

多様な自然文間のパターンマッチング

敬和学園大学

安藤司文

概要

意味言語において意味構造の結合関係を変更する事によって多種多様な自然文が生成できるが、意味言語では意味が同じであれば基本的に同じ意味構造で表現される。そのための自然文を意味言語に変換して意味言語でパターンマッチングを行えば、自然言語の多様性が解消され多様な自然文間でのパターンマッチングを簡単に行うことができる。

Pattern Matching between Natural Sentences with Analogous Meanings

Shimon Ando

Keiwa College

ABSTRACT

Natural sentences can be expressed in various ways even when the sentences have the same meaning. But in Meaning Language, different natural sentences expressing the same meaning can be expressed in a single meaning sentence. Therefore, pattern matching between different but analogous natural sentences can be carried out easily by using meaning sentences transformed from natural sentences.

1.はじめに

人間は自然言語を用いて、質問応答、知識獲得、推論、物語理解、知識獲得、翻訳などの知的活動をおこなっている。そのため自然言語の中にそのような知的な活動を支えているメカニズムが存在すると考えられる。しかし、自然言語は個別言語によって決定的に支配され、さらに話し手の状況認識や運用上の判断によって著しく影響を受けるので、自然言語の表現をそのまま全面的に信用して自然言語の研究を行うことはできないと考えられる。話し手が自然言語を表出するためには、話し手の頭の中にその自然言語で表現しようとしている意味構造、即ち概念構造が存在しているはずである。その概念構造を記号化して取り出したものを意味言語と名付け(1, 16, 17)、これを用いて人間の頭の中で展開されている知的なプロセスを解明するのとともに自然言語の言語構造や文法理論について再検討をおこなっている。

これまで日本文の使役文と受け身文の言語構造(2)、日本文における論理構造(3, 4)、自然言語での多様な文の生成(5)、主語(6)などについて言語関係の学会で報告し、質問応答(7, 8)、対話と学習(9)、多言語間の機械翻訳(10)、知識の獲得(11, 12, 13)、仮説の生成(14)など言語に内在する知的な機能については情報処理関係の学会で報告している。

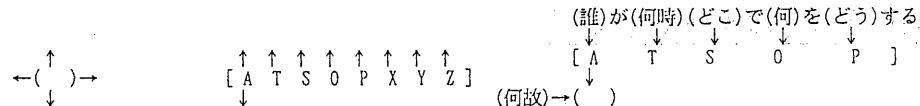
人間が行う推論、知識獲得、物語理解、学習は、聞き手と話し手との知的な対話によって行われるが、同時に話し手が過去の体験によって獲得した知識と頭の中で盛んに質問応答を行うことによって、このような知的な活動が行われていると考えられる。対話や質問応答は基本的には自然文と自然文とのパターンマッチングによって行われるので、自然文同士のパターンマッチングは知的な活動を支える基本的な機能である。これがどれだけ正確に行われるかによって、コンピュータ上で実現できる知的な機能のレベルが左右される。

ところが、自然言語では同じ内容でも実に多種多様に表現することができる。この自然言語の多様性が自然文同士のパターンマッチングを困難にしている。Schankはこの自然言語の多様性を避けるために、概念依存理論(Conceptual Dependency Theory CDT理論)(15)を提案し、明確な概念規定した記号で自然言語の意味を記述しようとしている。意味言語では自然言語でどのように多種多様に表現されていても意味が同じであれば、同じ意味構造で表現できるので、自然文の意味を意味言語で表現すれば自然言語の多様性から解放され、実質的な意味でパターンマッチングを行うことができる。しかも、意味言語はコンピュータと相性がよく、自然言語の意味をコンピュータ上でほぼ完璧に記述できるので、これによりコンピュータに上述の知的な処理を行わせることができる。

ここではどのようなプロセスで意味構造から多種多様な自然文が生成されるかを説明して、多種多様に表現される自然文の意味が同じであれば、何故同じ意味構造として表現できるかを明らかにする。

