

単語概念の深層格選好に基づく深層格推測手法

渋木 英潔^{†*} 荒木 健治^{††} 桃内 佳雄^{†††*} 柝内 香次^{††††}

A Method for Inference of Deep Case Based on Deep Case Preference
of Word Concept

Hideyuki SHIBUKI^{†*}, Kenji ARAKI^{††}, Yoshio MOMOUCHI^{†††*},
and Koji TOCHINAI^{††††}

あらまし 本論文では、構文解析済みの単文に対して深層格を推測する手法を提案する。推測に用いる知識は、網羅性とコストの問題に対処するため、一定量のタグ付きコーパスから得られた知識を中核として、タグなしコーパスから規則として学習する。本手法は、単語の概念ごとに、それぞれ特定の深層格に解釈される傾向（深層格選好）が存在するという仮定に基づき、一定量のタグ付きコーパスから深層格選好を計算した後、その値を用いてタグなしコーパスから多様な言語表現に対処するための規則を学習する。深層格選好を手掛りの主体とすることにより、本手法は言語の違いに依存せずに深層格を推測することが可能となる。言語非依存性を検討するため、日本語と英語の EDR コーパスを用いて、一つの動詞と二つの名詞で構成される単文を対象に実験を行った。その結果、クローズドデータにおける日本語の精度 81.2%（再現率 78.0%）、英語の精度 78.5%（再現率 77.8%）、オープンデータにおける日本語の精度 69.5%（再現率 62.1%）、英語の精度 73.5%（再現率 73.3%）となり、従来手法と同程度の精度で日本語と英語の深層格を推測することができた。

キーワード 深層格推測, 意味解析, 深層格選好, 機械学習, 単語クラスタリング

1. ま え が き

機械翻訳における曖昧性の解消などのために、文の骨子を表現する手段の一つとして、深層格を利用することが考えられる。深層格を推測するための知識を人手で与えることは、網羅性やコストの点で問題が残るため、コーパスから学習する手法が望まれている。このような背景から、大石らの手法 [1] をはじめとして、文献 [2] ~ [7] などに挙げられる手法が提案されている。

これらの手法では、すべての学習データに深層格のタグが付与された文を用いており、タグが付与されていないデータから学習する点に関して考慮されていない。人手によるタグの付与にはコストがかかるため、コスト軽減という観点からは、タグなしの文から学習できる機構を備えた手法が望ましい。そこで我々は、一定量のタグ付きコーパスから得られた知識を中核として、タグなしコーパスから深層格推測の知識を学習する手法を提案する。これにより、多大なコストをかけずに網羅的に知識の拡充を図ることを試みる。

我々も従来手法の多くと同じく、深層格を推測するためには、最終的に助詞や語順などの表層表現と対応づけられた知識が必要であると考えている。しかしながら、係助詞のように深層格を決定できない表層表現や、「水が飲みたい」のように格助詞であっても共起する動詞により一般に解釈される深層格と異なる場合など、深層格と表層表現との対応関係は複雑であり、人手で網羅的に与えることは困難であると考えられる。この問題は、タグ付きデータを作成、利用する場合にも生じる問題である。我々は「水が飲みたい」と「水を飲みたい」が同じ深層格と判断されるのは、「水」と

[†] 北海学園大学大学院経営学研究科, 札幌市
Graduate School of Business Administration, Hokkai-Gakuen University, 4-1-40 Asahi-cho, Toyohira-ku, Sapporo-shi, 069-0813 Japan

^{††} 北海道大学大学院情報科学研究科, 札幌市
Graduate School of Information Science and Technology, Hokkaido University, Kita 14, Nishi 9, Kita-ku, Sapporo-shi, 060-0814 Japan

^{†††} 北海学園大学工学部, 札幌市
Faculty of Engineering, Hokkai-Gakuen University, Minami 26, Nishi 11-1-1, Chuo-ku, Sapporo-shi, 064-0926 Japan

^{††††} 北海学園大学経営学部, 札幌市
Faculty of Business Administration, Hokkai-Gakuen University, 4-1-40 Asahi-cho, Toyohira-ku, Sapporo-shi, 069-0813 Japan

* 現在, 北海学園大学ハイテク・リサーチ・センター

「飲む」という名詞と動詞の概念によるものであると考え、この傾向は言語の種類を問わずに普遍であると仮定した。このような考えから、提案手法では、名詞や動詞の概念ごとに、それぞれ特定の深層格に解釈される傾向が存在するという仮定に基づき、タグ付きコーパスから深層格の傾向を計算した後、その傾向を用いてタグなしコーパスから多様な言語表現に対処するための知識を学習する。この深層格の傾向を深層格選好と呼び、深層格選好に基づいて深層格を推測する本手法を DCAPR (Deep Case Analysis based on deep case Preference and Regularization) と呼ぶ。単語概念の深層格選好を手掛りとすることで、表層表現の違いを事前に考慮することなく学習することを可能としている。本論文では、言語における表層表現の差異を取り上げ、日本語と英語を対象として実験を行う。

2. で DCAPR の概要を説明し、3. で DCAPR が扱う知識表現を記述する。その後、4. で処理の詳細を述べる。5. で日本語と英語を対象とした実験を行い、6. で DCAPR の言語非依存性について考察を行う。7. では、学習データ量と精度の関係や分野の違いによる影響などの調査を行う。8. は結論である。

2. 概 要

2.1 基本的な考え方

本手法は、人手による労力の軽減という目的から、機械的に処理できる情報から深層格を推測することを試みている。それゆえ、深層格は、格支配される名詞、格支配する動詞、格支配される際の表層表現によって推測されると仮定した。本論文では、名詞と動詞を手掛りとした知識を深層格選好、表層表現を手掛りとした知識を深層格推測規則と定義する。

まず、名詞と動詞の深層格選好について説明する。例として「太郎」と「東京」の二つの名詞が、ある文の中で使われた場合に、これらの名詞が担うであろう深層格を考える。文脈によって、どちらの単語も様々な深層格を担うことが可能であるが、一般的に agent を担う傾向が強いのは「東京」よりも「太郎」の方である。同様に、place を担う傾向が強いのは「東京」の方と考えられる。また、これらの傾向は、動詞との関係を考慮することにより、更に顕著なものとなる。例えば「住む」という動詞が二つの名詞をとることを考えた場合、agent と place を担う二つの名詞を同時に取る傾向の方が、object と source を同時に取る傾向よりも強いと考えられる。したがって、これらの傾向

を用いることで、「東京、太郎、住む」のような助詞が欠落した文に対して「東京」が place、「太郎」が agent と判断し「東京に太郎が住む」と復元することが可能である。本論文では、この傾向を数値化したものを深層格選好と定義し、3.1 に述べる。

次に、深層格推測規則について説明する。深層格選好を手掛りとした推測は、一般的な傾向を反映したものであるため、多くの事例を正しく推測できる一方で、例外的な事例には対処できない場合があると考えられる。例えば「次郎を公園で見る」という文の推測を行う場合「次郎」「公園」「見る」の深層格選好に基づいて推測するため「次郎が公園で見る」と同じように推測される可能性が高い。この問題を解決するために、機能語や語順といった、深層格を推測する手掛りとなる表層的な表現を考慮する必要がある。しかしながら、1. で述べたように、表層表現と直接対応づける知識を人手で網羅的に与えることには問題があるため、タグなしコーパスから自動的に学習する仕組みが必要である。表層表現がどのように深層格と対応しているかを学習するために、本手法では、深層格選好による推測結果の多くは正しく、推測結果に共通した表層表現は深層格推測の手掛りとなるという仮定に基づいて学習を行う。例えば「花子が服を買う」と「部屋で本を読む」という二つの文が深層格選好に基づいて以下のように推測されたと仮定する。1 番目の文が「花子」「服」「買う」に基づいて、2 番目の文が「部屋」「本」「読む」に基づいてそれぞれ推測され「服」と「本」の推測結果が object と正しく推測されたとする。このとき、両方の文において object と推測された単語に共通した表層表現として「を」が得られることとなる。したがって、object を推測する表層表現として「を」を学習することにより「次郎を公園で見る」における「次郎」の深層格推測に役立たせることができる。本論文では、表層表現と深層格を対応づける知識を深層格推測規則と定義し、3.3 に述べる。

2.2 全体の流れ

本手法の全体の流れを図 1 に示す。学習データとテストデータはすべて構文解析済みの単文である。また、図中の番号は処理の順序を示しており、以下、番号に従って記述する。最初に (1) すべてのデータで使用される名詞と動詞のクラスタリングを、3.2 に述べる表層情報に基づいて行う。これは、タグ付き学習データに含まれない未知語の深層格選好を、作成したクラスタから類推するための処理である。表層表現に基づく

