している。先行研究において我々は、入力のある二文間の接続関係を大規模なテキストデータによる統計的处理によって推定する手法を提案した。このときの正解率は 65%であった。更なるシステムの向上のために 3 人の被験者に、テキスト全体を見た場合と二文だけに情報を限定した場合の二通りについて実際に接続関係を判断してもらい、比較検討した。この結果から各接続関係の近さをビジュアル化し、さらに全ての接続関係を組み合わせて作成した SVM との相関から、我々が示した各接続関係の近さが妥当なものであることを示した。人間の接続関係の推定結果から得られた知見を基に今後のシステムの精度向上を図る。

キーワード 接続関係, 人手による判断, SVM

Analysis on Manual Identification of Inter-sentential Discourse Relation

Manami SAITO[†] Kazuhide YAMAMOTO[†] and Satoshi SEKINE[‡]

[†]Nagaoka University of Technology 1603-1 Kamitomioka, Nagaoka-shi, Niigata, 940-2188 Japan

[‡]New York University 715 Broadway, 7th floor, New York, NY, 10003 USA

E-mail: [†]{saito, ykz}@nlp.nagaokaut.ac.jp, [‡]sekine@cs.nyu.edu

Abstract We aim to make a system which identifies a discourse relation between sentences automatically. We previously proposed a method which identifies the discourse relation between two successive sentences using statistics given by a huge corpus, and the accuracy was 65%. In this paper, in order to improve the accuracy, we observe human identification results of this task. Three subjects identify the discourse relation by two ways: one uses whole text, the other uses only two sentences. Our observation shows the relationship for each discourse relation in visually. Moreover, relevance of the relationship was explored with comparing the result of SVM.

Keyword discourse relations, manual identification, SVM

1. はじめに

接続関係を同定することは談話解析やマルチドキュメント要約、質問応答など多くの応用分野において重要となる。そこで、先行研究として我々は与えられた二文の接続関係を同定するシステムを提案した[6,7]。これは入力として与えられた二文から単語や構文情報を取り出し、何もタグ付けされていない大量のコーパスを用いて統計的に適切な接続関係を選択する。システムは与えられた二文と同じ関係を持つと思われる文をコーパス中から探し、接続関係を同定する。このと

きのシステムの正解率はおおよそ 65%であった。

先行研究[6,7]では文頭の接続詞は直前の文との接続関係を示しているという前提のもとで行なわれていたが、その接続詞がテキスト中で常に直前の文との関係を示しているとは限らない。一つ前の文とのつながりを表しているかもしれないし、あるいは文ではなく段落ごとのつながりを表しているということも考えられる。このことから、我々は入力を二文に限定せず、テキスト全体から接続関係を判断することでシステムの精度向上につながると考えた。

システムの設定をどのようにすべきか、という問題もある。システムの評価においては、元のテキストで用いられている接続詞が属する接続関係を正解と定義している。しかし人間が接続関係を読み取る際に、人間の出力とこの正解とははずれが生じており、正解そのものを見直す必要もある。また、人間が判断する場合においても接続関係毎の難しさは存在するのだろうか。さらに、人間が接続関係を一つではなく複数読み取る場合のシステムが出力する接続関係は一つでよいのだろうか。

文間接続関係の自動同定のため、本稿では実際に人間に接続関係を判断してもらい、その結果を比較検討することでこれらの疑問に対する見解を求めシステム向上の糸口を探す。

2. 人手による接続関係の推定

3人の被験者(A,B,C)に育空文庫(1)より旧字体のものを除いてランダムに選んだ23テキストを与え、文頭に接続詞をもつ文を対象に300個の接続部分を空欄として穴埋め形式で適当な接続関係を選んでもらった。被験者にはあらかじめ指定された接続関係と、その接続関係に属する代表的な接続詞を教えておく。テキストの長さにはかなりばらつきがあり、短いテキストでは1テキスト中に穴埋め箇所は2箇所、長いテキストでは43箇所あるものもある。このとき選ぶ接続関係は指定した6種類のうちいくつ選んでもかまわないが、複数選択する場合には優先順位をつけてもらうことにする。ここで指定した6種類の接続関係は先行研究[7]にて使用したものと同じものを定義している。ただし、先行研究[7]において複数の接続関係に分類される接続詞は本実験では除いている。ここで取り除いた接続詞は1割程度である。表1に各接続関係と今回使用した問題300問中の割合を示す。

