論　文

帰納的学習による表層文から意味表現への変換規則の自動獲得と適用

森　英悟[†]　荒木　健治[††]　宮永　喜一[†]　栃内　香次[†]

Automatic Acquisition and Application of Translation Rules from Sentences to Semantic Representations Using Inductive Learning

Eigo MORI[†], Kenji ARAKI[††], Yoshikazu MIYANAGA[†], and Koji TOCHINAI[†]

あらまし　帰納的学習による表層文から意味表現への変換規則の自動獲得手法を提案する．提案手法では，システムはあらかじめ人手により与えられた表層文と意味表現の対の集合を，既知の規則の再帰的な適用や部分規則の推定によって一般化することにより，表層文から意味表現への変換規則を獲得する．規則の選択手法として重み付きリンクを経由する方式を採用している．リンクは規則獲得を通じて，表層文の一部や単語概念，意味表現，シソーラスから規則を想起するように構築される．シソーラスは文外の知識としてあいまい性解消のために用いられる．リンクの重みは，想起した規則の適用の正否に応じて規則獲得過程で自動的に設定される．解析は想起した規則が入力文に対して適用可能であることを検証する形で進むために規則検索時間が短く，また従来の形態素解析，構文解析，意味解析にあたる解析は同時に行われるため，規則数が増加しても高速な解析が可能である．本論文では，本手法で用いた意味表現の構成，規則の獲得および適用手法について述べ，実験システムを構築して行った実験結果から本手法の有効性を示す．

キーワード　規則獲得，意味表現，格フレーム，連想リンク

1. まえがき

自然言語理解のように意味に踏み込んだ深い解析を行う際には，表層文のもつ意味を定式化し，これを文脈構造や談話構造の解析などの上位のアルゴリズムに対する入力とすることが一般的である[1],[2]．従来の伝統的な手法での表層文から意味表現への変換は，形態素解析，構文解析，意味解析などの解析過程からなり，それぞれの解析における規則は人手で設計されたものが用いられてきた．実際の多様な言語表現を解析するためには，規則系の充実が欠かせないが，規則の無矛盾性を保ちながら規則系を拡大することは容易ではない．

本論文では，こうした問題を解決するために，帰納的学習によるこれらの規則の自動獲得手法を提案する．本手法における帰納的学習とは，人手によりあらかじめ与えられた表層文と意味表現の対の集合を，既知の規則の再帰的な適用や部分規則の推定によって一般化することによって，表層文から意味表現への変換規則を自動的に獲得することを言う[3],[4]．

近年，大量データからさまざまな情報を自動獲得する研究が盛んになっており，対訳コーパスから動詞の格フレームを獲得する手法[5],[6]，機械翻訳において対象領域の変化に応じてシステムが規則や辞書を適応的に変化させる手法[7]などが提案されている．これらの研究はデータとして対訳のコーパスを用い，手法の応用としては主に機械翻訳を想定したものである．これに対し我々の手法は，自然言語理解システムへの応用を目的とし，表層言語と意味表現のデータ対からその変換規則を獲得するものである．同様の規則の獲得手法としては言語獲得の計算機モデル構築の観点からの錦見らの研究[8]があるが，本研究では工学的な実現性に重点をおき，適応的な自然言語理解システムの構築を目標としている．

一般に，多くの規則の中から適切な規則を選択することは重要な課題であるが，本手法では重み付きのリンクによって規則を選択する方式を採用している[9],[10]．リンクは規則獲得を通じて，表層文の一

[†] 北海道大学大学院工学研究科，札幌市
Graduate School of Engineering, Hokkaido University, Sapporo-shi, 060-8628 Japan
[††] 北海学園大学工学部，札幌市
Faculty of Engineering, Hokkai-Gakuen University, Sapporo-shi, 064-0926 Japan

部や単語概念，シソーラス，意味表現から規則を想起するように張られ，規則適用時には重みの大きいリンクをたぐることによって規則を想起する．リンクの重みは，想起した規則の適用の正否に応じて規則獲得過程で自動的に設定され，適応的に変化する．規則は表層表現から意味表現への直接的な変換規則であり，形態素情報も同じ規則に含まれる．解析は想起した規則が入力文に対して適用可能であることを検証する形で進むため規則検索時間が短く，また従来の形態素解析，構文解析，意味解析にあたる解析は同時に行われるため，規則数が増加しても高速な解析を可能にしている．

自然言語の意味を定式化する際には，自然言語の含むあいまい性の扱いが大きな問題となる．一般にあいまい性は発話の文脈，一般知識などさまざまな情報から解消されるが，本手法では一般知識としてシソーラス情報を与え，規則獲得時に行われるリンク構築と重みの設定を通じてシソーラス情報を経由したあいまい性の解消法を導入している．

以下，まず本手法で用いた意味表現の構成について述べ，続いて規則獲得と適用のアルゴリズムについて述べる．更に，実際に実験システムを構築して行った実験の結果から本手法の有効性を示す．

2. 意味表現の構成

2.1 格フレーム

本手法では，意味表現として深層格フレームを採用し，システムが対象とする自然言語を英語としている[11],[12]．格フレームは素性-素性値対の集合として，以下のように定義する．

$$Ss \equiv [[F1, V1], [F2, V2], \ldots, [Fn, Vn]] \qquad (1)$$

素性 Fi は，大きく三つに分類される．

$$Fi \in \{\text{概念素性}, \text{格素性}, \text{付加素性}\} \qquad (2)$$

一般的な制約として意味表現は以下の事項を満たすものとする．

[制約 1]　必ず概念素性を一つもつ．

[制約 2]　複数の同じ素性を同時にもたない．

以下，各素性の素性値 Vi について述べる．

2.2 各素性の素性値

2.2.1 概念素性

概念素性（HEAD）の素性値は文法理論における主要部の語義に相当し，文には動詞，名詞句には名詞の語義が素性値となるものとする[13]．本手法では，単語と概念の対応関係は単語-概念辞書としてあらかじめシステムに与えるものとしている．以下単語-概念辞書の構成方法について述べる．