のは、人手による労力を軽減するため、機械的に処理できる情報に限定したためである。クラスタリングの詳細は、4.1 に述べる。次に、(2) 作成した単語クラスタを参照し、単語及び単語クラスタの深層格選好を、タグ付き学習データに付与された深層格タグに基づいて学習する。深層格選好は、ある深層格と解釈される傾向を確率として表現したものである。その後、(3) 学習された深層格選好をもとに、タグなし学習データの深層格を推測し、推測された深層格とタグなし学習データ中の表層表現とを対応づけることで深層格推測規則を学習する。本来であれば、深層格選好の学習に用いられるタグ付きデータとは別に、規則を学習するためのタグなしデータを用意することが望ましいが、本論文では利用可能であったデータ量の関係から、深層格選好学習に用いたデータからタグを除いたものを規則学習のデータとした。深層格選好学習と深層格推

測規則学習の詳細は、4.3 と 4.4 にそれぞれ述べる。最後に、(4) 作成されたクラスタ及び学習された深層格選好と規則に基づいて、テストデータの深層格を推測する。深層格推測の処理は 4.5 に述べる。

3. 知識表現

3.1 深層格選好

名詞と動詞の深層格選好の例を表 1 と表 2 にそれぞれ示す。名詞の深層格選好は、ある動詞に格支配される際にその名詞が担う深層格の傾向を表し、動詞の深層格選好は、ある名詞を格支配する際にその名詞が担う深層格の傾向を表す。深層格選好は、それぞれの深層格ごとに、その深層格と解釈される確率で表現される。また、学習データにおける単語のスパースネス問題に対処するため、「太郎」や「Tokyo」のような個々の単語ごとに深層格選好を学習するのに加えて、4.1 に述べるクラスタリングを行い、クラスタごとにも深層格選好の学習を行う。本論文では、日英における比較のために EDR コーパスを用いるため、対象とする深層格を EDR の概念関係子に基づくこととした。大石ら [1] と小山ら [7] においても、EDR の概念関係子に基づいていたため、それぞれが対象とした深層格を併せ、表 3 に示す 14 種類とした。

動詞は複数の名詞を格支配するため、動詞の深層格選好は表 2 に示すように、一つの深層格ではなく、同時に格支配する名詞群が担う深層格の組として表現する。深層格の組で表現する理由は、深層格の推測において、以下のような効果を期待するからである。便宜的に「行く」の深層格選好が、「太郎」が [agent] 東京に [goal] 行く」と「東京から [source] 大阪に [goal] 行く」のように、[agent,goal] と [source,goal] の組み合わせとしないとし、「札幌から小樽に行く」の推測を行うとする。このとき、「札幌」の深層格が source だと推測

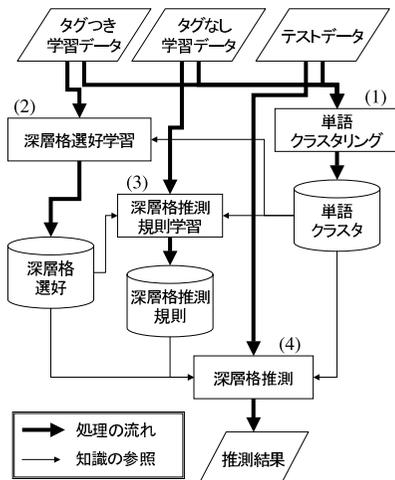


図 1 全体の流れ

Fig.1 Outline of the proposed method.

表 1 名詞の深層格選好の例

Table 1 An example of the deep case preference of noun.

NOUN	agent	object	goal	condition	...	quantity
太郎	0.8	0.0	0.2	0.0	...	0.0
Tokyo	0.0	0.0	0.4	0.0	...	0.0
(ClusterN1)	0.0	0.5	0.4	0.0	...	0.1

表 2 動詞の深層格選好の例

Table 2 An example of the deep case preference of verb.

VERB	agent,agent	agent,object	agent,goal	agent,condition	...	quantity,quantity
行く	0.0	0.0	0.6	0.1	...	0.0
live	0.0	0.0	0.0	0.0	...	0.0
(ClusterV1)	0.0	0.5	0.4	0.0	...	0.0

表 3 深層格の一覧
Table 3 A list of deep cases.

深層格	意味	例
agent	有意志動作を引き起こす主体	父が食べる
object	動作・変化の影響を受ける対象	りんごを食べる
goal	事象の主体または対象の最後の位置	東京に行く
condition	事象・事実の条件関係	雨が降ったので家に帰った
implement	有意志動作における道具・手段	ナイフで切る
material	材料または構成要素	牛乳からバターを作る
place	事象の成立する場所	部屋で遊ぶ
scene	事象の成立する場面	ドラマで演じる
source	事象の主体または対象の最初の位置	京都から来る
cause	原因	病気で倒れる
purpose	目的	映画を見に行く
basis	比較の基準	バラはチューリップより美しい
beneficiary	利益・不利益の移動先 [受益 [者] と被害 [者] の両方を含む]	父に買ってあげる
quantity	物・動作・変化の量	3kg 痩せる

表 4 深層格推測規則の例
Table 4 An example of the deep case inference rules.

RULE	agent	object	goal	condition	...	quantity
[が:F2][を:F1][2]	0.0	0.7	0.0	0.1	...	0.0
[φ:F1][in:B1][1]	0.9	0.0	0.0	0.0	...	0.0

できた場合、深層格単位で表現されていたとするならば、「小樽」の深層格として agent と goal が考えられるのに対して、深層格の組による表現ならば、source と組である goal と一意に決定することができる。この深層格の組を深層格パターンと定義する。本論文では、二つの名詞と一つの動詞からなる文を対象としたため、動詞の深層格は表 2 に示すように二つの深層格の組に対して表現される。

3.2 表層情報

本手法では、ある名詞の深層格を推測するための表層的な手掛りとして、その名詞に接続する機能語、その名詞の文中での位置、文全体で使用されているすべての機能語の並び、という三つの情報に着目する。第 1 の名詞に接続する機能語は、日本語においては助詞、英語においては前置詞とした。機能語が存在しない場合には空の文字列を代用とする。第 2 の文中での位置は、動詞を基準とした文節または名詞句単位での相対的な位置で表現する。例えば「太郎が本を読む」における「太郎」の位置は「読む」の二つ前に位置するので“F2”という略号で表記する。また、“Taro reads the book”における“book”は、“read”の一つ後ろに位置するので“B1”となる。位置を考慮することにより、「8時に札幌に行く」のような同じ機能語が連続する文の機能語を識別することが可能となる。第 3 のすべての機能語の並びとは、「太郎が本を読む」を例にとると [[F2:が][F1:を]] のように、それぞれの名詞

における機能語と位置の組を並べたものである。本論文では、これを機能語パターンと定義する。機能語パターンでは、位置を考慮するため、「太郎が本を読む」と「本を太郎が読む」では別の機能語パターン [[F2:が][F1:を]]、[[F2:を][F1:が]] となる。

表 3 に示した深層格が助詞などの表層情報とどの程度一意に対応しているかを調査するために、EDR 日本語コーパスを用いて調査を行った。助詞が省略されておらず、コーパス中で 3 回以上出現する助詞をもつ 5,508 の係受け関係を抽出し調査対象とした。また、抽出された係受け関係に含まれた助詞は 52 種類であった。その結果、ある助詞が同一の深層格に解釈される割合は平均 76.6%であったが、位置と機能語パターンを考慮することで平均 83.8%まで向上させることが確認できた。また、ある助詞が何種類の深層格と解釈され得るかの調査では、平均して 4.3 種類から 1.9 種類へ絞り込むことが可能となった。

3.3 深層格推測規則

3.2 で述べた表層情報と深層格を対応づけた知識を深層格推測規則とする。3.2 で述べたように三つの表層情報を考慮することで、深層格の推測候補を絞り込むことが可能となるが、1 対 1 まで絞り込めるわけではない。したがって、深層格選好と同様に、規則を適用した場合に解釈される深層格の傾向を確率として表現することとした。規則の例を表 4 に示す。三つの [] でくくられた規則の適用条件のうち、最初の二つは機

能語パターンを示し, 3 番目の [] は規則を適用する名詞の位置を表す. 表 4 の最初の規則は, 例えば「太郎が本を読む」において「本」の深層格を推測するために利用され, 次の規則は「I live in Tokyo.」において「I」の深層格を推測するために利用される.

