自然言語処理 120－9
(1997. 7. 24)

係り受け関係を用いた副詞の分類と分類要素についての実験的評価

乾伸雄，小谷善行，西村恕彦
東京農工大学工学部電子情報工学科コンピュータサイエンスコース
〒183 東京都小金井市中町2－24－16
TEL. & FAX. 0423-87-4607
{ nobu, kotani, nisimura }@cc.tuat.ac.jp

あらまし
副詞は，対話システムのような自然言語インタフェースにおいて重要な役割を果たすが，意味的な側面及び構文的な側面から，その取り扱いは難しい．本論文では，副詞と副詞の係る文の関係について，いくつかの副詞の事例を検討し，副詞のシソーラスを構築する手法を提案する．副詞概念は，既存のシソーラスを用いて作成された分類要素付きシソーラスと呼ぶ決定木用いて，推定される．実験の結果，助動詞，動詞，助詞が推定に大きな影響を与えることがわかった．

キーワード 分類学習，コーパス，シソーラス，自然言語処理，副詞

Classifying Adverbs using Dependency Relations, based on an Existing Thesaurus

Nobuo Inui, Yoshiyuki Kotani, Hirohiko Nisimura
Dept. of Computer Science, Tokyo University of Agriculture and Technology
2-24-16 Nakamachi Koganei, Tokyo, 183
TEL. & FAX. 0423-87-4607
{ nobu, kotani, nisimura }@cc.tuat.ac.jp

Abstract

Adverbs play an important role in natural language interfaces. It is, however, hard to process them with proper syntactic and semantic considerations. To facilitate this process, in this paper, we characterize the relation between adverbs and sentences modified by them and propose a method for classifying adverbs into a hierarchy, using decision trees, called 'thesaurus with classification elements'. Experimental results show that the classification accuracy is greatly influenced by verbs, auxiliary verbs and particles.

key words Classification Learning, Corpus, Thesaurus, NLP, Adverb

1．はじめに

これまで，日本語に関する自然言語処理・学習の研究において，他の品詞に比べて，副詞は余り研究対象になることが多くないように思われるが，話者の主観的な表現を理解する上で，重要な働きを持っている[2]．副詞の意味を理解する上で，シソーラスを構築することは重要である．本報告では，コーパスから副詞のシソーラスを構築する手法について述べる．

コーパスから副詞のシソーラスを構築する上で問題となるのは，なにに基づいて副詞を分類し，シソーラスを構築するかということであろう．名詞や動詞は比較的良く研究されている．これは，動詞の取り得る成分が，人間が理解可能な上位下位概念に基づくシソーラスを反映していると考えられるからである[3]．また，名詞については共起確率に基づく手法が提案されている[7]．これに対して，副詞には未知の問題が多い．

そのため，本研究では既存のシソーラスの分類を得るために重要となる評価要素について，実験的に検証する．そのため，いくつかの副詞に関して分析を行い，その副詞の意味を特徴づける要素に関する検討を行う．この分析では，副詞を伴った文の自然さがどのような要素に決定されるかを明らかにする．

本研究では，副詞の係り先に存在する単語で副詞を分類する手法を用いた．これらの単語を用いて，既存のシソーラスからシソーラスを構築するための決定木を生成する．この手法では，データのスパース性の解消が問題となる．最も望ましい解決手法は，なんらかの分類に基づいて，係り先の単語を分類し，これによって一般化するという手法である．しかし，どのような一般化が望ましいかについては，未知の部分が多い[4]ので，本研究では，スパース性に関する議論を行うに留める．

本論文の構成は以下の通りである．2 章において，いくつかの副詞に関して，どのような要素で分類するのがよいのかを考察する．3 章では，本稿で提案する分類手法について述べ，4 章で実験方法を報告する．5 章で，実験結果の考察を行って，最後にまとめる．

2．副詞の分類要素に関する考察

本節では，副詞「まだ」，「もう」と「まったく」に関する意味的な考察について述べ，それが副詞の修飾する文に対して，どのような影響を持つのか考察する．

2．1　「まだ」の場合

副詞「まだ」は，文法的な分類としては「情態」，「程度」，「陳述」といった要素を持つ複雑な副詞である．意味的には，常にというわけではないが，次節で述べる「もう」の反対語となる．「まだ」は一般的には次のような意味を表す[5]．