表1. 接続関係と接続詞の例

接続関係	代表的な接続詞の例	割合[%]
累加	また、そして、しかも…	24.7
加反	しかし、でも、ところが…	41.3
因果	だから、なので、すると…	10.0
並列	つまり、すなわち、それとも…	8.67
転換	さて、ところで、では…	4.67
例示	例えば	10.7

また、実験はテキスト全体を与えた場合と穴埋め箇所の前後一文ずつを取り出した二文だけを与えた場合の二通り行った。最初にテキスト全体での推定をしてもらい、その3日後に二文だけを用いた場合の実験を行った。二文だけを用いた実験では、テキスト全体での出力結果を被験者が思い出さないように全てのテキストから取り出した二文のペアをランダムに並べ替えて提示した。これらの結果を比較し、人間の接続関係の判断についての見解を以下に述べる。

2.1. 正解率の比較と一致率

各被験者のテキスト全体を見て判断した場合(Text)と二文の情報のみで判断した場合(2文)での正解率と、二種類の出力の一致率を表2に示す。被験者には優先順位をつけて複数選

択してもらっているため、それを考慮するため質問応答でよく使われているMRR(Mean Reciprocal Rank)の評価手法を応用して正解率を求めた。具体的にはN番目の出力が正解した場合に1/Nのポイントを与え、その合計を全問題数300で割った値を正解率としている。

表2. 各正解率と二つの出力の一致率

被験者	Text	2文	一致率
A	0.604	0.607	0.631
B	0.508	0.488	0.512
C	0.628	0.622	0.645

我々は、テキスト全体を見たときの方が二文だけを見て判断するより精度が良くなるという予測のもとに実験を行った。しかし人間が接続関係を推定するときにはテキスト全体を見て判断した場合と二文のみで判断した場合の正解率にあまり差はなく、我々の期待と反した結果となった。

しかし、テキスト全体を見て判断した場合の人間の出力と二文のみで判断した場合の出力の一致率はおよそ6割と高くない。つまり正解率は変わらないが正解している問題は少なからず異なっているといえる。このことから、テキスト全体を見た方が正しく判断できるものと逆に情報を二文に限定した方がよいものがそれぞれ存在するという仮説が立てられる。

2.2. 接続関係毎の正解率

テキスト全体を見た場合と二文だけを見た場合の正解率の異なりを接続関係毎に分析し、接続関係の持つ特徴を捉える。図1、図2にそれぞれ3人の被験者の平均による接続関係毎の正解率と適合率を示す。

図1、図2より「転換」においては二文だけで判断するよりテキスト全体を見たほうが正解率と適合率が高い。特に「転換」という接続関係は話題の移り変わりを表すので、直前の一文だけでは判断できないと考えられる。また、「加反」や「例示」はテキスト全体でも二文でもほとんど差がなく、しかも他の接続関係に比べて正解率、適合率共に高い。このことから、「加反」や「例示」は直前の文とのつながりを表しやすいため。実際にシステムを作る際にはこのような接続関係のもつ特徴も考慮する必要がある。このとき、場合によって分類手法を変えるなどの対応が考えられる。