4. 処 理

4.1 単語クラスタリング

本手法の単語クラスタは, 学習データに含まれない未知語の深層格選好を類推するために利用される. したがって, 深層格選好が類似した単語群を同じクラスタにまとめる手法が必要である. これまで, Caraballo [11] など, 表層的な情報を手掛りに, 意味的に類似した単語のクラスタリングを行う手法が提案されている. 3.2 で述べたように, 本論文の表層情報は深層格と比較的対応関係がとれているため, 表層情報を利用して深層格推測に特化したクラスタを作成することができると考えられる. このような背景から, 我々は, 3.2 で述べた表層情報に基づいて単語をクラスタリングする手法を提案しており, 日本語を対象とした深層格推測実験では, 分類語彙表 [12] を用いた結果よりも自動的に作成されたクラスタを用いた結果の方が高い精度であったことを確認している [9]. クラスタリングされる単語の定性的傾向は, 名詞, 動詞ともに, 類似した文における類似した構成要素として用いられる単語群がクラスタリングされることになる. したがって, 対象となる名詞に付属する表層情報の頻度分布が類似の傾向にあるだけでなく, 共起する名詞に付属する表層情報の頻度分布も類似の傾向にあることになる.

本クラスタリングは, 図 2 に示す流れで行われる. 最初に, 名詞と動詞それぞれの単語ベクトルを作成する. 単語ベクトルの要素は, コーパス中でその単語とともに出現した表層情報の頻度である. 表層情報は, 機能語, 位置, 機能語パターンの三つであるが, 動詞の場合は, 接続する助詞や前置詞がないため機能語の情報が存在せず, 位置情報も動詞を基準とした相対的な位置であるため有益な情報とならない. それゆえ, 動詞ベクトルの要素数は機能語パターンにのみ依存するとし, 名詞ベクトルの要素数は三つの表層情報の組合せに依存するとした. 本論文では, コーパスにおける頻出機能語上位 30 種類を用いてベクトルを作成した. したがって, 機能語パターンは二つの機能語の組合せであるため, 動詞ベクトルの要素数は $30 \times 30 = 900$, 二つの名詞を格支配する文を対象としているため, 名

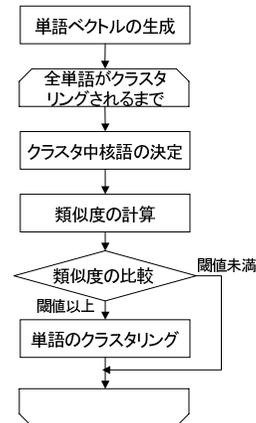


図 2 単語クラスタリングの流れ
Fig. 2 Flow of the word clustering.

詞ベクトルは, 機能語パターンに名詞の位置情報 2 通りを加えて $30 \times 30 \times 2 = 1,800$ となる.

次に, コーパス中での最頻出単語をクラスタの中核語として, 中核語との類似度がしきい値 T_1 以上となる単語を同一のクラスタにまとめる. 単語間の類似度は, Caraballo [11] と同様に, 二つの単語ベクトルの余弦で定義する.

$$\text{sim}(\vec{v}, \vec{w}) = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}| |\vec{w}|} \quad (1)$$

しきい値以上の単語をすべてクラスタリングした後, 残りの単語の中で最も出現頻度が高い単語を次のクラスタの中核語とし, 残りの単語の中でしきい値以上の類似度をもつ単語をクラスタリングする. 以上の処理を, すべての単語がクラスタリングされるまで繰り返す. クラスタリングのしきい値 T_1 は, 文献 [9] の結果から 0.8 とした.

4.2 ゆ う 度

2.1 に述べたように, 本手法の深層格 d は, 格支配される名詞 n , 格支配する動詞 v , 格支配する際の表層表現 g によって推測される. それぞれの要素から深層格 d となる確率 $P(d, n, v, g)$ を式 (2) のように近似する.

$$P(d, n, v, g) = P(d|n) \times \sum_{i \in D} P(d|v, i) \times P(d|g) \quad (2)$$

D は, すべての深層格の集合であり, 動詞における確率 $P(d|v, i)$ は, 深層格パターンを考慮するため, 共起する深層格 i も条件に含まれている.

しかしながら、深層格推測において、ある名詞句は必ず何らかの深層格を一つだけ担うことが前提であるため、 $P(d, n, v, g)$ の値によって、深層格を一意に決定することができないという問題がある。例えば、agent と object に対する $P(d, n, v, g)$ がともに 0.9 であるような場合には、どれほど高い確率値であったとしても一意に決定することはできず、逆に agent に対する $P(d, n, v, g)$ が 0.1 しかなくとも、object を含む他の深層格に対する $P(d, n, v, g)$ が 0.01 のように相対的に低い値であった場合には agent であると推測することが妥当であると考えられる。このような考えから、 $P(d, n, v, g)$ をベクトル要素として、単位ベクトル化することにより、 $P(d, n, v, g)$ 間の相対的な大小関係を表すゆう度 $pl(n, v, g, d)$ に変換することとした。ゆう度 $pl(n, v, g, d)$ は、以下の式 (3) により計算される。

$$pl(n, v, g, d) = \frac{P(d, n, v, g)}{\sqrt{\sum_{i \in D} P(i, n, v, g)^2}} \quad (3)$$

4.3 深層格選好学習

2.1 で仮定した深層格を推測する手掛りのうち、名詞と動詞の深層格選好を、タグ付きコーパスから以下の手順で学習する。ある名詞 n にタグ付きコーパス中で付与されている深層格 d の頻度を $f_{qN}(n, d)$ とする。以下の式 (4) に従って条件付き確率となるよう深層格選好 $P(d|n)$ を計算する。

$$P(d|n) = \frac{f_{qN}(n, d)}{\sum_{i \in D} f_{qN}(n, i)} \quad (4)$$

D は表 3 に示す深層格の全集合である。この値は単語ベクトルの各要素の頻度を正規化したものである。

動詞 v の場合も同様に、二つの深層格 d, d_2 の頻度を $f_{qV}(v, d, d_2)$ とし、以下の式 (5) に従って深層格選好 $P(d|v, d_2)$ を計算する。

$$P(d|v, d_2) = \frac{f_{qV}(v, d, d_2)}{\sum_{i, j \in D} f_{qV}(v, i, j)} \quad (5)$$

また、4.1 で作成した名詞クラスタ nc 及び動詞クラスタ vc の深層格選好 $P(d|nc)$ 及び $P(d|vc, d_2)$ を、以下の式 (6) 及び式 (7) に基づいて $f_{qNC}(nc, d)$ 及び $f_{qVC}(vc, d, d_2)$ を求めた後、同様にして計算する。

$$f_{qNC}(nc, d) = \sum_{i \in nc} f_{qN}(i, d)$$

$$P(d|nc) = \frac{f_{qNC}(nc, d)}{\sum_{i \in D} f_{qNC}(nc, i)} \quad (6)$$

$$f_{qVC}(vc, d, d_2) = \sum_{i \in vc} f_{qV}(i, d, d_2)$$

$$P(d|vc, d_2) = \frac{f_{qVC}(vc, d, d_2)}{\sum_{i, j \in D} f_{qVC}(vc, i, j)} \quad (7)$$

4.4 深層格推測規則学習

タグなしデータから、深層格と対応づけられる表層情報を学習するため、まず、深層格選好を手掛りとして、タグなしコーパスの深層格を推測する。この推測では、表層情報を用いることができないため、格支配される名詞と格支配する動詞の深層格選好のみを用いることとした。したがって、動詞 v に格支配されている名詞 n が深層格 d となる確率 $P(d, n, v)$ を式 (8) のように近似する。

$$P(d, n, v) = P(d|n) \times \sum_{i \in D} P(d|v, i) \quad (8)$$

4.2 に述べたように、ゆう度を $P(d, n, v)$ 間の相対的な大小関係として表現する必要があるため、ゆう度 $pl_L(n, v, d)$ を以下の式 (9) により計算する。

$$pl_L(n, v, d) = \frac{P(d, n, v)}{\sqrt{\sum_{i \in D} P(i, n, v)^2}} \quad (9)$$

このゆう度は、表層表現と対応づけられる深層格を推測するために用いられるものであり、ゆう度がしきい値 T_2 以上となった深層格を学習データにおける推測結果とする。このとき、複数の深層格のゆう度がしきい値以上となる場合や、どの深層格のゆう度もしきい値未満となる可能性があるため、推測結果が一意に定まるとは限らない。本論文では予備実験の結果から、しきい値 T_2 の値を 0.7 とした。ゆう度を単位ベクトルとしてとらえている本手法において、0.7 というしきい値は、推測結果を最大でも二つまでしか出力できず、基本的には一つの深層格を出力することを意味している。

深層格選好は個々の単語に関する知識であり、文単位における深層格の関係を考慮していない。文単位における深層格の関係を扱った原理として、一文一格の原理 (the one-instance-per-clause principle) が広く知られている [13]。一文一格の原理を考慮して学習することで、学習された規則は、深層格選好のみでは考慮できない文単位の関係を補う知識となることが期待できる。したがって、ある動詞に格支配されているすべての名詞の深層格が一意に推測され、かつ、推測された深層格が一文一格の原理を満たしている文を抽出