単語は一般に複数の語義（概念）をもち，これらは別の語義として意味表現上に反映させる必要がある．他言語での語義のマッピングの差異を利用して，語義定義を行う手法として Dagan らの研究[14]があるが，本手法でも日本語と英語での語義のマッピングの差異を利用し，以下の方法によって語義を分類し，同時にシソーラスデータを作成している[15]．

[手順 1]　英単語の日本語訳を英和辞書を参照しながら付加（複数）し，これを語義とする[16],[17]．

[手順 2]　各語義に日本語シソーラス（分類語い表）を用いてシソーラス情報を付加する[18]．

分類語い表には日本語単語が 6 層に構造化して表記されており，対応する日本語訳の分類番号をそのままシソーラス情報として与えた．手順 2 によって，日本語訳の等しい語義は同じシソーラス情報をもつことになる．語義定義の例を図 1 に，シソーラス情報付加の例を図 2 に示す．

本手法における変換規則は，表層的な言語現象と意味表現の対応関係を記述するものである．本手法では，英語における表層的な言語現象として，語順と形態素変化（語尾変化）を取り上げる．英語における語尾変化には明らかな規則性があり，数も少数である．そこで，単語-概念辞書には各単語の表層上のすべての表現を見出し語として用いることとし，形態素情報として対応概念が同じ見出し語に通し番号を付加した．通し番号は規則的に与えている（原形：1，-s 形：2，-ing 形：3 など），但し，後述のように，システムはあくまでも表層的な一致を手掛りとして動作し，あらかじめ与えられた文法知識などは用いらない．単語-概念データ，シソーラスデータは形態素情報と合わせて単

```
paper → paper_A （紙）
paper → paper_B （論文）
thesis → thesis_A （論文）
```

図 1　語義定義の例
Fig. 1　Examples of word meaning definition.

```
paper: paper_A 1.4.1.1.0.1
paper: paper_B 1.3.1.5.4.6
thesis: thesis_A 1.3.1.5.4.6
```

図 2　シソーラス情報付加の例
Fig. 2　Examples of thesaurus data addition.

語–概念辞書としてシステムに与えられる．各データは表層単語を見出し語に以下のフォーマットで与えられる．

表層単語/語義/形態素情報/シソーラス情報　(3)

単語–概念辞書の一部を図3に示す．本手法では解析に品詞情報を使用しないため，辞書には品詞の情報は付加されない．但し，多品詞語の各品詞に対応する語義は別のものとして，概念の接尾記号（_A,_B,...）で区別する．代名詞は語義をPROとし，熟語はそのもの全体が一つの概念に対応するものとしてlook-for_Aのように取り扱う．

2.2.2 格素性

格素性は素性値として意味表現自体をもつ素性で，動詞が下位範ちゅう化する格情報などが相当する．格素性は深層的な意味役割に対応して定義する．表1に格素性を示す[19]．

2.2.3 付加素性

付加素性は態，相，数，人称など付加的なさまざまな情報を与える素性である．素性値は各々に定義された複数の素性値から選択される．表2に付加素性と素性値を示す．

paper/paper_A/1/1.4.1.1.0.1
papers/paper_A/2/1.4.1.1.0.1
paper/paper_B/1/1.3.1.5.4.6
papers/paper_B/2/1.3.1.5.4.6
write/write_A/1/2.3.1.5.0.1
writes/write_A/2/2.3.1.5.0.1
writing/write_A/3/2.3.1.5.0.1
wrote/write_A/4/2.3.1.5.0.1
written/write_A/5/2.3.1.5.0.1

図3　単語–概念辞書の一部
Fig. 3　Part of word-concept dictionary.

表1　格素性と素性値
Table 1　Case features and their values.

素性	略記	意味
Agent	AGT	動作主
Theme	THEME	主題
Patient	PAT	動作の対象
Manner	MNR	動作の様態
Experiencer	EXP	経験者
Recipient	REP	受取人
Destination	DST	終点
Frequency	FRQ	頻度
Degree	DGR	程度
Time	TIM	時
Location	LOC	場所
Modifier	MDF	属性，修飾値
Possession	POSS	所有
Instrument	INS	道具
Connection	CNC	接続関係
Relation	REL	文脈関係

2.3 意味表現の例

本手法における意味表現の例を図4に示す．全文に対する概念素性は，likeの語義になり，格素性として，EXP，THEME，DGRが付加される．それぞれの格素性は対応する名詞句の名詞，副詞句の副詞の語義を概念素性としてもつ意味表現を素性値としてもつ．付加素性として，全文の意味表現には時制，文属性，格素性（EXP，THEME）の素性値には，数，定が付加されて意味表現を構成している．

3. 処理過程

3.1 変換規則とリンク表現

本手法における変換規則を一般化された表層表現と意味表現間の対応付けとして定義する．計算機上での扱いを容易にするため，表層表現はリスト構造で表し，変換規則は以下のように表記する．

$$[W1, W2, \ldots, Wn] \Rightarrow [[F1, V1], [F2, V2], \ldots, [Fn, Vn]] \quad (4)$$

但し，

$$Wi \in \{Words, @n, < m >\} \quad (5)$$

$$Vi \in \{\text{付加素性値}, @n\} \quad (6)$$

表2　付加素性と素性値
Table 2　Peripheral features and their values.