「話者の立場から見て，現実の状態が想定していた状態に達していない」

「まだ」の表す主観的な意味は，状態遷移の点から説明される[1]．例えば，「まだ，彼は本を読んでいる」という文では，話者が「彼は本を読み終わっている」という状態を想定しているのに対して，現実は「読んでいる」という状態にあることを示している．「まだ」は，「ない」と呼応関係にあり，実際そのような用法が多いが，これは「まだ」の意味を決定する上であまり意味がない．単に，現在「ない」という状況下で，話者が「ある」ことを想定しているということが多いためであると考えられる．

「まだ，(現在) 彼は生きていない」という文は「まだ，彼は生きている」という文に比べて不自然な文である．しかし，「その時代に」という節を補ったり，末尾に「だろう」というムードを表す語をつけると，自然さが増すように思われる．

このような文の自然さは，「生きる」という動詞の意味に起因すると思われる．「まだ，生きる」のような文が自然であることを考えると，元来「生きる」には「まだ」が係るために必要とされる要素を持っているが，「ていない」がつくことにより，この要素が失われる．しかし，この性質は，動作動詞「読む」の場合には成り立たない．このことを考えると，「まだ」は，動詞に「ていない」がついたとき，完了の意味を

要求すると考えられる．「読んでいない」は「読む」動作の未完了を意味するとき自然である．通常，「生きていない」は「生きる」ということの未完了ではなく，否定を表すため，「まだ，（現在）彼は生きていない」は不自然であると分析される[1].

2．2　「もう」の場合

「もう」は，話者の主観的な判断と状態の移り変わりからみた場合，「まだ」の反対語となり，次のような意味を基本的には表す[5].

「話者の視点から見て，現実の状態が想定した状態を越えている」

例えば，「もう，彼は本を読んでいる」という文では，話者は「まだ，（現在）彼は本を読んでいない」ということを想定していたということになる．「まだ」の場合と同じで，「もう，彼は死んでいない」という文が不自然なことは，この「もう」の意味から説明することができる．「もう」に関しても，修飾する述語に対して一定の制約があることがわかる．

2．3　「まったく」の場合

「まったく」は，一般的には，係り先の述語がもつ否定要素に呼応して，否定を強調する．また，係り先が肯定の場合は，聞き手の意見に同意する表現となる．「全然」や「さっぱり」は「まったく」の同義語になるが，「まったく」はこれらの語よりも広い範囲で使われるといわれている[5].

否定表現の場合，「まったく」は広い範囲の述語に係ることができるが，例外がないわけではない．例えば，「まったく，彼は生きていない」という表現は不自然である．強調する場合は通常「本当に」が使われる．これに対して，「まったく，彼は生きられない」が自然な文であることを考えると，動詞「生きていない」の場合は，可能表現化することによって，「まったく」が係ることが可能になることがわかる．これは，「まったく」によって強調される事柄に制約があるためであると推測される．

しかし，この現象も動詞の種類に依存すると思われる．例えば，「読む」の場合は非可能，可能の状態両方で自然な文が作られる．

2．4　考察

以上，三つの副詞について，係ることが可能な述部の表現について概略を述べた．この述部の表現は，副詞の持つ意味に関係すると考えられる．副詞が係る述部の要素によって，副詞を分類することが可能であることが推測される．また，似ている述部要素を持つ副詞を集めることで，決定木を構築することが可能であるとも考えられる．これは，動詞の成分によって，動詞を意味的に分類することが可能なことに相当する．しかし，副詞の場合は，意味的に分類する述部の要素が何であるか，動詞の場合ほど明らかではない．次節以降では，副詞が係る文あるいは節に含まれる要素によってどのような分類が得られるか，実験的に評価する．そして，「良い」副詞の分類要素に関しての考察を行う．

3．副詞の分類手法

本節では，シソーラスとコーパスを用いた副詞の分類手法について提案する．従来の動詞に関する研究では，コーパスだけからシソーラスを構築することが多い．これは，動詞の意味的な分類が，その成分に強く依存しているという考えに基づいている．しかし，副詞に関しては，どのような要素に基づいて分類可能かはよくわかっていない．助動詞であったり，動詞の意味であったり，様々である．そのため，本稿で述べる手法は，いくつかの副詞に関しては，シソーラスがすでに与えられていると仮定する．そのシソーラスを生成するための分類要素を，コーパスを用いて求めることにする．これによって，特定の分類方法に基づくシソーラスが得られると期待される．