し、これらの文を用いて規則の学習を行う。抽出された文中の名詞 n に付属する表層情報を g とし、深層格 d が推測された頻度 $f_{qR}(g, d)$ を求める。深層格選好と同様に条件付き確率となるよう以下の式 (10) に基づいて計算し、表層情報 g を満たす名詞が d となる傾向を示す確率 $P(d|g)$ を深層格推測規則として学習する。

$$P(d|g) = \frac{f_{qR}(g, d)}{\sum_{i \in D} f_{qR}(g, i)} \quad (10)$$

4.5 深層格推測

最終的な深層格の推測は、4.3 と 4.4 で学習された深層格選好と深層格推測規則の両方を考慮して行われる。ゆう度を式 (2) と式 (3) に従って計算する。ただし、 n や v が学習データに存在しない場合には、 $P(d|n)$ や $P(d|v, d_2)$ の代わりに、クラスタ単位での深層格選好 $P(d, nc)$ や $P(d, vc, d_2)$ を用いて $P(d, n, v, g)$ を求める。 $P(d, n, v, g)$ は、 $P(d, n, v)$ と $P(d|g)$ の積で近似されているが、 $P(d, n, v)$ は、名詞と動詞の深層格選好の積であるため、結果としてすべての深層格に対するゆう度が 0 となる場合が考えられる。その場合には、どの深層格にも解釈され得るとして、 $P(d, n, v)$ の値を 1 として計算する。 $P(d|g)$ に関しても学習データから有効な値が学習できなかった場合が考えられる。その場合には、 $P(d, n, v)$ と同様に、 $P(d|g)$ の値を 1 として計算する。また、 $P(d, n, v)$ と $P(d|g)$ の積を計算した結果、すべての深層格に対するゆう度が 0 となった場合には、表層情報による規則を優先し、 $P(d|g)$ の値を最終的なゆう度とする。4.4 で述べたように、しきい値 T_2 以上のゆう度となった深層格を推測結果とした。

5. 日英比較実験

本手法が、対象となる言語に依存せずに深層格を推測できることを確認するために、EDR の日本語コーパスと英語コーパスを用いた実験を行った。実験で使用した文は、日本語と英語それぞれに対して以下の手順で抽出した。コーパスから表 3 に示す深層格を含む意味フレームを抽出する。二つの名詞を格支配している意味フレームの単語をもとに、構文木から対応する文節または句を特定し、二つの名詞と一つの動詞で構成される文を再構築する。この際、「れる」や「せる」などの深層格に変化を及ぼす助動詞を含む文、及び、be 動詞 + “-ed” などの受動態の文を除外した。再構築

された文の中で、それぞれの文を構成する、名詞、動詞、機能語の出現頻度がすべて 3 以上の文のみを実験で使用する文とした。最終的に、日本語 2,492 文、英語 7,485 文が抽出された。日本語のデータに関しては量的に十分ではないため、11 分割交差検定法による実験を行った。英語のデータに関しては、抽出された文の中から、11 分の 1 に相当する 681 文をランダムに選択しテストデータとし、残りの 6,804 文を学習データとした。日英ともに学習データは同一の文を 2 組用意し、一方は深層格推測規則学習のために深層格のタグを取り除き、もう一方は深層格選好学習のためにタグをそのまま残した。

推測結果の評価は、名詞単位で行い、以下の式に従って再現率、精度、F 値を求めた。

$$recall = \frac{NCO}{NA} \quad (11)$$

$$precision = \frac{NCO}{NO} \quad (12)$$

$$F \text{ value} = \frac{2 \times recall \times precision}{recall + precision} \quad (13)$$

NA はコーパス中で付与されている深層格の総数であり、 NO は推測された深層格の総数である。 NCO は、推測された深層格が正解であった数である。正誤の判断として、EDR コーパスで付与されている概念関係子と一致している推測結果を正解とした。

本手法は、人手による労力の軽減を目的とした Minimally supervised learning であり、深層格推測を対象として、Minimally supervised learning を用いた従来手法は存在していない。したがって、Supervised learning である、正解タグを用いて深層格推測規則を学習する手法をベースライン手法として設定し、提案手法がどの程度 Supervised learning に迫れるかという観点から評価することとした。

6. 結果と考察

6.1 日英間の比較

日本語と英語における実験結果を表 5 と表 6 にそれぞれ示す。クロードデータである学習データに対する結果を上段に、オープンデータであるテストデータに対する結果を下段に記している。また、本手法の推測は、式 (3) に示すように、深層格選好 $P(d, n, v)$ による推測と、深層格推測規則 $P(d|g)$ による推測に大別できるため、それぞれ単独で推測した場合の結果についても、ベースライン手法による結果とともに併

表 5 日本語における実験結果
Table 5 Result of Japanese sentences.

		対象数	出力数	正解数	精度	再現率	F 値
クローズドデータ	提案手法	49,840	47,840	38,858	81.2%	78.0%	79.6%
	深層格選好のみ	49,840	48,034	35,374	73.6%	71.0%	72.3%
	深層格推測規則のみ	49,840	46,046	35,478	77.1%	71.2%	74.0%
	ベースライン手法	49,840	48,234	39,783	82.5%	79.9%	81.1%
オープンデータ	提案手法	4,984	4,458	3,097	69.5%	62.1%	65.6%
	深層格選好のみ	4,984	4,703	2,970	63.2%	59.6%	61.3%
	深層格推測規則のみ	4,984	4,249	3,117	73.4%	62.5%	67.5%
	ベースライン手法	4,984	4,427	3,251	73.4%	65.2%	69.1%

表 6 英語における実験結果
Table 6 Result of English sentences.

		対象数	出力数	正解数	精度	再現率	F 値
クローズドデータ	提案手法	13,608	13,487	10,589	78.5%	77.8%	78.2%
	深層格選好のみ	13,608	13,389	10,047	75.0%	73.8%	74.4%
	深層格推測規則のみ	13,608	13,303	10,006	75.2%	73.5%	74.4%
	ベースライン手法	13,608	13,282	10,487	79.0%	77.1%	78.0%
オープンデータ	提案手法	1,362	1,360	999	73.5%	73.3%	73.4%
	深層格選好のみ	1,362	1,350	953	70.6%	70.0%	70.3%
	深層格推測規則のみ	1,362	1,336	982	73.5%	72.1%	72.8%
	ベースライン手法	1,362	1,362	1,054	77.4%	77.4%	77.4%

記している。

6.1.1 日本語の結果

まず、日本語の結果について考察する。表 5 の対象数と出力数の関係を見ると、クローズドデータとオープンデータとも同程度の値であり、精度と再現率の間に大きな差は存在しない。ベースライン手法との F 値の比較において、クローズドデータでは 1.5 ポイント、オープンデータでは 3.5 ポイントの差であった。Minimally supervised learning と Supervised learning の一般的な差に関する知見は存在しないため、主観的な判断をせざるを得ないが、1.5 から 3.5 ポイント差というのは、タグなしデータから規則を学習する提案手法が、タグ付きデータから規則を学習したベースライン手法に迫る結果であると考えられる。

本手法が、人手による労力の軽減を目的として、二つの名詞を格支配する動詞句に対象を限定しているのに対し、大石らの手法 [1] は、より複雑な統語構造をもつ文も対象としているため、直接比較することは適当ではない。しかしながら、同じ深層格推測に関する研究であるため、参考として大石らの結果について以下に言及する。大石らの結果では、学習による知識のみを用いた推測結果が EDR コーパスと一致した正解率は 80.32% であった。本手法のクローズドデータに対する精度は 81.2%、再現率は 78.0% であり、対象の違いを考慮する必要があるが、大石らの手法と大きな

差はない。大石らの手法は表層的な表現を学習の手掛りとしているため、日本語以外の言語に適用するためには、コーパスとは別に表層表現を事前に与えなおす必要がある。本手法は、言語ごとにコーパスを用意するだけでよく、汎用性に優れている。対象の違いによる影響はあるが、このように汎用性の高い枠組みの中で、従来手法と同程度の精度であったことは本手法の有効性を示すと考えられる。

本手法のオープンデータに対する精度は 69.5%、再現率は 62.1% であり、ベースライン手法と比較すると精度で 3.9 ポイント、再現率で 3.1 ポイントの低下であり、クローズドデータにおける差と同程度であった。