素性	略記	意味	素性値（例）
Number	NUM	数	SG, PL
Person	PERS	人称	1, 2, 3
Gender	GEN	性	MASC, FEMI
Case	CASE	格	NOM, GEN, ACC
Definite	DEF	定	+, −
Distance	DISTANCE	距離	LONG, SHORT
Tense	TENSE	時制	PRESENT, PAST
S-ATTRibute	S-ATTR	文属性	A, N, Q
Aspect	ASPECT	相	PROGRESSIVE
Modality	MODAL	態	POSSIBILITY, REQUEST

Judy likes sukiyaki very much

[[HEAD,like_A],
[EXP,[[HEAD,Judy_A],[NUM,SG],[DEF,+]]],
[THEME,[[HEAD,sukiyaki_A],[NUM,UC],[DEF,-]]],
[DGR,[[HEAD,much_A],[DGR,[[HEAD,very_A]]]]],
[TENSE,PRESENT],[S-ATTR,A]]

図4　意味表現の例
Fig. 4　An example of semantic representation.

とし，*Words* は表層単語を示す．以下，変換規則中の表層表現側，意味表現側をそれぞれ表層表現片，意味表現片と呼ぶものとする．@n は一般化され他の表層表現，意味表現で置き換えられる部分（置換えマーカ）を示す．同じ番号の置換えマーカは両者の間に対応関係のある表層表現と意味表現で置き換える．$<m>$ は形態素情報を示し，表層表現片中で直前の置換えマーカに対して作用する．この場合，直前の置換えマーカを置き換える表層表現は単語となり，m はその単語が単語–概念辞書内にもつ形態素情報となる．変換規則の例を図5に示す．図5のルールが適用されると，双方の@1の部分は対応する表層表現と意味表現で置き換えられる．表層表現片の@2の部分は意味表現片のHEADの値（@2）を概念としてもつ単語となる．このとき，表層表現部の@2を置き換える単語は，直後の形態素情報 <2> を満たすものでなくてはならない．すなわち，意味表現片に代入される概念と一致し，かつ表層表現片に記載された形態素情報をもつ表層単語のみにこの規則は適用できる．例えば，writesには適用できるが，write，writingなどには適用できない．(図3参照)

変換システムが実際に入力された表層文に対して規則の適用を行う際には，多くの規則の中から適用する規則を選択する必要がある．しかし，適切な規則を選択するのは必ずしも容易なことではない．本手法では，適用規則の決定にいくつかの手掛りから規則を想起する方法を用いている．規則はリンクをたぐることで重みの大きいものから順に採用されるため，適用規則を全数検査によって選定するプロセスは行われない．この方法による適用規則の選択を規則の想起と呼ぶものとする．規則想起の種類を以下に挙げる．

[想起法1]　表層文の一部を手掛りに規則を想起
[想起法2]　単語から辞書リンクを通して概念を想起し，概念から直接，あるいはシソーラスリンクを経由して上位の規則を想起

```
表層表現片: [@1,@2,<2>,@3]
              ⇓
意味表現片: [[HEAD,@2],
        [AGT,@1],[THEME,@3],
        [TENSE,PRESENT],[S-ATTR,A]]
```

図5　変換規則の例
Fig. 5　An example of transformation rule.

[想起法3]　意味表現片を手掛りに上位の規則を想起

想起手掛りから規則へのリンクを連想リンク，規則関係にある表層表現片と意味表現片を結ぶリンクを規則リンク，シソーラス階層を構成するリンクをシソーラスリンク，単語と概念を結ぶリンクを辞書リンクと呼ぶ．シソーラスリンクはシソーラスのノードであることを示す“THP”の後に，各概念のもつシソーラス情報を付加したものをリンクするものである．リンクは概念と結合される末端のTHP.x1.x2.x3.x4.x5.x6から最上位のTHP.x1まで，単語–概念辞書中の6層のデータをそのまま階層構造として表現するように構築される．例えば，hit_Aからは，

hit_A
→ THP.2.1.5.6.3.1
→ THP.2.1.5.6.3
→ THP.2.1.5.6
→ THP.2.1.5
→ THP.2.1
→ THP.2

のリンクが作られる．

それぞれのリンクは単一のリンク系中に表現され，各リンクのもつ属性と重みによって種類と優先度が区分される．手掛りからリンクをたどることによって，まず表層表現片または意味表現片を得ると，更に規則リンクをたぐって変換規則の対が得られる．連想リンクを含めた変換規則系の一部を図6に示す．図中[a]からの想起が想起法1，THP.2.1.5.6.3.1からの想起が想起法2，[[HEAD,@1],[NUM,SG],[DEF,-]]からの想起が想起法3による規則想起である．

各リンクは直接のリンク元からの関係でなく，更に一段前の想起元の履歴（下段想起履歴）を記憶し，リンク選択の精度を高めている，下段想起履歴の一部を図7に示す．図中association link2が下段想起履歴を示すリンクである．“he”に対する意味表現を規則適用によって作成しているケースを考える．heから，dictionary linkをたどって概念PROが選ばれ，更に，association linkをたどって意味表現片が選択される．意味表現片の選択時には複数の選択肢があるが，リンクの重みだけでは有効なリンク選択ができない．これは“he”に，PRO以外の情報が含まれているためであり，意味表現片の選択時には，PROを想起した想起元(“he”or“I”...)が何であったかが有効である．これが下段想起履歴であり，規則獲得時に収集され規則適用