2. 意味言語の概要

意味言語を表現する要素として概念を表す語要素(META WORD)と基本的な意味単位を表す文要素P S (PRIMITIVE SENTENCE)を設定している。MWを” () ”で、P Sを” [] ”の記号で表す。単語はある概念を表すのでMWに単語を埋め込むと、MWは概念を表現することになる。P SにはA格(主格)、T格(時間格)、S格(空間格)、O格(対象格)、P格(述語格)の主要な五つの格と補助的な格、即ち、X格(補助格)、Y格(応答格)、Z格(全体格)を持っている。MWが図1(a)に示すように4本、P Sは(b)に示すように9本の結合用の手を持っている。横結合(論理結合)を” → ” “ ← ”で、縦結合(格結合)を” ↑ ” “ ↓ ”で示す。P SのA T S O Pの格にMWを結合させ、それをMWで受けさらにこれを別のMWと論理結合させると、(c)に示すように、1 H 5 Wの概念構造が得られる。これを基本的な意味単位とする。P格に” (あ) る ” “ (あ) る ” “ と ” “ と ” “ す ” “ る ”を埋め込むと、(d~f)に示すような” ~がある ” (P S - E X) “ と ” “ ~である ” (P S - I S) “ と ” “ ~をする ” (P S - D O) の三種類の基本意味単位が得られる。この意味単位が縦横に組み合わせて、複雑な意味が表現される。我々人間は日常生活の中であるまとまった概念構造を頻繁に用いているが、これに動詞、形容詞などを割り当てている。これ



(a) 語要素MW

(b) 文要素P S

(c) 最小意味単位 1 H 5 W

- (d) P S - E X 基本意味単位の”～がある”
 ↓ ↓ ↓ ↓
 [A T S O P] PS-EX
- (e) P S - I S 基本意味単位の”～である”
 ↓ ↓ ↓ ↓
 [A T S O P] PS-IS
- (f) P S - D O 基本意味単位の”～をする”
 ↓ ↓ ↓ ↓
 [A T S O P] PS-DO

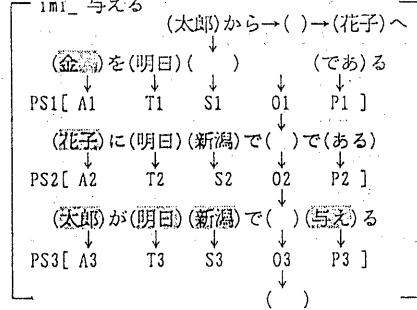


図2 “太郎が明日新潟で花子に金を与える”

の意味言語による記述

図1 意味言語の構成要素と基本意味単位

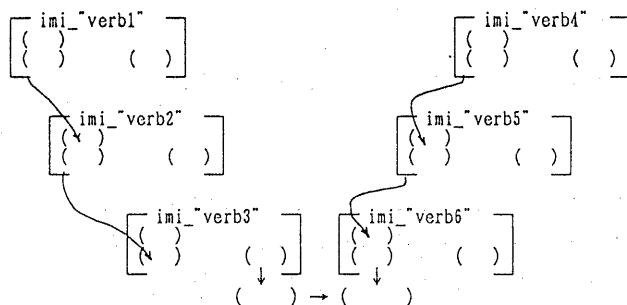


図3 文の意味構造

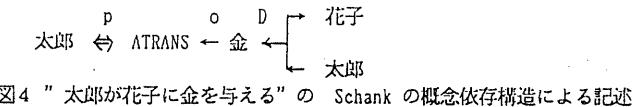
を意味フレームと呼んでいるが、図2にimi_“与える”の意味フレームを示す。文章の意味は図3に示すように意味フレームが縦横に結合されたような構造で表現される。

意味フレームのimi_“与える”は図2に示すような構造で表現されるが、スペースを節約するために、次に示すように単語が埋め込まれたスロットや説明に必要なスロットのみを意味フレームで指示された結合関係に従って並べたものをもちいることにする。この略記法に従えば、”太郎が明日新潟で花子に金を与える”は次のようになる。

[(太郎)が(明日)(新潟)で(花子)に(金)を(与える)]

[太郎が明日新潟で花子に金を与える]

3. Schank の概念依存構造と意味言語との比較



Schank の概念依存構造では”太郎が花子に金を与える”はく金が太郎から花子へ移動(ATRANS)するように太郎がする>という意味であるとして、同図のように表記する。それに対して意味言語では図2に示すように記述される。意味フレームimi_“与える”は3階層のPSで構成されている。PS1でく金が明日太郎のところから花子のところに移動する>ことを示し、PS2以上では、<花子は明日新潟で[金が明日太郎のところから花子のところに移動する]>という状態である>ことを示し、PS3以上では、<太郎が明日新潟でそのような状態にする>ことを表している。意味言語では自然文の単語や助詞と意味構造とが一一対応しているので、意味構造の変化に対応した多種多様な自然文の生成プロセスが手にとるように分かる。