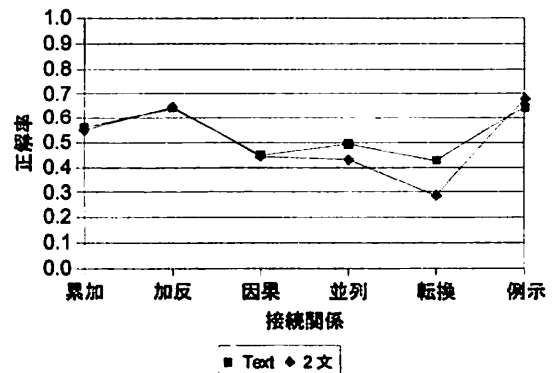


図1. 接続関係毎の正解率の平均

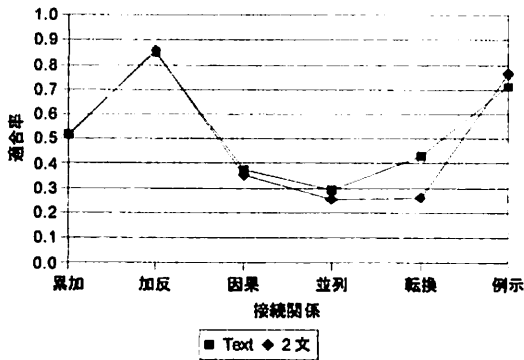


図2. 接続関係毎の適合率の平均

2.3. テキストと二文の出力の一致度と正解の関係

テキスト全体を見て判断したときと二文だけを見て判断したときの接続関係が一致した問題数と、そのときの正解率から考察を行う。ここでは、優先順位の最も高い接続関係を被験者の唯一の出力としている。

表3は各被験者の出力結果の内訳である。表の○、×はそれぞれ正解、不正解を示しており、aとbは二つの実験における被験者の出力結果が一致した場合、c~eは不一致の場合を表している。また、表4は二文だけで判断したときには誤っているがテキスト全体を見た場合では正解している問題と、逆にテキスト全体では誤っているが二文だけで判断したときには正解している問題の接続関係毎の割合を示している。

表3. 出力の一致からみた人間と正解の関係

	出力	Text	2文	A	B	C	Total
a	一致	○	○	140	101	150	391
b		×	×	52	61	45	158
c	不一致	○	×	38	48	36	122
d		×	○	42	39	36	117
e		×	×	28	51	33	112

表4. 一方のみ正解している問題

接続関係	Textで正解		2文で正解	
	問題数	割合	問題数	割合
累加	35	0.29	30	0.26
加反	39	0.32	45	0.39
因果	18	0.15	18	0.15
並列	13	0.11	9	0.08
転換	8	0.07	3	0.03
例示	9	0.07	12	0.10
Total	122	-	117	-

表4より、テキスト全体を見た場合に正解した問題と二文だけを見た場合に正解した問題に多少の偏りは見られる。しかし、この結果からは接続関係によってテキスト全体を見た方が正しく判断できるものと情報を二文に限定した方がよいものに分けることはできない。

今回のようなタスクで長い文章から接続関係を判断するとき、まず人は二文だけを見て決めようとする。そこでうまく接続関係が決められない場合には見る範囲を広げていく。しかしそこにテキストがある以上、二文だけで判断可能な場合でも人は他の文章の影響を受けずにはいられない。そのためもし二文だけで接続関係を判断するための情報が十分含まれているとしたら、テキスト全体を与えることでかえって判断を迷わす結果となっていると考えられる。また、接続関係を定めるための情報が二文では不十分であるときは人間の判断も自信のないものとなっている可能性が高い。システムにおいても二文で判断した際の信頼度によって見る範囲を変えられれば、より精度の高い文章理解が得られるであろう。

また表3において、テキスト全体を見た場合と二文だけを見た場合の出力が一致した問題(a,b)中での正解率(a/(a+b))は62~77%であり、3人の平均では71%であった。表2に示す3人の被験者それぞれの正解率と比較するとおよそ1割強高い。これは、全体の正解率と比べて出力が一致している問題中での正解率の方が高くなるであろうという人間の直感とも一致する。しかし全体の17%の問題(b)については正解とは異なった関係を選んでいるにもかかわらず、テキスト全体を見て判断した接続関係と二文だけで判断した接続関係が一致していることから、人は迷いなくその関係を選んでいると考えられる。出力がどちらも同じであるということは、接続関係を判断するためには二文の情報だけで十分であるといえる。では人はなぜ本来の接続関係と異なった関係を選んでいるのだろうか。これについては被験者全員が一致して誤った接続関係を出力した例をもとに2.6節で考察する。

2.4. テキストと二文を組み合わせたときの正解率

表5では二つの正解率とともに、テキスト全体を見た場合と二文だけで判断した場合の結果を組み合わせたときの正解率を示す。ここでは、テキスト全体を見た場合と二文だけで判断した場合の両方の出力のうち少なくとも一方に答えが含まれていれば正解としている。正確な比較を行うため、テキスト全体を見た場合と二文だけで判断した場合における正解率も、選択された接続関係の中に答えが含まれていれば優先順位に関係なく1ポイントとしている。