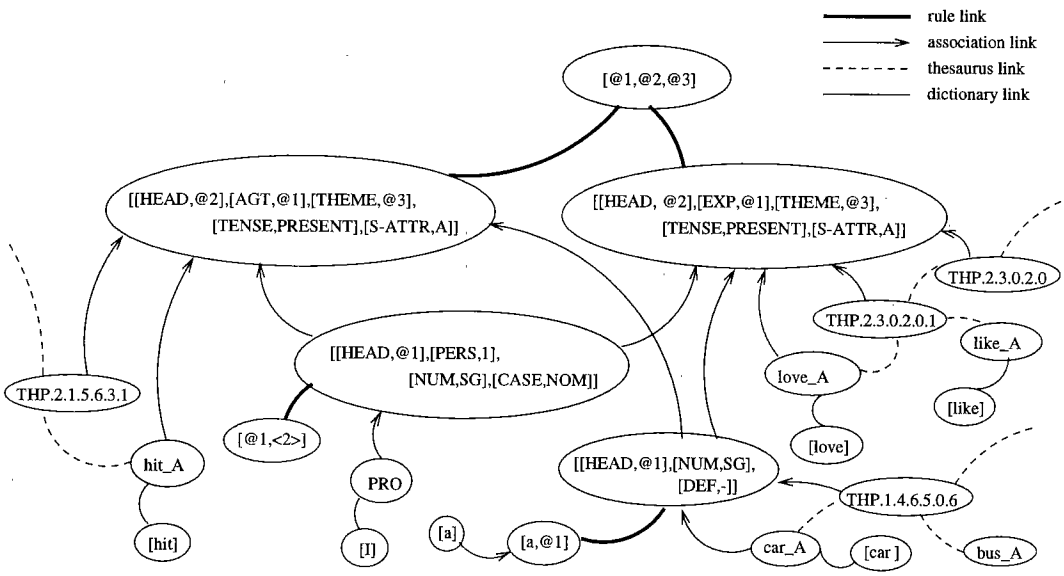


図 6　規則体系の一部
Fig. 6　Part of link system.

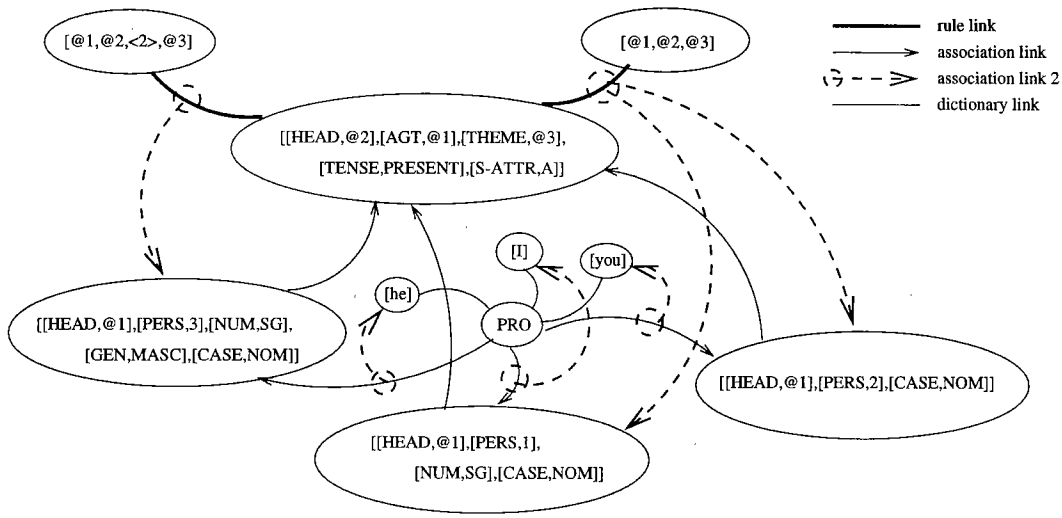


図 7　下段想起履歴の例
Fig. 7　Part of links from lower nodes.

時に利用される．例では PRO からの連想リンクの重みだけでなく，その連想リンクがもつ PRO の想起元に対する下段想起履歴を考慮して連想リンクが選択される．同様のケースには，数や人称の素性値が表層表現片の選択（三人称単数現在形の-s 化）に影響を及ぼすケースがある．

実験システムでは規則想起の優先度は直接のリンク重み lw および下段想起履歴リンクの重み xw の積 $lw \times xw$ によって決定されるものとした．但し下段想起履歴がない場合には，あらかじめ決めた最小値 $xw0 > 0$ を用い，$lw \times xw0$ によって重みの積を求めるものとした．

3.2　変換規則の適用プロセス

本システムは規則の獲得と適用を大量のデータに対

して行うことによって，適応的に規則の獲得を試みるシステムである．従って，規則の獲得と適用のアルゴリズムは密接に関連している．変換規則の適用アルゴリズムは独立して実行することもでき，このときには表層文を入力として受け取り，意味表現を出力する．本節では変換規則の適用アルゴリズムを規則の適用を単独に行うプロセスに従って説明する．後述の規則獲得プロセスでは，その時点で獲得済みの規則を適用することによって対応する意味表現の作成を試み，それが失敗に終わったときに新しい規則の獲得を行う．伝統的な形態素解析，構文解析，意味解析に相当する解析は対応規則の適用によって同時に行われることになる．