4. 多様な自然文の例

[金は1億円である] (S1)

[太郎が明日新潟で花子にその金を与える] (S2)

[次郎はそのように思っている] (S3)

[私はそのように確信している] (S4)

S1～S4の文を次のように結合するとS5のような文が得られる。

[1億円]の
 [太郎が明日新潟で花子に(↓)金を与える]と
 [次郎が(↓)思っている]と
 [私は(↓)確信している]

”私は太郎が明日新潟で花子に1億円の金を与えると次郎が思っていると確信している”。 (S5)

ところが、次に示すS6～S10の文を人間はほぼ同じ意味であると理解することができる。

”太郎が明日新潟で花子に1億円の金を与えると次郎が思っているのは確かに” (S6)

”確かに太郎が明日新潟で花子に金を1億円与えると次郎が思いました” (S7)

”確かに太郎が明日新潟で花子に金を与えるのは1億円であると次郎が思った。” (S8)

”太郎ならば明日新潟で花子に1億円の金を与えると次郎が思ったのは確かに” (S9)

”次郎が花子にならば太郎が明日新潟で金を1億円を与えるという考えを持ったのは確かに” (S10)

自然文では上述のように同じ意味でも多様に表現できるが、これははその内のごく一部である。もう少し詳しく見るために上に示した單文毎にみると次のようになる。”金は1億円である”は”1億円の金”，”金1億円”と表現できるが、いずれも同じ意味である。さらに”太郎が明日新潟で花子に1億円の金を与える”という文で”1億円の金”的部分だけでも、次のように表現することができる。

”太郎が明日新潟で花子に金を1億円与える”

”太郎が明日新潟で花子に与える金は1億円である”

”金ならば太郎が明日新潟で花子に1億円与える”

”1億円の金は太郎が明日新潟で花子に与える”

”太郎が明日新潟で花子に1億円の金を与える”という文についても次のように表現することができる。

- ”太郎ならば明日新潟で花子に1億円の金を与える”
- ”明日になれば太郎が新潟で花子に1億円の金を与える”
- ”花子に対して太郎が明日新潟で1億円の金を与える”
- ”新潟に行って、太郎が明日花子に金1億円を与える”

また、”次郎はそのように思った”は”次郎はそのような考えを持った”，”次郎はそのような考えを持っている”とほぼ同じ意味である。この文と”私はそのように確信している”という文と組み合わせると、次のように多様に表現できる。

- ”私は次郎がそのような考え方を持っていると確信している”
- ”次郎がそのように思っているのは確かである”
- ”確かに次郎がそのように思っている”
- ”次郎がそのような考え方を確かに持っている”

そのような文を組み合わせるとかなりの数の文が得られることになる。わずか4つの単文からだけでもこのように多様な文が得られるので、これらの文をそのまま用いて、パターンマッチングを行おうとすると膨大な組み合わせになる。次にどのようなメカニズムでこのような多様な自然文が生成されるかについて説明する。

5. 多様な自然文の生成

この意味言語の意味構造は図3に示したように基本的には意味フレームが縦結合（格結合）と横結合（論理結合）によって多層に結合されているので、結合関係を変更することによって多種多様な自然文が生成できる。

5.1 論理結合による自然文の生成

5.1.1 一つの文を二つに分けて論理結合させる場合

[(太郎)が(明日)(新潟)で(花子)に(金)を(与え)る]という文から”(太郎)”を取り出し、[(太郎)(である)]という文にし、それと元の文を”()ならば→()”という論理関係で結合すると、次のような文が得られる。この文で”太郎”は2回表現されるが、意味として重要性が低いものを表現禁止にする。これを反転文字で示す。■は表現しても、しなくて良い文字列を示す。

[(太郎)(である)] [(明日)(新潟)で(花子)に(金)を(与え)る]
(↓)ならば→(↓)

となり、”太郎ならば、明日新潟で花子に金を与える”となる。”(明日)”を取り出して、[(明日)に(な)る]とすれば、次のような文になる。

[(明日)に(な)れ] [(太郎)が■■■■(新潟)で(花子)に(金)を(与え)る]
(↓)ば→(↓)