表5. テキストと二文の組み合わせによる正解率

被験者	Text	2文	組み合わせ
A	0.613	0.607	0.747
B	0.520	0.510	0.663
C	0.637	0.623	0.747

全ての被験者において出力を組み合わせたときの正解率がテキスト全体、または二文のみを判断材料として限定したどちらの場合の正解率よりも15%程度高くなっている。このことから、システムが接続関係を判断する場合においても両者を組み合わせることが必要であるといえる。問題によってテキスト全体を見るべきか、二文だけで判断すべきかの区別ができれば大幅に精度の向上が期待できる。

2.5. 複数選択

表2と表3では正解率の計算方法を変えているが、それらの正解率にはあまり差がみられない。つまり、タスクとして接続関係の複数選択を可能にしたことによる影響は少ない。3人の被験者の出力の総数と問題数の割合（選択した接続関係の和/300）を表6に示す。

表6. 3人の被験者の出力の総数と問題数の割合

	A		B		C	
	Text	2文	Text	2文	Text	2文
総数	318	302	349	340	312	305
総数/300	1.06	1.01	1.16	1.13	1.04	1.02

個人差はあるが、一つの問題に対して複数の接続関係が得られるとしている問題は、多い人でも全体の15%程度である。少ない人では数問程度で、ほとんどの問題に対して出力は一つである。

人間が複数の接続関係を読み取るとき、その文の間にはそもそも複数の正解が存在しているという仮説が立てられる。もしこの仮説が正しければ、システムも複数選択に対応させる必要がある。しかし、人間が複数の接続関係を出力することはほとんどなく、3人の被験者の出力結果を比較してみても共通の問題で複数の同じ接続関係を選択している例はなかった。そのため、接続関係を自動的に判定する場合でも出力は一つとしてよいといえる。

直感的には与えられる情報が二文に限定されている方が接続関係の判断に迷い、複数選択することが多くなると思われるが、表6を見ると被験者全員二文だけを見た場合の方が複数の接続関係を選択することが少ない。この原因に関しては二つの実験において差が少ないことから一概に述べることはできないが、情報を限定することによって曖昧性が生じるといえる。

2.6. 人間の評価と書き手の示す関係との違い

例1はテキスト全体を見た場合と二文だけを見た場合で出力が一致し、なおかつ答えとは異なる接続関係が選ばれたときのテキストの抜粋である。空欄には「すなわち」という「並列」の接続詞が入るが、被験者は全員「例示」とだけ答えた。

例1. 書き手と読み手の違い

例えば、甲の眼にはAとBとが、よく似ているように見える。ところが、乙に云わせると、ちっとも似ていないじゃないかと云う。これは甲と乙とで着眼点がちがうためだと云えばそれまでである。□、甲にとってはAとBとの二人の顔の中で、例えば眼だけが注意の焦点となるのに、乙には眼はそんなに問題にならないで口許が特に大切な特徴となって印象される、という場合がそれである。

例1で、空欄の直前の文の内容は「甲と乙の着眼点は違う」ということである。ここでは甲と乙の着眼点が違うということとを次の文で例をあげながら言い換えている。書き手として

は対象の文は直前の文で言っている「甲と乙の着眼点が違う」ということがどういうことを言い直しているだけで、二つの文は同等のことについて述べているため、この二文の関係は「並列」であるとしている。しかし、読み手としては具体的な例を述べているのだから接続関係としては「例示」とであると判断したわけである。だがこの場合被験者らの出力が誤りであるとはいえない。むしろ読み手が一致して同じ関係を導くのなら、システムの正解としてはそちらを起用すべきである。

同様に被験者全員が一致して誤った例として、テキスト全体を見ても二文だけを見ても「転換」であったとしたもの文章の一部を例2に示す。元の文章では空欄には「加反」の接続詞「しかし」が入る。

「しかし」という接続詞は多くの場合前の文の内容に反した事柄を後の文で述べる「加反」の関係を表すが、時には話題を変える「転換」の場合にも用いられる。現在は接続関係を元のテキスト中で使われている接続詞が属する関係に自動的に振り分けているため、このように正解と人間の出力に違いが出てくる。