ここでは，規則獲得プロセスによって規則の獲得とリンク重みの設定が完了しているものと仮定する．入力文への規則適用アルゴリズムの手順を以下に示す．

[手順 1] 文頭の単語を受け取り，変換規則を想起（想起法 1or2）

[手順 2] 後続の単語に想起中の変換規則を適用

[手順 3] 後続の単語を受け取るたびに想起中の規則の妥当性を検証

[手順 4] 妥当性の低い規則の適用は中断し，バックトラックして新しい規則を想起

[手順 5] 想起中の規則適用の完了と同時に入力単語が終了したとき，規則の適用が完了

[手順 6] 規則適用完了後も入力単語があるときには上位規則を想起（想起法 3）

[手順 7] 規則中の概念素性に対する置換えマーカには辞書リンクを参照

[手順 8] 規則中の格素性に対する置換えマーカには再起的に規則を適用

システムは入力を受け取ると，まず先頭単語のリンク状況を一時記憶用の領域（作業記憶）に複写する．作業記憶上のリンクは活性状態を示す変数をもち，活性/未活性/活性済みの三つの値をとる．作業記憶に複写されたリンクは初期状態では未活性であり，規則想起に用いられると活性状態になる．想起中の規則の適用が中断すると，バックトラックを行って次の優先度をもつ未活性リンクから規則を想起し，活性中だったリンクは活性済みの状態になる．作業記憶を用いることによって，1文で同一の規則が複数回適用されたときに，リンクの活性状態が相互に誤作用することを防ぐことができる．

想起中の規則を入力文に適用する妥当性は，後続の単語を手掛りとして現在想起中の規則を想起可能かを検証することで判断される．検証過程では文外の知識としてシソーラスデータが用いられ，意味表現作成におけるあいまい性の解消が図られる．規則適用中断の条件を以下に示す．

[中断条件 1] 表層表現片と入力単語の不一致．形態素情報が不一致だったケースも含まれる．

[中断条件 2-a] 意味表現片の格素性の置換えマーカに埋め込むために作成された意味表現から，意味表現片への連想リンクがないとき

[中断条件 2-b] 意味表現片の概念素性の置換えマーカに埋め込むために作成された概念素性から，意味表現片への連想度が設定値より低いとき

但し概念素性からの連想度は，概念から想起中の意味表現片へ直接の連想リンクが存在する場合を最大，概念からシソーラスリンクを経由して最上位階層までさかのぼっても連想リンクがない場合を最低とする．シソーラスリンクを経由して意味表現片に到達したとき，そのシソーラス階層があらかじめ決めた階層より上位だった場合には 2-b により規則適用が中断する．

バックトラックを含む規則適用の例を図 8 に示す．図 8 の例では先頭入力 "I" から想起法 2 によって，辞書リンクをたどって概念 PRO と意味表現片，表層表現片が想起される．更に想起法 3 によって上位の規則の想起が行われるが，like に対する概念 like_A から意味表現片への連想度が低いために規則適用が中断され次の規則が想起される．新たに想起された規則に対しては，like_A からシソーラスリンクを経由して連想リンクが存在する．入力 "a car" に対しては "a" から想起法 1 を用いて想起した規則による再帰的な規則の適用が行われる．バックトラックであらためて想起された規則は後続の入力単語からの適用の妥当性が認められ，規則適用は最後まで成功している．

3.3 変換規則と連想リンクの獲得プロセス

変換規則と連想リンクを獲得するプロセスは，表層文と意味表現の対（実例データ）を集めたデータベースに対して行われる．図 9 に規則獲得過程のフロー図を示す．システムはデータベースから 1 組の実例データを受け取ると，表層文から意味表現への変換がその時点までに獲得した規則の適用によって可能であるかを検査する．表層表現への既知規則適用時の動作は基本的に規則適用プロセスと同じであるが，規則適用の中断は作成された意味表現と実例データの意味表現と不一致が発生したときに発生する．規則の適用が部分

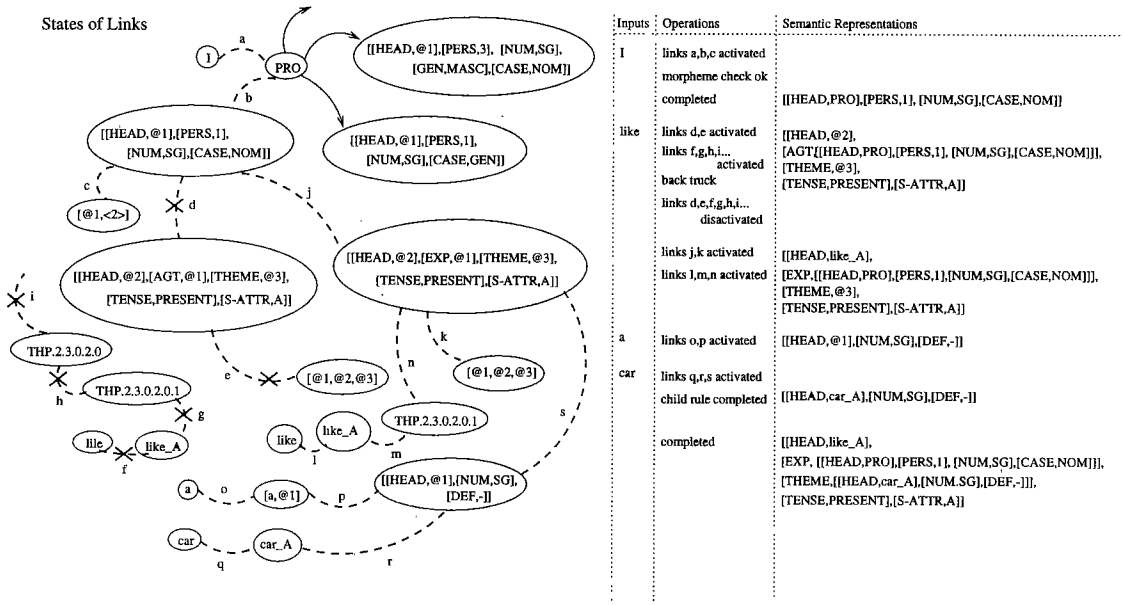


図 8　規則適用過程の例
Fig. 8　An example of rule application process.