深層格選好のみを用いた推測の F 値を調べると、クローズドデータで 72.3%、オープンデータで 61.3% という値であり、単語の概念がもつ深層格の傾向を手掛りとするので、6 割以上の精度で推測可能であることが確認できる。この深層格選好の推測結果から学習された深層格推測規則のみを用いた推測の F 値を調べると、クローズドデータで 74.0%、オープンデータで 67.5% という値であった。このことは、一文一格の原理を考慮することで、深層格選好よりも精度の高い規則をタグなしコーパスから学習できたことを示している。

深層格選好と規則を併用した提案手法の F 値は、クローズドデータにおいては最も高い値を示したが、

オープンデータにおいては規則を単独で用いた場合よりもわずかに低い値となった。しかしながら、クロズドデータとオープンデータにおけるベースライン手法の結果の差から、今回の実験で用いた学習データの量が、オープンデータをクロズドデータと同程度の精度で推測するには不足していたと考えられる。したがって、学習データの量を増加させることにより、提案手法が規則を単独で用いた場合よりも向上することが期待できる。今回の実験では、十分な量の学習データを確保できなかったが、今後、学習データの量を増加させ精度が向上するか調査したいと考えている。

6.1.2 英語の結果

次に英語の結果について考察する。表 6 に示すように、クロズドデータに対する本手法の精度は 78.5%、再現率は 77.8%であり、オープンデータに対する精度は 73.5%、再現率は 73.3%であった。日本語の場合と同様に、精度と再現率の間に大きな差はない。また、ベースライン手法との F 値の差は、提案手法がクロズドデータで 0.2 ポイント上回り、オープンデータで 4.0 ポイント下回った結果となり、日本語と比較して大きな差はないと考えられる。

深層格選好のみを用いた推測の F 値は、クロズドデータで 74.4%、オープンデータで 70.3%であり、7 割以上の精度で推測可能であることが確認できる。日本語の場合と比較して、オープンデータとクロズドデータの間に顕著な差が現れなかったのは、英語の方が学習データの異なり数が多く十分に学習できた結果と考えられる。学習が比較的十分に行われたと考えられるクロズドデータでの F 値を比較すると、日本語の場合と同程度であり、単語概念の深層格選好を手掛りとした推測は、日本語と英語の違いに影響を受けないことが確認できた。

深層格推測規則のみを用いた推測の F 値は、クロズドデータで 74.4%、オープンデータで 72.8%であった。この結果は、日本語の場合と同じく、深層格選好よりも精度の高い規則が学習できたことを示している。深層格選好と規則を併用した場合の F 値と比較すると、日本語の場合と異なり、クロズドデータ、オープンデータともに、併用した手法の方が高い値となった。これは、日本語のときと異なり、十分な量の学習データが存在したために、クロズドデータと同程度の精度でオープンデータを推測することが可能になったためだと考えられる。

6.1.3 言語非依存性の考察

以上の結果から、本手法は、日本語と英語の違いによる影響をほとんど受けることなく、7 割から 8 割程度の精度で推測可能であることが確認できた。また、言語非依存性を実現するための中核となる深層格選好による推測は、クロズドデータに対して 7 割以上、オープンデータに対して 6 割以上の精度で推測可能であり、その結果をもとにタグなしコーパスから学習された規則を併用することで、すべてのデータにおいて精度と再現率を向上できたことも確認できた。本実験では、日本語と英語のみを対象としたが、日英の比較において顕著な差が確認できないことから、他の言語に対しても適用できる可能性がある。今後の課題として、他の言語に対しても実験を行い、日英の場合と同程度の精度で推測可能であるか調査したいと考えている。

6.2 推測結果の例と今後の課題

本手法の問題点を明らかにし今後の改善に役立てるため、テストデータ中の具体的な推測結果を例示して検討する。

6.2.1 正解例

まず、深層格選好による推測の段階で正解であった例として、日本語では「父を [object] 戦場で [place] 失い」や「私たちは [agent] 握手を [object] 交わした」など、英語では “Japan[agent] started from scratch[source]” や “those[agent] refuse appointment[object]” などが挙げられる。ただし、ここで例示する文は名詞、動詞、機能語から再構築した文であり、前後の文脈や、形容詞、副詞、冠詞などは省略されている。これらの例は、日本語で 2,613 例、英語で 913 例が存在した。

これらの例において、“Japan” などは object や place としても解釈されていたが agent として解釈される率が高かったため、深層格選好のみを用いて正しく推測することができた。また、規則による推測も、深層格選好による推測と同じ深層格を導くものであるが、式 (3) に示すように深層格選好との積としてゆる度求めた場合に異なる深層格を導くようなものではなかった。

6.2.2 改善例

次に、深層格選好による推測の段階では誤りであったが、規則と併用することにより改善された例を示す。日本語では、「消費者が [agent] 商品を [object] 買うとき」や「自然を [object] 相手に [goal] する」など、

英語では “I[agent]’m going to bank[goal]” や “sell them[object] in Japan[place]” などが挙げられる。このような例は、日本語で 484 例、英語で 86 例が存在した。

ほとんどの例は、ゆう度計算において規則との積をとった結果、深層格選好による誤った推測結果が抑制されたものであった。例えば、「消費者が商品を買うとき」において、深層格選好による推測段階では「消費者」のゆう度は object が最も高い値であり、正解である agent のゆう度は 2 番目であった。しかしながら、[[が:F2][を:F1]] の機能語パターンにおいて [が:F2] に適用される規則は、agent, object, goal の順に高い値であり、object と goal に対する値はほとんど 0 に近い値であった。したがって、規則との積を計算することで object のゆう度が 0 に近い値となった結果、agent のゆう度が object のゆう度を超えることとなり、「消費者」の深層格を正しく推測することができた。

また、改善された結果の中で、深層格選好と規則をそれぞれ単独で用いた場合には誤りとなるが、両方を併用した場合にのみ正解となった例は、日本語で 6 例、英語で 2 例が存在した。日本語では「電話で [implement] 注文を [object] 出し合い」など、英語では “reduce stress[object] at work[scene]” などが挙げられる。これらの例ではすべて、規則との積でゆう度を求めた結果、併用した場合にのみ正しい推測を行うことができた。規則との積によるゆう度の抑制以外で改善された例としては、規則との積をとった結果、すべての深層格に対するゆう度が 0 となり、4.5 で述べたように規則のゆう度を優先したため正解となった例などが挙げられる。

6.2.3 悪化例

反対に、深層格選好による推測の段階で正解であったにもかかわらず、規則と併用することにより誤りとなった例は、日本語で 357 例、英語で 40 例が存在した。日本語では「地方に [place(goal)] 本拠を [object] おく」や「日本では [place] 恩赦は [object(agent)] 及ばなかったが」など、英語では “Kaifu[agent] becomes prime-minister[goal(object)]” や “it[object(agent)] turns profit[goal(object)]” などが挙げられる。例文中の () 内に記されたイタリック体の深層格は、誤って推測された深層格である。

悪化した原因は、改善された原因とすべて同じ理由によるものであった。すなわち、「地方に」のように、規則との積をとることで正解であった深層格のゆう度

が抑制されたことによるものか、「恩赦は」のように、すべての深層格のゆう度が 0 となったために規則のゆう度を優先した結果として誤りとなったものである。したがって、この問題を解決するためには、規則の学習精度を向上させる必要があり、一文一格の原理以外の制約を考慮した学習を行う必要があると考えられる。また、英語の例のように、機能語が存在せず語順のみしか表層情報として利用できない場合には、規則の粒度が粗いものとなり精度の低下を招くため、規則を細分化する手段が必要になると考えられる。

ただし、規則の併用により改善された場合と異なり、それぞれ単独で用いた場合には正解であるが、併用した場合のみ誤りとなるような例は、日英ともに存在しなかった。

6.2.4 誤り例

最後に、深層格選好による推測の段階で誤りであり、規則と併用することによっても改善されなかった例を示す。日本語では「交通事故で [scene(cause)] けがを [object] して」や「トヨタに [beneficiary(agent)] エンジンを [object] 供給したことも」など、英語では “imports[object] are increasing from Southeast Asia[source(place)]” や “you[agent] stay for dinner[scene(object)]” などが挙げられる。誤りであった例のほとんどはこのグループに属し、日本語で 1,530 例、英語で 323 例が存在した。

改善できなかった原因の一つとして、scene や beneficiary など比較のまれな深層格が正解であったことが挙げられる。深層格選好による推測は、統計的手法の一種であるため、まれな事例に対する推測は一般的な事例に対する推測と比較して低い精度となる傾向にある。また、規則の学習も、深層格選好による推測結果に基づくため、まれな事例に対する影響を免れない。本手法は、深層格選好や規則を単独で用いた場合に低いゆう度であっても、両者の積をとった後に単位ベクトルに再変換するため、まれな深層格が推測されないわけではない。例えば、「交通事故で」における cause のゆう度は、深層格選好のみの場合には 0.02、規則のみの場合には 0.04 であるが、併用することで 1.00 にゆう度が向上している。しかしながら、「交通事故で」の scene や「トヨタに」の beneficiary などは、深層格選好と規則のどちらのゆう度も 0 であるため、積を計算しても対処できなかった。また、“from Southeast Asia” の場合には、規則による推測では source のゆう度が存在していたが、深層格選好による推測では

place にしか値が存在しなかった．このような統計上の不備を解決するために，統計的に得られる情報以外の知識を人手で与えることが考えられるが，1. で述べたように，網羅性やコストの面で問題が残る．それゆえ，一般的な原理や制約から統計情報以外の知識を学習する機構が必要となるが，この点に関しては今後の課題である．

6.2.5 規則の学習に対する考察

深層格選好のみを用いた推測結果から，深層格推測規則を併用することで，改善または悪化した例の数を表7にまとめる．表7に示すように，日英ともに悪化した例よりも，改善された例の方が多かった．日本語，英語ともに，改善効果のある規則を学習できたという結果は，本手法の学習の有効性を示すと考えられる．

表7 規則による改善と悪化の数

Table 7 Number of sentences improved or spoiled by application of rule.