つまり、”明日になれば、太郎は新潟で花子に金を与える”という文になる。”(新潟)”を取り出して、[(新潟)に(行)く]にすれば、次のような文になる。

[(新潟)に(行)って] [(太郎)が(明日)■■■■(花子)に(金)を(与え)る]
(↓), →(↓)

つまり、”新潟に行って、太郎は明日花子に金を与える”という文になる。

5. 1. 2 格結合されている文を分割して、論理関係で結合する場合

”太郎が明日花子に金を与えるのは確かである”という文は次のような意味構造を持っている。”

[(太郎)が(明日)(新潟)で(花子)に(金)を(与え)る]の
↓
[()は(確か)(である)]

”[(太郎)が(明日)(新潟)で(花子)に(金)を(与え)る]”という文が<[(A)は(確か)(である)]>の(A)格に埋め込まれたものである。ここで、縦の関係を解除して、横の関係で結合すると次の文が得られる。

[()は(確か)(である)] [(太郎)が(明日)(新潟)で(花子)に(金)を(与え)る]
(↓)に→ (↓)

つまり、”確かに、太郎は明日新潟で花子に金を与える。”となる。

5. 2 格結合による自然文の生成

5. 2. 1 疑問文の生成

”～は～である”というP S - I S <[(A)は(O)(である)]>の(A)格に”(誰)”を、(O)格に文全体を埋め込めば、意味構造は次のようになる。

[(誰)が(明日)(新潟)で(花子)に(金)を(与え)る]の
↓
[()は(誰)(ですか?)]

(A)格に文全体を、(O)格に”(誰)”を埋め込めば、次のような意味構造になる。

[(誰)が(明日)(新潟)で(花子)に(金)を(与え)る]の
↓
[()は(誰)(ですか?)]

”誰が明日新潟で花子に金を与えるのですか？”や”明日新潟で花子に金を与えるのは誰ですか？”という文はこのようなプロセスで生成されたことが分かるが、何もこれは日本語だけではなく、英語などでも同じである。

5. 2. 2 一つの文から二つの文を生成し、それらをP S - I Sで結合する場合

”[太郎が明日新潟で花子に金を与える]”という文から”(花子)”を取り出し、[(花子)に(関)する]という文にして、P S - I S の(A)格に[(花子)に(関)する]を埋め込み、元の文を(O)格に埋め込むと、次のような文が得られる。

[(花子)に(関)して] [(太郎)が(明日)(新潟)で(金)を(与え)る]の
↓ ↓ ↓
[()は()] [()] (である)る
つまり、”花子に関しては、太郎が明日新潟で金を与える。”という文が得られる。”金”を取り出せば、次のような文になる。

[(金)に(関)して] [(太郎)が(明日)(新潟)で(花子)に(与え)る]の
↓ ↓ ↓
[()は()] [()] (である)る

つまり、”金に関しては、太郎が明日新潟で花子に与える。”という文が得られる。

5. 2. 3 格結合されている文を分解し、P S - I Sで結合する場合

[(確か)(である)]を(A)格に埋め込み、残りの文を(O)格に埋め込めば、次のような文になる。

[(確か)(である)]の [(太郎)が(明日)(新潟)で(花子)に(金)を(与え)る]こと
↓ ↓
[()は()] [()] (である)る

つまり，“確かに太郎が明日新潟で花子に金を与えることである。”という文が得られる。

5. 2. 4 繰り返し文による強調

”太郎が花子に金を与える”という同じ文を P S – I S の (A) 格と (O) 格に埋め込めば、次のような文になる。

(O) 格に埋め込んだ文の内、特に強調したい”（与え）る”という単語以外は表現禁止になる。このようなプロセスによって、”太郎が花子に金を与えるのは与えるのですが～”という文が得られる。”与えて→与えた”のように同じ文を論理関係で結合しても意味を強調することはできるが、”与えるのは与えるが～”のようにP S - I Sで結合しても、このように意味を強調する事ができる。

”1億円の金を与える”と”金を1億円与える”的関係については、副詞の意味構造から説明しなければならない。これにはもう少し複雑な説明が必要になるので別の機会に改めて報告する。