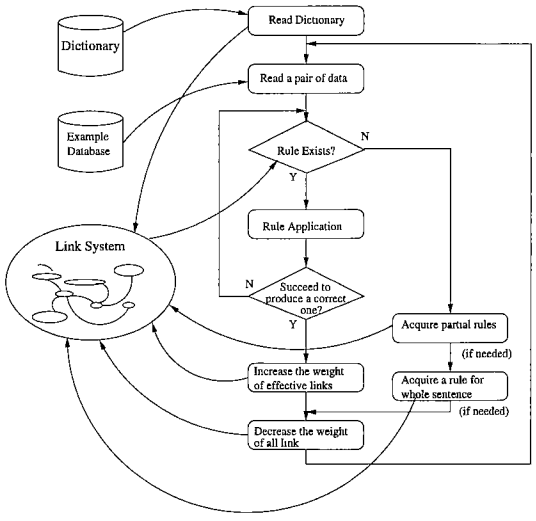


図 9　規則獲得のフロー図
Fig. 9　Flow chart of rule acquisition stage.

的にでも成功したときには，その解析時に用いられたリンクの重みを強化する．

既知の規則のみで与えられた意味表現に到達できなかったときには，そのケースに応じて必要な新規則の獲得を行う．与えられた意味表現中の格素性の素性値（意味表現）に生成できないものが存在した場合は必要な部分規則が未獲得であると判断し，部分規則の獲得を行う．新部分規則の獲得の必要のないとき，およびすべての部分規則の獲得が終わったときには，更に全文に対する規則の獲得を行う．これらの規則獲得作業は再帰的に行われる．

規則獲得時には，実例データを 1 組処理するごとに，すべてのリンクの重みを低下（忘却）させる方法を採用した．誤った規則が獲得された場合でも，この仕組みによって無効なリンクの重みは規則獲得が進むと共に徐々に減少する．

適用成功によるリンクの重みの増分 aw，時間経過による減分 dw はその時点での重みの大きさによって定まるものとする．但し，すべてのリンク重みは 0〜1 の値をとるものとする．現在のシステムでは増分は重みにかかわらず一定（図 10 (a)）とし，

$$aw = t1 \tag{7}$$

減分については，図 10 (b) のように 2 次式により決定する方法を採用した．

$$dw = ax^2 + bx + c \tag{8}$$

とし，最大減分 $t2$，最小減分 $t3$，および 2 次係数 $a \leqq t2 - t3$ を定めると

$$b = t3 - t2 - a \tag{9}$$

$$c = t2 \tag{10}$$

となる．減分に2次式を用いたために，リンク重みが最大値近くまで増えたときには，容易に淘汰されなくなる．これは，人間においても定着した記憶は忘れにくい傾向にあることから着想して採用した．効果の詳細な検討は今後の課題である．

部分規則の獲得が必要なときには，以下の手順で表層表現中から格素性の素性値に対応する部分の抽出を行う．対応部分の抽出には，単語-概念辞書を手掛りとして用い，#n1-n2-n3-f の構造のマークを使用する．ここで n1 は推定中の格素性値の通し番号（位置番号），n2 は対応概念の直接存在する意味表現の深度のレベル，n3 はそのレベルでの通し番号（位置番号）である．f は推定中の格素性を示す．

[手順1] 単語-概念辞書を参照し，概念素性値をもつ単語の前にマークを付ける．

[手順2] 既知の変換規則の適用を仮定することによって，同じ位置番号（n1）のマークをもつ単語を含む部分が，表層表現片として獲得可能であるかを検査．

[手順3] 手順2が失敗したときには位置番号（n1）が等しいマークを含み，異なる位置番号（n1）をもつマークや置換えマーカを超えない最大部分を表層表現片と推定．

解析対象の意味表現と，推定された表層表現は概念素性値を置換えマーカで置き換えられ，新しい規則として記憶される．解析対象の意味表現が格素性をもつときには上記の操作は再帰的に行われ，更に新しい規則を得た上で，そのレベルでの規則が得られる．こうして，すべての格素性の素性値を置換えできるように部分規則を作成した後，格素性の素性値および抽出された表層表現部は同じ置換えマーカで置き換えられる．置き換えられてできた意味表現片と表層表現片の関係が未知の場合には新ルールとして記憶される．各部分規則の意味表現片からは上位規則の意味表現片に対しての連想リンクが，また概念からは直接およびシソーラスリンクを経由した連想リンクが上位規則の意味表現片に対して張られる．表層表現片の先頭単語が置換えマーカでなかったときには先頭単語から表層表現片への連想リンクが張られる．この部分規則推定法は必ずしも妥当な規則を推定できるとは限らない．例えば，同一単語が複数回，1文中に出現したときには辞書検索による概念-単語関係の推定を誤ることが予想される．しかし，誤って獲得されたルールは以降の規則獲得過程では有効に動作されずにリンクの重みが減少し，次第に淘汰されていく．

部分規則推定の例を図11に示す．図11の例では格素性 THEME の素性値に対する置換えが成功しなかったために対応表層表現を抽出し部分規則を獲得している．既知の規則の表層表現片から，“computer games” が部分規則の対応表層表現と推定される．更に，MDF の素性値に対して再帰的に解析が行われ，“computer” が対応表層表現と推定される．概念素性と下位の部分

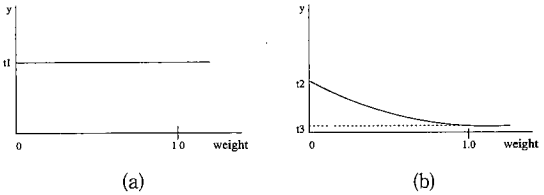


図10 (a) リンクの強化量 (b) リンクの忘却量
Fig. 10 Link (a) Thickening rate. (b) Weakening rate.