例2. 接続詞が意味する接続関係の曖昧性

…彼女は婦人雑誌でいう教養も、話題に困らないほどはゆたかで、女として人間として活潑であるものの面白さを自分の身に添えて表現する技術も理解し、わきまえている近代的な若い婦人であるということが出来よう。
□、この頃よくあるように、家庭のある男の人と恋愛めいたいきさつがはじまったようなとき、あるいは、そういういきさつが自分の人生に起りかかっているのを自覚したとき、そのひとは、現実の人生問題をゆたかな文学的教養とむすびあわせて、どのように身を処していくだろうか。トルストイの小説の中には、…

テキスト全体を見た場合と二文だけで判断した場合とで出力が同じであるとき、元のテキストで与えられた接続関係と違うものを人間が選んでいたとしてもそれが不正解であるとは言いがたい。現在のタスクでは元のテキストでの接続関係と一致していなければ不正解としているが、これらの例からもわかるように、元の関係でなくとも文として関係の成り立つものは他にも存在する。2.5節で述べているように人間は複数の選択肢で迷うことはあまりないが、だからといって正解が常に一つであるとは限らない。接続関係の正しさを評価する場合、その問題が複数の解を持つならどれを選んででも正解とすべきである。

2.7. 接続関係の距離

ある接続関係に対して人間がどの接続関係と判断しているか、接続関係毎にまとめ、表7に示す。本節ではすべて優先順位の最も高い接続関係を被験者の唯一の出力としている。被験者ごとに数値の揺れはあるが全体的な傾向は同じであるため、ここではテキスト全体を見て判断した場合の被験者全員の結果を合成したものを示す。また、全ての合計が $300 \times 3 = 900$ となっていないのは、被験者Aが指定された接続関係

に適切なものがないと判断した問題が一間あったためである。表 8 は表 7 よりそれぞれの接続関係のペア毎に求めた人間と正解の κ 値である。

これらの結果から、元のテキストにおいて本来「加反」であるものは「例示」とは認識されにくく、逆に「例示」が「加反」となることもほとんどないといえる。このように関係の遠いものはこの他に「並列」と「転換」、「因果」と「転換」、「因果」と「例示」、「累加」と「例示」がある。また逆に関係の近いものは、「累加」と「並列」、「累加」と「因果」である。

表 7. 各接続関係における人間の出力結果と正解

		人間の出力						Sum1
		累	加	因	並	転	例	
正解	累	122	21	28	44	2	4	221
	加	58	236	34	22	18	4	372
	因	21	9	40	15	0	5	90
	並	21	5	4	37	0	11	78
	転	6	8	5	4	17	2	42
	例	14	1	5	9	5	62	96
Sum2		242	280	116	131	42	88	899

表 8. 各接続関係間の κ 値

	加反	因果	並列	転換	例示
累加	0.615	0.454	0.330	0.778	0.805
加反		0.571	0.680	0.516	0.951
因果			0.609	0.814	0.814
並列				0.844	0.648
転換					0.776

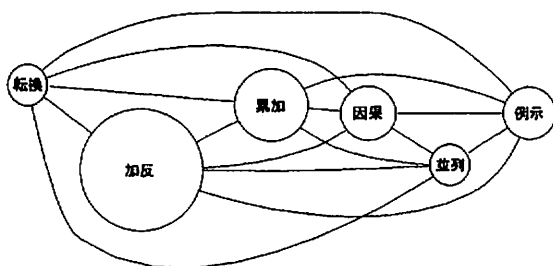


図 3. 各接続関係の近さ

これらの結果による接続関係の近さを図 3 に示す。関係の近さの表示には Graphviz(3) を用い、表 8 に示す κ 値をノード間を結ぶエッジの weight として与えた。この weight が高いほど短く、直線的なエッジとなる。今回使用したデータの各接続関係の割合によって円の大きさを変えている。

特に分類が難しいのは「累加」とその周りに位置する「並列」、「因果」、「加反」である。それらの関係と「累加」の境界はあいまいで、そのまま 6 分類の中からひとつの関係を選