6. 自然文同士のパターンマッチング

もし、テキスト文が“確かに太郎が明日新潟で花子に金を与えるのは1億円であると次郎が思った。”というような文であっても、意味言語では、”次郎が太郎が明日新潟で花子に1億円の金を与えると思ったのは確かである”として、格納されている。また質問文が“次郎が花子にならば太郎が明日新潟で金を1億円を与えるという考えを確かに持ったのですか？”という文であっても、意味言語では”次郎が太郎が明日新潟で花子に1億円の金を与えると思ったのは確かであるか？”という文として格納されている。両者は全く同じ文であるので、その文同士のパターンマッチングはきわめて簡単になる。

意味言語ではどのような複雑な意味を表現している自然文であっても、図3に示すように意味フレームの縦と横の結合として表現され、意味フレームも図2に示したような単純で明確な意味構造を持っているので、パターンマッチングは簡単に行える。

7. まとめ

意味言語では意味が同じであれば同じ意味構造で表現されるので、自然文を意味言語に変換し、意味言語でパターンマッチングを行えば、自然言語の多様性が解消され、多様な自然文間のパターンマッチングを簡単に行うことができる。次のように意味構造の結合関係を変更することによって、多様な自然文が生成できる。

(1) 一つの文から単語を取り出してもう一つの文を作つて二つの文にしたり、格結合されている二つの文の結合関係を解除してできた二つの文を論理関係によって結合することによって、多様な文が生成できる。

(2) 一つの文から単語を取り出してもう一つの文を作って二つの文にしたり、論理関係や格関係によって結合されている二つの文の結合関係を解除してできた二つの文を P S - I S の (A) 格や (O) 格に結合することによって、多様な文が生成できる。

8. 参考文献

- (1) 安藤司文. 1991. [意味言語の提案] 情報処理学会, 自然言語処理研究会, 情処研報, Vol. 91, No. 80, pp. 73-80
 (2) 安藤司文. 1992. [意味言語から見た日本文の構造—使役文と受け身文], 国語学会平成4年度春季大会要旨, pp.

- (3) 安藤司文, 1992, [日本語文における論理構造], 国語学会平成4年度秋季大会要旨, pp. 44-49
- (4) 安藤司文, 1992, [自然言語に内在する論理構造], 情報処理学会 自然言語研究会 情報研報, Vol. 92, No. 2, pp. 59-66
- (5) 安藤司文, 1993, [格構造と論理構造との間での変換による多様な文の生成] 第106回日本言語学会大会配布資料
- (6) 安藤司文, 1992, [意味言語から見た主語について] 第104回日本言語学会大会配布資料
- (7) 安藤司文, 1991, [意味言語による質問応答]、人工知能学会、ヒューマンインターフェイスと認知モデル研究会、研究資料SIG-HICG-9102, PP. 43-52
- (8) ANDO, S., 1992, The Questioning and Answering Based on Meaning Language, International Symposium on Natural Language understanding and AI (ISKIT'92), p. 203-210, Kyushu Institute of Technology, Fukuoka, Japan.
- (9) 安藤司文, 1991, [意味言語による対話と学習]、認知科学会、学習と対話研究分科会、SIGLAL 91-2, 学習と対話, Vol. 91, No. 2, PP. 1-10
- (10) 安藤司文, 四元健仁, 1992, [意味言語による多言語間機械翻訳] 情報処理学会, 自然言語処理研究会, 情報研報, Vol. 92, No. 55, pp. 49-56
- (11) 安藤司文, 1991, [自然言語に内在する知識獲得のメカニズム] 情報処理学会、データベースシステム研究会、情報研報Vol. 91, No. 84, PP. 1-10
- (12) Ando, S., Masutani, M., 1993, [Knowledge Acquisition from Natural Sentences], Natural Language Processing Pacific Rim Symposium Fukuoka, Japan
- (13) 安藤司文, 益谷真, 守屋秀洋, 1993, [多様な自然文からの知識獲得], 情報処理学会 自然言語研究会 情報研報, Vol. 93, No. 101, pp. 25-32
- (14) 安藤司文, 1991, [自然言語に内在する仮説生成機構]、情報処理学会, 人工知能研究会, 情報研報, Vol. 75, No. 80, pp. 1-10
- (15) Schank, R. C., 1975 [Conceptual information Processing], North_Holland
- (16) 安藤司文, 1989, [認知プロセスと自然言語—自然言語の構造(1)] 認知科学学会第6回大会 発表論文集, pp. 1 28-129
- (17) 安藤司文, 1991, [自然言語から意味言語への変換] 情報処理学会第43回全国大会講演論文集(3), pp. 3-137

~138