○ **Input data from database**

I like computer games

[[HEAD,like_A],
[EXP,[[HEAD,PRO],[PERS,1],
,[NUM,SG],[CASE,NOM]]],
[THEME,[[HEAD,game_A],[NUM,PL],[DEF,-],
[MDF,[[HEAD,computer_A]]]]],
[TENSE,PRESENT], [S-ATTR,A]]

○ **Result of known rule application**

[@1,@2,computer,games]

[[HEAD,@2],
[EXP,@1],
[THEME,[[HEAD,game_A],[NUM,PL],[DEF,-],
[MDF,[[HEAD,computer_A]]]]],
[TENSE,PRESENT], [S-ATTR,A]]

○ **Concept matching using dictionary**

[@1-EXP,@2-HEAD,#1-2-1-MDF,computer,#1-1-1-THEME,games]

○ **Candidate rule for generalization (known rule)**

[@1,@2,@3] [[HEAD,@2],[EXP,@1],[THEME,@3],
,[TENSE,PRESENT],[S-ATTR,A]]

○ **Selected block using candidate rule**

[computer,games] [[HEAD,game_A],[NUM,PL],[DEF,-],
[MDF,[[HEAD,computer_A]]]]

○ **Result of known rule application**

[computer,@1,<2>] [[HEAD,@1],[NUM,PL],[DEF,-],
[MDF,[[HEAD,computer_A]]]]

○ **Concept matching using dictionary**

[#1-1-1-MDF,computer,@1-HEAD,<2>]

○ **No candidate rule for genelarization**

○ **Selected block by inference**

[computer] [[HEAD,computer_A]]

○ **Acquired rules**

[@1] [[HEAD,@1]]

[@1,@2,<2>] [[HEAD,@2],[NUM,PL],[DEF,-],[MDF,@1]]

図11 部分規則推定を含む規則獲得
Fig. 11 Rule acquisiton including partial rule inference.

規則推定によって推定された格素性の素性値を置換えマーカで置き換えて，新しい規則を得ている．前述のように実際にはこれら新しい規則を想起するための連想リンクが張られ規則はリンク体系中に取り込まれる．

4. 実　　験

4.1 実験システム

本手法に従って，実験システムをC++を用いて構築した．単語–概念辞書から設定されるシソーラスリンク，辞書リンクの重みの初期値は1.0とし，規則獲得によって得られる規則リンク，連想リンクの重みの初期値は0.8とした．また，リンクの増分 $t1 = 0.1$，最大減分 $t2 = 0.005$，最小減分 $t3 = 0.00001$，2次係数 $a = 0.00499$ とした．また，概念素性からの連想度による規則適用の中断はシソーラスの深度4以下で連想できない場合に起こるものとした．下段想起履歴リンクが存在しないときの下限値は $xw0 = 0.1$ とした．各パラメータの値は予備実験の結果から最適値と考えられる．

4.2 規則獲得と適用の実験

実例データとして2種類の初級の英語テキストの英文（全文，但し本文のみ）751文に対して，人手で意味表現および単語–概念辞書を作成した[20],[21]．変換規則が全くない状態から，実例データベースの各実例データに対する規則適用と規則獲得を交互に行った．実験では，システムはまず実例データの表層表現に対し規則適用プロセスを実行して意味表現の作成を試み，作成された意味表現が実例データの意味表現と完全に一致したときには正変換が行われたとするものとした．更に，規則適用の正否にかかわらずその実例データによる規則獲得プロセスを行い，これを連続的に全データに対して行った．入力20文ごとの正変換率の推移を図12に示す．但し，正変換率は次式で定義される．

$$\text{正変換率} = \frac{\text{正変換数}}{\text{入力文数}} \times 100 \qquad (11)$$

4.3 考　　察

図12中A点は実例データが別のテキストに切り換わった点である．初級テキストの性質上，徐々に複雑な文が出現するために，既知ルールによる正変換率はテキストの後半では低くなるが，A点以後ではその傾向が緩やかであり，600文以降では正変換率が向上する傾向も見られ，規則獲得が有効に作用していることが確認できる．A点以前の平均正変換率は32.2%，以後は48.2%である．

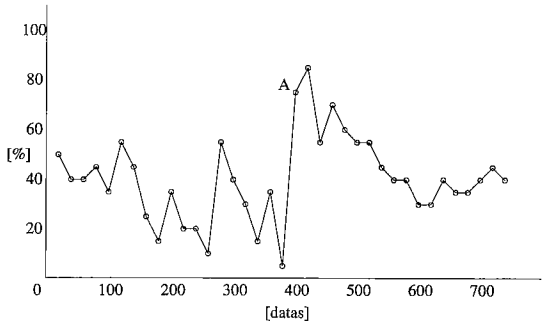


図12　正変換率の推移
Fig. 12　Correct translation rate.

誤変換率，未変換率を以下のように定義する．

$$\text{誤変換率} = \frac{\text{誤変換数}}{\text{入力文数}} \times 100 \qquad (12)$$

$$\text{未変換率} = \frac{\text{未変換数}}{\text{入力文数}} \times 100 \qquad (13)$$

但し，誤変換は生成された出力が間違っていた場合，未変換は変換結果が生成されなかった場合を言う．A点以前の平均誤変換率は9.1%，以後は13.8%であり，A点以前の平均未変換率は58.6%，以後は38.0%であった．未変換が多いのは中断条件2-a.が厳格な規則適用を要求するためと思われる．概念のシソーラス経由の想起（中断条件2-b.）と同様に意味表現間にも連想度を定義し，連想度の高い意味表現を経由した規則想起を可能とすれば，未変換は減少すると思われる．しかし，逆に妥当な中断を阻害し，誤変換増加が起きることも考えられるため今後の検討を要する．