	正解のまま	改善	悪化	誤りのまま
日本語	2,613	484	357	1530
英語	913	86	40	323

7. 調 査

7.1 タグ付きデータ量の変化による影響

本手法は，深層格選好を学習するために一定量のタグ付きコーパスを必要とする．学習に必要なタグ付きコーパスの量が少ないほど，人手による労力を軽減することができるため，タグ付きコーパスの量を変化させた場合の精度と再現率を調査することとした．タグ付きコーパスの割合を1~10割に変化させ，それぞれの割合における，提案手法による推測結果，深層格選好のみによる推測結果，深層格推測規則のみによる推測結果を表8と表9に示す．また，クローズドデータとオープンデータにおけるF値の推移を図3と図4にそれぞれ示す．ちなみに，5. の条件では，深層格選好の学習データは深層格推測規則の学習データとタグの有無を除いて同一の事例であるため，10割に相当する．

タグ付きデータを用いて学習される深層格選好の推測結果は，タグ付きデータの割合が減少するにつれて低下している．しかしながら，その深層格選好の推測結果をベースとして学習された深層格推測規則による推測結果は，深層格選好の推測結果よりも高い値で，

表8 タグ付きデータの割合を変化させた場合のクローズドデータの推測結果

Table 8 Results of closed data using the various rates of tagged data.

		タグ付きデータの割合									
		10割	9割	8割	7割	6割	5割	4割	3割	2割	1割
提案手法	精度	81.2%	80.6%	79.8%	78.9%	78.2%	76.8%	75.4%	73.4%	70.5%	69.0%
	再現率	78.0%	77.3%	76.5%	75.9%	75.2%	73.7%	72.3%	69.6%	66.2%	63.1%
	F 値	79.6%	78.9%	78.1%	77.4%	76.7%	75.2%	73.8%	71.5%	68.3%	65.9%
深層格選好	精度	73.6%	73.3%	73.0%	72.2%	71.6%	71.0%	69.9%	67.6%	65.5%	64.8%
	再現率	71.0%	70.2%	69.5%	68.5%	67.4%	65.8%	63.3%	59.9%	55.9%	46.9%
	F 値	72.3%	71.7%	71.2%	70.3%	69.4%	68.3%	66.5%	63.5%	60.3%	54.4%
深層格推測規則	精度	77.1%	77.1%	77.1%	77.2%	77.7%	77.3%	78.0%	77.6%	77.2%	77.2%
	再現率	71.2%	71.1%	71.0%	71.4%	71.3%	69.7%	70.5%	69.6%	68.7%	66.2%
	F 値	74.0%	74.0%	73.9%	74.2%	74.3%	73.3%	74.1%	73.4%	72.7%	71.3%

表9 タグ付きデータの割合を変化させた場合のオープンデータの推測結果

Table 9 Results of open data using the various rates of tagged data.

		タグ付きデータの割合									
		10割	9割	8割	7割	6割	5割	4割	3割	2割	1割
提案手法	精度	69.5%	68.9%	68.6%	68.6%	68.0%	67.7%	66.7%	65.8%	64.7%	65.1%
	再現率	62.1%	61.5%	61.1%	61.0%	59.9%	59.3%	59.2%	57.9%	56.8%	55.0%
	F 値	65.6%	65.0%	64.7%	64.6%	63.7%	63.2%	62.8%	61.6%	60.5%	59.6%
深層格選好	精度	63.2%	63.0%	62.9%	62.0%	61.5%	61.5%	60.5%	59.2%	58.5%	59.8%
	再現率	59.6%	59.0%	58.3%	57.2%	56.3%	55.4%	53.4%	50.9%	48.8%	42.2%
	F 値	61.3%	61.0%	60.5%	59.5%	58.8%	58.3%	56.7%	54.8%	53.2%	49.5%
深層格推測規則	精度	73.4%	73.0%	74.7%	74.9%	75.3%	75.3%	74.4%	74.1%	73.4%	75.0%
	再現率	62.5%	62.0%	63.2%	63.6%	63.4%	61.4%	63.1%	62.1%	61.8%	59.3%
	F 値	67.5%	67.0%	68.5%	68.8%	68.8%	67.7%	68.3%	67.6%	67.1%	66.2%

かつ、安定した結果となった。このことは、学習される深層格推測規則の信頼性の高さを示すものであると考えられる。

深層格推測規則と深層格選好の両方を考慮する提案手法の推測結果は、クローズドデータにおいて5割以上をタグ付きデータとして用いた場合には、三つの手法の中で最も良い結果となった。しかしながら、4割以下の場合には深層格推測規則を単独で用いた場合が常に最も良い結果となり、オープンデータにおいては、深層格推測規則を単独で用いた場合が最も良い結果となった。また、図3と図4から、提案手法の結果は深層格選好の結果に影響される傾向にあることが確認された。以上から、信頼できる深層格選好を学習するために十分な量のタグ付きデータを用意できる場合には、深層格選好を考慮することで精度を向上させることが可能であるが、そうでない場合には、深層格選好は深層格推測規則を学習するためのベースとして用いるにとどめた方がよいと考えられる。したがって、本手法の頑健性を高めるためには、学習データの量に依

じて、深層格選好の利用方法を使い分けることが必要であると考えられるが、このような使い分けを自動的に行うための機構については今後の課題である。

7.2 分野の違いによる影響

分野の違いに対する本手法の影響を調査するために、以下の実験を行った。この実験の目的は、ある分野の表層情報と対応づけられた深層格推測規則を学習するために、他の分野の用例から学習された深層格選好を利用できるかを調査することである。

本論文で用いたEDRコーパスは、新聞・雑誌・辞書などを出典として、政治経済から語釈文まで様々な分野から用例を収録したものである。それゆえ、EDRコーパスに含まれていない分野として、芥川龍之介の「蜘蛛の糸」[14]を用いることとした。JUMAN [15]とKNP [16]を用いて形態素解析と構文解析を行い、その結果を第1著者が人手で校正・整形した事例の中から、二つの名詞句を格支配する68の動詞句、136事例を抽出し対象データとした。

まず、EDR日本語コーパスの4,984事例をタグ付きデータとし、深層格選好の学習に用いることとした。次に、別分野である上述の「蜘蛛の糸」の136事例を、タグを取り除いたEDRコーパス4,984事例に加えた、5,120事例を深層格推測規則の学習データとした。「蜘蛛の糸」を加えた5,120事例の学習データを用いて単語クラスタリングを行った後、タグ付きデータ4,984事例から深層格選好を、タグなしデータ5,120事例から深層格推測規則を学習した。最後に、学習された知識を用いて「蜘蛛の糸」136事例の解析を行った。

正解かどうかの判断は、データ作成時に第1著者が判断した深層格と一致しているかどうかで行った。この第1著者の判断の信頼性を調査するために、EDRコーパスから無作為に抽出した100事例に対し、EDRコーパスのタグが分からない状況で、第1著者から第4著者が相談せずに深層格をそれぞれ判断した結果を用いて評価者間一致率を求めることとした。第1著者と他の著者との一致率(κ 統計量)は、84%(0.80)、81%(0.78)、81%(0.77)となり、極めて高い一致率を示した。また、それぞれの著者とEDRコーパスとの一致率を同様に調査したところ、それぞれ、75%(0.69)、71%(0.63)、64%(0.57)、70%(0.63)であり、EDRコーパスとは高～中程度の一致率を示した。したがって、第1著者の判断には一定の信頼性があると考えられる。

提案手法、深層格選好のみを用いた推測手法、第1

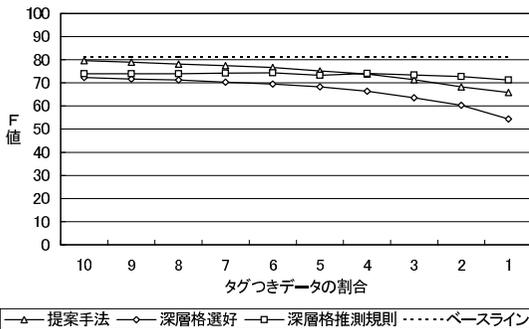


図3 クローズドデータにおけるF値の推移
Fig. 3 Change of F-value in closed data.