ぼうとすると分類は非常に難しい。先行研究[7]においても「累加」の正解率は最も低かった。しかし我々は、図 3 に示す各接続関係の近さを利用して SVM 等の二値分類器を組み合わせることで、関係の近いものでも分類が可能であると考えられる。それぞれの分類器の結果に対して分類器毎の信頼度を付与し、それらを総合することで最も信頼度の高い接続関係を選ぶことが可能である。

表 7 は被験者がテキスト全体を見て判断した場合の結果であるが、二文だけで判断した場合も傾向は同じであった。両者を比較した場合の被験者の傾向として、二文に限定した場合、他の接続関係に比べ「加反」を出力することが多くなり、「転換」の出力数が減少したことがあげられる。このことから 2.2 節で述べたように、「加反」は二文間の接続関係を表しやすく、「転換」は二文間の関係を表すことがあまりないという傾向が読み取れる。

3. SVM による予備実験

2.7 節で述べたように、SVM を用いて表 8 や図 3 に表される接続関係間の関係の近さを利用して多値分類を行う際、SVM で人間と同じような傾向がつかめるかどうかの問題である。全ての接続関係の組み合わせに対する SVM(2) の一対一学習の精度の比較からこの問題を調査する。まずテキスト中で接続詞によってつながっている二文のペアを接続関係毎に 500 セットずつ用意した。全ての接続関係の組み合わせ ($C_2=15$ 通り) に対して SVM を用意し、訓練データは正例負例の合計 1000 セットを各 SVM に対して用いた。テストセットでは各分類器において正例負例 100 セットずつの 200 セットを用いた。訓練データおよびテストセットには新聞記事を使用している。SVM の素性として「固有名詞以外の名詞」、「動詞」、「形容詞」を出力形のまま使用した。また、同じ単語であっても一文目と二文目に含まれる単語はそれぞれ区別している。

表 9 に全ての接続関係の組み合わせに対する SVM の結果を示す。

表 9. SVM による各二値分類の精度

	加反	因果	並列	転換	例示
累加	0.550	0.630	0.595	0.670	0.640
加反		0.625	0.710	0.695	0.750
因果			0.575	0.660	0.715
並列				0.610	0.655
転換					0.690

表 8 と表 9 の結果を比較すると、ほぼ同じ傾向を示していることがわかる。図 4 に人間の誤りやすさと SVM の誤り率の相関を示す。このときの相関係数は 0.51 であった。これにより、簡単な SVM を用いても人間が判断した結果に対して比較的高い相関が得られることがわかる。SVM の実験では A と B のどちらの関係であるかという二値分類であったため、選択肢は常に二つであったが、実際に被験者に判断してもらったときは 6 種類の接続関係から選んでもらった。そのため人間の各接続関係間の κ 値と SVM の精度は正確には対応し

ていないことになる。しかしこの二つの相関を見ることによって SVM を用いた学習においても図 3 に示したような各接続関係の「関係の近さ」を確認することはできる。

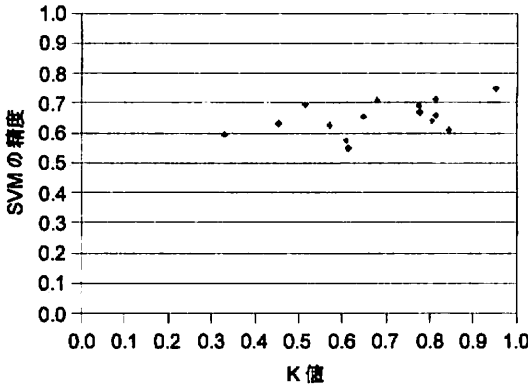


図 4. κ 値と SVM の精度の相関図

これらの結果を実際に 6 分類のタスクに適用させるには工夫が必要であるが、「転換」、「例示」は他の接続関係と距離が離れているため比較的分類しやすい。そのため、多値分類器等で「転換」、「例示」、「その他」の 3 値分類をしてから残りの 4 種類の接続関係をそれぞれの関係との二値分類による多数決等で決定するという方法も考えられる。今回は「固有名詞以外の名詞」、「動詞」、「形容詞」の 3 種類を素性として用いたが、SVM を使用するのであれば、これらの素性についてももう少し考える必要がある。SVM の精度向上も含めて、どのように 6 分類に適用させるかは今後の課題となる。