5. む　す　び

本論文では，表層文と意味表現の対応データからその変換規則を自動獲得し，獲得規則を用いて言語表現を意味表現に変換する手法を提案した．獲得した規則は重み付きのリンク体系として表現される．規則適用時には入力文からリンクを通して想起された規則を適用することによって高速な規則適用を可能としている．また，実際に実験システムを構築し，実験を行って手法の有効性を確認した．本手法の有効性を更に検証するには大規模なデータに対する実験が必要である．本手法で用いるデータには，各文に意味表現が付加されている必要があり，これを大量に用意することは容易ではない．こうした大量データを実際に作成した例として，EDRコーパスがあり，約13万文の英文に対し

て意味情報を付加している[11]．現在，我々は意味表現作成を支援するための簡単なツールを用いているが，今後更にこれを発展させることで，大量データの作成を可能にしたいと考えている．また，本手法がもつ動的な規則獲得やリンク重みの設定の特徴を利用し，適応性の高い言語理解システムの構築などの応用を計画している．

文　献

[1] J. Greene, "Language Understanding," Open University Press, 1986.
[2] 石崎　俊，"自然言語処理，" 昭晃堂，1995.
[3] 荒木健治，栃内香次，"帰納的学習による語の獲得および確実性を用いた語の認識，" 信学論（D-II），vol.J75-D-II, no.7, pp.1213–1221, July 1992.
[4] E. Mori, K. Araki, Y. Miynaga, and K. Tochinai, "A Language Acquisition Model on a Computer using Partial Pattern Matching," AMLaP96 Conference Programme and Abstracts, p.45, Sept. 1996.
[5] H. Tanaka, "Verbal case frame acquisition from a billingual corpus: Gradual knowledge acquisition," Proc. of COLING94, pp.727–731, 1994.
[6] 宇津呂武仁，松本裕治，長尾　眞，"二言語対訳コーパスからの動詞の格フレーム獲得，" 情処学論，vol.34, no.5, pp.913–924, May 1993.
[7] S. Yamada, H. Nakaiwa, K. Ogura, and S. Ikehara, "A method of automatically adapting a MT system to different domains," Proc. of TMI95, pp.303–310, 1995.
[8] 錦見美貴子，松原　仁，中島秀之，"複数の領域間の関係に基づいて概念を獲得するシステム Rhea，" 人工知能誌，vol.7, no.6, pp.1096–1106, Nov. 1992.
[9] 森　英悟，荒木健治，宮永喜一，栃内香次，"言語獲得モデルへの連想記憶リンクの導入，" 1996 信学秋季全大，電子情報通信学会，Sept. 1996.
[10] 森　英悟，荒木健治，宮永喜一，栃内香次，"連想記憶リンクを用いた言語獲得の計算機モデル，" 平9北海道連大，Oct. 1996.
[11] 日本電子化辞書研究所，"EDR 仕様説明書，" 日本電子化辞書研究所，1995.
[12] E. Charniak and Y. Wilks, "Computational Semantics," North-Holland Publishing Company, 1976.
[13] B.P. Sells, "Lectures on Contemporary Syntactic Theories," CSLI (Stanford Univ.), 1985.
[14] I. Dagan, A. Itai, and U. Schwall, "Two language are more informative than one," Proc. of 29th Annual Meeting of ACL, pp.130–137, 1991.
[15] 森　英悟，荒木健治，宮永喜一，栃内香次，"自然言語—意味構造対応ルールの獲得と適用，" 信学技報，NLC95-42, Oct. 1995.
[16] 柴田徹士，"アンカー英和辞典，" 学習研究社，1972.
[17] 研究社辞書編集部，"新リトル英和和英辞典，" 研究社，1987.
[18] 国立国語研究所，"分類語彙表，" 秀英出版，1964.
[19] B.J. Blake, "Case," Cambridge University Press, 1994.
[20] 長谷川潔，秋保慎一，宇都　裕，大喜田由馬，沖原勝昭，加藤行夫，坂田俊策，佐瀬廸子，J.R. Bowers, 田中輝雄，田部　滋，原田昌明，藤掛庄一，松島　健，山口喜佐夫，教育出版株式会社編集局，"教育出版ワンワールド 1，" 教育出版，1991.
[21] 太田　朗，伊藤健三，日下部徳次，監修，"ニューホライズン 1，" 東京書籍，1991.

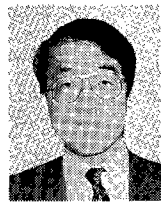
（平成9年5月9日受付，12月24日再受付）

森　英悟　（正員）

平4北大・工・電子卒．平6同大大学院修士課程了．現在同大大学院博士後期課程在学中．自然言語処理の研究に従事．情報処理学会会員．

荒木　健治　（正員）

昭57北大・工・電子卒．昭63同同大学院博士課程了．工博．同年，北海学園大学工学部電子情報工学科助手．平1同講師．平3同助教授．機械学習を用いた自然言語処理の研究に従事．情報処理学会，日本認知科学会，人工知能学会，IEEE，ACL，AAAI各会員．



宮永　喜一　（正員）

昭54北大・工・電子卒．昭56同大大学院修士課程了．昭59米国イリノイ大学客員研究員．現在，北大・大学院工学研究科・電子情報工学専攻教授．主として並列信号処理，並列計算機アーキテクチャ，適応信号処理の研究に従事．工博．情報処理学会，日本音響学会，IEEE各会員．

栃内　香次　（正員）

昭37北大・工・電気卒．昭39同大大学院工学研究科電気工学専攻修士課程了．現在，北大・大学院工学研究科・電子情報工学専攻教授．計算機システムおよび自然言語処理の研究に従事．工博．情報処理学会，日本音響学会各会員．