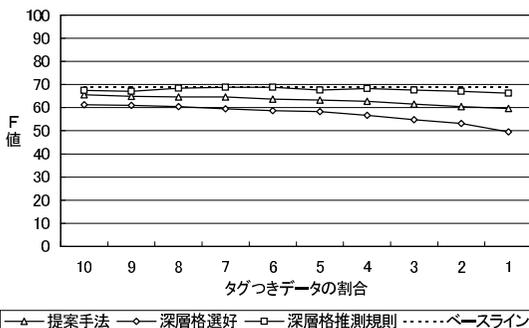


図4 オープンデータにおけるF値の推移
Fig. 4 Change of F-value in open data.

表 10 蜘蛛の糸の推測結果
Table 10 Results of “Kumo no ito.”

		対象数	出力数	正解数	精度	再現率	F 値
蜘蛛の糸	提案手法	136	113	70	62.0%	51.5%	56.2%
	深層格選好のみ	136	134	50	37.3%	36.8%	37.0%
	ベースライン手法	136	130	92	70.8%	67.7%	69.2%

表 11 助詞ごとの推測結果
Table 11 Results of each particle.

助詞「を」							
		対象数	出力数	正解数	精度	再現率	F 値
クローズドデータ	提案手法	14,700	14,275	13,898	97.4%	94.5%	95.9%
	深層格選好	14,700	14,459	13,465	93.1%	91.6%	92.4%
	ベースライン	14,700	14,700	14,260	97.0%	97.0%	97.0%
オープンデータ	提案手法	1,470	1,386	1,289	93.0%	87.7%	90.3%
	深層格選好	1,470	1,425	1,250	87.7%	85.0%	86.4%
	ベースライン	1,470	1,428	1,383	95.9%	94.1%	95.5%
助詞「に」							
		対象数	出力数	正解数	精度	再現率	F 値
クローズドデータ	提案手法	9,560	9,254	6,317	68.3%	66.1%	67.2%
	深層格選好	9,560	9,076	5,190	57.2%	54.3%	55.7%
	ベースライン	9,560	9,258	6,546	70.7%	68.5%	69.6%
オープンデータ	提案手法	956	907	422	46.5%	44.1%	45.3%
	深層格選好	956	886	361	40.7%	37.8%	39.2%
	ベースライン	956	869	483	55.6%	50.5%	52.9%
助詞「が」							
		対象数	出力数	正解数	精度	再現率	F 値
クローズドデータ	提案手法	7,230	7,133	6,324	88.7%	87.5%	88.1%
	深層格選好	7,230	6,992	5,773	82.6%	79.9%	81.2%
	ベースライン	7,230	7,045	5,149	73.1%	71.2%	72.1%
オープンデータ	提案手法	723	657	498	75.8%	68.9%	72.2%
	深層格選好	723	683	487	71.3%	67.4%	69.3%
	ベースライン	723	630	425	67.5%	58.8%	62.8%
助詞「は」							
		対象数	出力数	正解数	精度	再現率	F 値
クローズドデータ	提案手法	7,150	7,100	6,481	91.3%	90.6%	91.0%
	深層格選好	7,150	6,927	5,728	82.7%	80.1%	81.4%
	ベースライン	7,150	7,004	5,851	83.5%	81.8%	82.7%
オープンデータ	提案手法	715	687	543	79.0%	75.9%	77.5%
	深層格選好	715	676	487	72.0%	68.1%	70.0%
	ベースライン	715	692	522	75.4%	73.0%	74.2%
助詞「で」							
		対象数	出力数	正解数	精度	再現率	F 値
クローズドデータ	提案手法	3,150	2,953	1,219	41.3%	38.7%	40.0%
	深層格選好	3,150	3,026	1,124	37.1%	35.7%	36.4%
	ベースライン	3,150	3,061	1,801	58.8%	57.2%	58.0%
オープンデータ	提案手法	315	278	76	27.3%	24.1%	25.6%
	深層格選好	315	286	73	25.5%	23.2%	24.3%
	ベースライン	315	267	81	30.3%	25.7%	27.8%

著者の正解をもとに教師つき学習を行ったベースライン手法の結果を表 10 に示す。「蜘蛛の糸」の 136 事例というデータ量の少なさから、オープンデータの結果と F 値の比較を行うと提案手法が 9.4 ポイント低下しているが、深層格選好のみの場合には 24.3 ポイント低下しており、提案手法が分野の違いによる影響を比較的抑えていることが確認された。しかしながら、今

回の実験では十分なデータ量を用意できなかったため、分野の違いによる詳しい調査は今後の課題である。

7.3 助詞ごとの評価

助詞ごとの深層格推測の困難さを調査するために、上位 5 位の頻出助詞である「を」、「に」、「が」、「は」、「で」を対象に、それぞれの助詞の評価を行った。ベースライン手法、及び、深層格選好による推測手法と比

表 12 単語クラスタを用いない場合の推測結果
Table 12 Results of inference without the word cluster.

	対象数	出力数	正解数	精度	再現率	F 値
オープンデータ	4,984	4,315	2,880	66.7%	57.8%	61.9%

表 13 深層格パターンを用いない場合の推測結果
Table 13 Results of inference without the deep case pattern.

	対象数	出力数	正解数	精度	再現率	F 値
クローズドデータ	49,840	47,931	38,663	80.7%	77.6%	79.1%
オープンデータ	4,984	4,453	3,138	70.5%	63.0%	66.5%

較した結果を表 11 に示す。

5 種類の助詞すべての場合において、提案手法が深層格選好による推測手法よりも良い結果となった。特に、本手法の有効性を示すと考えられるのは、格助詞「が」と係助詞「は」の結果であり、クローズドデータ、オープンデータともに正解タグを用いて規則を学習したベースライン手法よりも良い結果となったことを確認した。これは、係助詞である「は」や、自動詞と他動詞の違いにより深層格が変化する「が」はその性質上、他の三つの格助詞と比較して、名詞や動詞の概念、ともに格支配される名詞の深層格など、包括的な判断が要求される可能性が高く、それらを考慮する本手法が有効であったからだと考えられる。

7.4 単語クラスタリングの効果

単語クラスタリングの影響に関して調査するために、クラスタを用いずに推測する手法との比較実験を行った。ただし、学習データに存在しない未知語の深層格選好を類推するためにクラスタを用いるため、クローズドデータの場合は同一の結果となる。それゆえ、オープンデータを対象とした結果のみを表 12 に示す。

クラスタを用いなかった場合と比較して、提案手法は、精度で 2.8 ポイント、再現率で 4.9 ポイント向上し、F 値が 3.9 ポイント向上した。再現率だけではなく精度においても向上が確認できたことから、本手法のクラスタリングが有効であったといえる。クラスタを用いない手法では正解を推測できず、提案手法でのみ推測できた事例は 306 事例存在した。逆に、クラスタを用いることで正解から誤りとなった事例は 89 事例存在した。クラスタの利用による悪化は、深層格推測規則を単独で用いることで正しく推測できていた事例が、クラスタの深層格選好が誤って考慮されるようになったために起こったと考えられる。しかしながら、改善の 306 事例に対して悪化は 89 事例であり、全体としてはクラスタを利用することで精度、再現率を向

上させることができた。

7.5 深層格パターンの効果

深層格パターンを用いることによる効果を調査するために、提案手法と、動詞の深層格選好に深層格パターンを用いなかった手法との比較実験を行い、その結果を表 13 に示す。

表 5 と比較すると、深層格パターンを用いることで、クローズドデータにおいて 0.5 ポイントの向上、オープンデータにおいて 0.9 ポイントの低下となった。深層格パターンの効果は小さく、データスパースネスの影響を受けやすいオープンデータにおいては低下するという結果であった。しかしながら、データ量を増加したときにわずかでも向上の効果が得られるように、本論文では、クローズドデータで良い結果となった深層格パターンを用いることとした。今後の課題として、学習データの量に応じてスパースネスの度合を考慮し、深層格パターンを利用するかどうか自動的に判断する機構を組み込むことが考えられる。

7.6 表層情報の効果

機能語、位置、機能語パターンの 3 種類の表層情報の効果を調査するために、位置と機能語パターンを用いた「機能語なし」手法、機能語と機能語パターンを用いた「位置なし」手法、機能語と位置を用いた「機能語パターンなし」手法を、提案手法と比較した実験を行った。

それぞれの場合に同一の表層情報とみなされる例を以下に示す。「位置なし」の場合には「太郎が で見た」と「 で太郎を見た」における「太郎」の表層情報が同一として扱われることとなり、「機能語パターンなし」の場合には「太郎が で見た」と「太郎が を見た」が同一として扱われることとなる。最後の「機能語なし」の場合、機能語パターンには対象となる名詞に付属する機能語の情報も含まれるため、機能語パターンに影響を与えずに機能語の情報のみを取

表 14 表層情報を用いない場合の推測結果
Table 14 Results of inference without one of the surface information.