4. 考察

今回の人手による実験では、テキスト全体での推定をもらった 3 日後に二文だけを用いた場合の実験を行った。テキスト全体での出力結果を思い出させないように二文のペアをランダムに並べ替えて提示したものの、被験者が先に行ったテキスト全体での実験の影響を受けていないか不安が残った。しかし、二つの実験結果ではかなり異なるものが得られたので、結果的に本稿の狙いは満足したものと考えている。

また今回は接続関係を 6 種類に分類して実験を行ったが、この接続関係の設定によって本稿で述べたような関係毎の傾向は異なってくるのが考えられる。2.6 節でも触れているように、「加反」に属す「しかし」という接続詞は「転換」の意味を含む場合があり、このように接続関係 A に属す接続詞が接続関係 B の意味も含んでおり、今回の 6 分類ではその切り方が悪かった等、接続関係の設定におけるそもそもの問題も存在する。しかし分類の数が 6 種類で適切であるかどうかは接続関係の同定結果を何に適用するかという目的によって異なってくる。今回の実験では先行研究[7]を受けて 6 分類とした。また、分類数が多くなればなるほど機械的な処理は難しくなってくる。我々は多くのタスクに適用可能で、比較的分類可能と思われるラインとして 6 分類を採用している。

5. まとめ

3 人の被験者にテキスト全体を見た場合と二文だけを見た場合の二通りについて適当な文間の接続関係の推定を行ってもらった。被験者には適当だと思われる接続関係を、優先順位をつけて複数選択してもらったが、実際に複数の出力が得られた問題は少なかった。テキスト全体を見た場合と二文だけを見た場合の出力の一致率、正解率から、二文だけに情報を限定した方が正解できる場合とテキスト全体の情報が必要となる場合があることがわかった。接続関係毎にみると、比較的「加反」、「例示」は二文の情報のみで十分であるが「転換」は二文だけでは判断しにくい接続関係であるといえる。

また、今回設定した接続関係には関係の近さに違いがあることを認識し、各接続関係の近さを図示した。さらに人間の出力と正解との接続関係のペア毎の κ 値と、全ての接続関係の組み合わせに対する SVM の精度の比較から、我々が提示した接続関係の近さを SVM によっても確認できた。多くの接続関係を同定するために、この関係の近さがどのように利用できるかについても考察した。今後の課題としては、これらの関係の近さを考慮した 6 分類のシステムの構築と SVM の精度向上などがあげられる。

使用した言語資源およびツール

- (1) 青空文庫, <http://aozora.gr.jp>
- (2) SVM 学習ツール, "Tiny SVM", Ver.0.09, 奈良先端科学技術大学院大学 松本研究室, <http://chasen.org/~taku/software/TinySVM/>
- (3) Graphviz, <http://www.graphviz.org>

文 献

- [1] B. Hutchinson, "Acquiring the Meaning of Discourse Markers," Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, pp.684-691, Barcelona, Spain, July 2004.
- [2] C. Sporleder, and A. Lascarides, "Exploiting Linguistic Cues to Classify Rhetorical Relations," Proceedings of Recent Advances in Natural Language Processing, pp.532-539, Borovets, Bulgaria, September 2005.
- [3] D. Marcu, and A. Echihiabi, "An Unsupervised Approach to Recognizing Discourse Relations," Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp.368-375, Philadelphia, USA, July 2002.
- [4] F. Wolf, and E. Gibson, "Representing Discourse Coherence: A Corpus-Based Study," Proceedings of Computational Linguistics, vol.31, no.2, pp.249-287, June 2005.
- [5] M. Lapata, and A. Lascarides, "Inferring Sentence-internal Temporal Relations," Proceedings of the North American Chapter of the Association of Computational Linguistics, pp.153-160, Boston, USA, May 2004.
- [6] M. Saito, K. Yamamoto, and S. Sekine, "Using Phrasal Patterns to Identify Discourse Relations," Human Language Technology conference - North American chapter of the Association for Computational Linguistics annual meeting, pp.133-136, New York City, USA, June 2006.
- [7] 齋藤真実, 山本和英, 関根聡, "大規模テキストを用いた 2 文間接続関係の同定," 言語処理学会第 12 回年次大会, pp.969-972, March 2006.