		対象数	出力数	正解数	精度	再現率	F 値
クローズドデータ	機能語なし	49,840	49,392	37,671	76.3%	75.6%	75.9%
	位置なし	49,840	49,368	37,566	76.1%	75.4%	75.7%
	機能語パターンなし	49,840	49,289	37,991	77.1%	76.2%	76.7%
オープンデータ	機能語なし	4,984	4,835	3,080	63.7%	61.8%	62.7%
	位置なし	4,984	4,829	3,239	67.1%	65.0%	66.0%
	機能語パターンなし	4,984	4,881	3,278	67.2%	65.8%	66.5%

り除くことは不可能である。それゆえ、共起する名詞に付属する機能語の情報を位置情報とともに用いた手法を「機能語なし」手法とした。したがって、「機能語なし」の場合には、「太郎が を見た」と「太郎を を見た」が同一として扱われることとなる。また、3種類の表層情報を用いる提案手法では、上記の例文はすべて別の表層情報として扱われることとなる。

結果を表 14 に示す。表 5 と比較すると、クローズドデータにおいて、機能語が 3.7 ポイント、位置が 3.9 ポイント、機能語パターンが 2.9 ポイントの低下、オープンデータにおいて、機能語が 2.9 ポイント、位置が -0.4 ポイント、機能語パターンが -0.9 ポイントの低下となった。したがって、機能語、位置、機能語パターンの順に効果が大きいと考えられる。オープンデータにおいて、位置と機能語パターンが悪影響を及ぼしているが、本論文では、データ量を増加したときにわずかも向上の効果が得られるよう、クローズドデータで向上の効果が確認された 3 種類の表層情報を用いることとした。しかしながら、深層格パターンの効果と同様に、学習データの量に応じて、位置や機能語パターンの利用を自動的に判断する機構が必要であると考えられる。これは、今後の課題である。

8. む す び

本論文では、構文解析済みの単文に対して深層格を推測する手法を提案した。推測に用いる知識は、網羅性とコストの問題に対処するため、一定量のタグ付きコーパスから得られた知識を中核として、タグなしコーパスから深層格推測の規則を学習する。本手法は、単語の概念ごとに、それぞれ特定の深層格に解釈される傾向（深層格選好）をもつという仮定に基づき、一定量のタグ付きコーパスから深層格選好を計算した後、その値を用いてタグなしコーパスから多様な言語表現に対処するための規則を学習する。深層格選好を手掛りの主体とすることにより、本手法は言語の違いに依存せずに深層格を推測することが可能となる。言語

非依存性を検討するため、日本語と英語の EDR コーパスを用いて、二つの名詞をとる単文を対象に実験を行った。その結果、クローズドデータにおける日本語の精度 81.2% (再現率 78.0%)、英語の精度 78.5% (再現率 77.8%)、オープンデータにおける日本語の精度 69.5% (再現率 62.1%)、英語の精度 73.5% (再現率 73.3%) となり、言語の差異による影響を回避する枠組みの中で、教師付き学習と同程度の精度による推測を行うことができた。また、オープンデータの内、規則と併用することにより、深層格選好のみを用いた推測と比較して、日本語では 484 例、英語では 86 例が全体として改善されており、タグなしコーパスから学習した規則の有効性を確認した。改善できなかった例の主な原因は、学習された統計情報の不備、または、規則の精度の低さによるものであり、網羅性とコストの問題を考慮した枠組みの中で、これらの問題を解決することが今後の課題である。

謝辞 本研究の一部は、北海学園大学ハイテク・リサーチ・センター研究費、及び、北海学園大学学術研究助成金による補助のもとに行われた。

文 献

- [1] 大石 亨, 松本裕治, “格パターン分析に基づく動詞の語彙知識獲得,” 情処学論, vol.36, no.11, pp.2597-2610, 1995.
- [2] 本木 実, 嶋津好生, 高橋直人, “階層型ニューラルネットによる深層格解析,” 情処学論, vol.41, no.10, pp.2852-2862, 2000.
- [3] D. Blaheta and E. Charniak, “Assigning function tags to parsed text,” Proc. 1st NAACL, pp.234-240, 2000.
- [4] D. Gildea and D. Jurafsky, “Automatic labeling of semantic roles,” Proc. 38th ACL, pp.512-520, 2000.
- [5] D. Gildea and D. Jurafsky, “Automatic labeling of semantic roles,” Computational Linguistics, vol.28, no.3, pp.245-288, 2002.
- [6] 原田 実, 田淵和幸, 大野高宏, 日本語意味解析システム SAGE の高速化・高精度化とコーパスによる精度評価, 情処学論, vol.43, no.9, pp.2894-2902, 2002.
- [7] 小山正太, 乾 伸雄, 小谷善行, “「名詞と表層格」パター

ンに対する深層格対応の推測；情処学研報，NL-154-22，2003.

- [8] 渋木英潔，荒木健治，栃内香次，“一文一格の原理と規則化に基づいた深層格の自動推測手法”，FIT2003 情報科学技術フォーラム情報技術レターズ，pp.91-92，2003.
- [9] 渋木英潔，荒木健治，桃内佳雄，栃内香次，“深層格の推測手法における自動クラスタリングの利用”，FIT2004 情報科学技術フォーラム情報技術レターズ，pp.79-80，2004.
- [10] (株)日本電子化辞書研究所，EDR 電子化辞書使用説明書，1995.
- [11] S.A. Caraballo，“Automatic construction of a hypernym-labeled noun hierarchy from text,” Proc. 37th ACL，pp.120-126，1999.
- [12] 国立国語研究所，分類語彙表，秀英出版，1964，(1993).
- [13] 長尾 真(編)，自然言語処理，岩波講座ソフトウェア科学 15，岩波書店，東京，1996.
- [14] 芥川龍之介，ちくま文庫芥川龍之介全集 2，筑摩書房，東京，1986.
- [15] 黒橋禎夫，長尾 真，日本語形態素解析システム JUMAN version 5.1，2005.
- [16] 黒橋禎夫，日本語構文解析システム KNP version 2.0 b6，2005.

(平成 17 年 5 月 6 日受付，10 月 11 日再受付)



渋木 英潔 (正員)

平 9 小樽商科大・商卒。平 11 同大大学院修士課程了。平 14 北大・工・電子情報工学専攻博士後期課程了。平 18 北海学園大学大学院経営学研究科経営学専攻博士後期課程了。現在，北海学園大学ハイテク・リサーチ・センター研究員。博士(工学)。博士(経営学)。自然言語処理に関する研究に従事。言語処理学会，日本認知科学会各会員。



荒木 健治 (正員)

昭 57 北大・工・電子卒。昭 63 同大大学院博士課程了。工博。同年，北海学園大学工学部電子情報工学科助手。平元同講師。平 3 同助教授。平 10 同教授。平 10 北大・工・電子情報工学専攻助教授。平 14 同教授。現在，北大・情報科学・メディアネットワーク専攻教授。自然言語処理，特に，機械翻訳，音声対話処理などの研究に従事。情報処理学会，言語処理学会，人工知能学会，認知科学会，ACL，IEEE，AAAI 等各会員。



桃内 佳雄 (正員)

昭 40 北大・工・精密卒。昭和 42 同大大学院修士課程了。同年(株)日立製作所入社。昭 47 北大大学院博士課程単位取得退学。昭 48 北大大学院情報工学専攻助手。昭 61 助教授。昭 63 北海学園大学工学部電子情報工学科教授。自然言語の理解と生成に関する研究に従事。工博(北大)。情報処理学会，言語処理学会，計量国語学会，人工知能学会，日本認知科学会，ACL 等各会員。



栃内 香次 (正員)

昭 37 北大・工・電気卒。昭 39 同大大学院修士課程了。北大・工・電子情報工学専攻教授を経て，現在，北海学園大学経営学部教授。主として音声情報処理，自然言語処理の研究に従事。工博。情報処理学会，日本音響学会，言語処理学会等各会員。