本稿で述べる分類方法は，「副詞のシソーラスは副詞の係り先の要素（これを分類要素と呼ぶ）によって決定される」，および「係り先の要素が類似する副詞は意味的に近い関係にある」という仮説に基づいてい

る．これは，単純過ぎる仮説であり，必ずしも成り立つとは限らないだろう．この点は，実験的にどの程度の分類が可能なのかを評価することで検証する．

分類実験システムは次の構成要素からなる．

a　副詞についての分類木の生成

b　副詞の分類推定

3．1　副詞についての分類木の生成

まず最初，既存のシソーラスを生成するために必要な分類要素を，コーパスから決定する．シソーラスは，次のように，上位下位関係を表した規則の集合によって表される．

$$\exists\{R,T,NT,P\}\{\ l \leftarrow r \in P \mid l \in R \cup NT,\ r \in T \cup NT\}$$

ここで，Rはシソーラスの最上位概念の集合，Tは副詞の集合，NTは中間段階の概念の集合，Pは概念の上位下位概念規則の集合をそれぞれ表す．このシソーラスは一般的なもので，上位下位概念がサイクルになることは禁止されるが，副詞から最上位概念までの経路が複数あってもかまわない．このとき，シソーラスを次のように再構成する．このシソーラスのことを分類要素付きシソーラスと呼ぶことにする．

$$\exists\{R,T,NT,P,E\}\ \{f(E1,E2), l(E1), r(E2), l \leftarrow r \in P \mid l \in R \cup NT,\ r \in T \cup NT,\ E1,E2 \subset E\}$$

Eは分類要素の集合，f(E1,E2)は二つの分類集合間の関係，l(E1)，r(E2)は，l，rという概念がE1，E2という分類要素によって構成されることを示している．fには様々な関係が設定できる．例えば，概念的な上位下位関係の場合は，下位概念が意味的には上位概念を包含しているので，lを構成する概念の集合をC1，rのそれをC2とおけば，$f(C1,C2)=C1 \subseteq C2$という関係が成り立つ．しかし，今回考えているEは分類要素なので，$f(E1,E2)=E1 \supseteq E2$のように考えることにする．つまり，上位概念に相当する架空の副詞が存在するとすれば，上位概念の副詞になればなるほど，それが係る先の表現に関する制約が緩くなるということを示す．

ここで，要素ベクトルという概念を導入する．コーパスから獲得すれば，分類要素に頻度情報を付随させることができる．この頻度情報を用いて，分類要素がn個あった場合は，n次元のベクトルを考え，これを要素ベクトルと呼ぶことにする．上位概念の要素ベクトルは下位概念の要素ベクトルの和で表すことにする．ここで，要素ベクトルの類似度を規定する．これは，二つの概念が類似しているかどうかを判定する基準として使われる．二つの要素ベクトルで共有する次元に関して取り出したm次元（$n \geqq m$）の要素ベクトルuとvの類似度を次のように表す．

$$\text{類似度} = m * \left| (u/|u|)+(v/|v|) \right|$$

類似度は，二つの単位ベクトルの和の距離として表現される．ただし，共有する次元が多いほど似ていると思われるので，mをかける．uとvが同じ方向のベクトルの時，類似度は大きくなる．お互いに共有する要素が少ない場合は小さくなる．

3．2　副詞の分類推定

分類要素付きシソーラスが構築されれば，それを決定木と考えることによって，新たな副詞の分類を求めることができる．今回のシステムでは，実験で用いるすべての副詞に関して分類がわかっているので，作成された分類木とのマッチングを行うことが可能である．これをコーパスなし推定と呼ぶことにする．コーパスなし推定は，シソーラスと副詞の最上位概念までの経路を最上位概念からマッチングすることで求める．そのマッチングした経路の中で最長経路のものを最良推定経路とする．これは，シソーラスで得られる階層に関する基準を与えるものである．

これに対して，コーパスあり推定は，要素ベクトル間の類似度に基づいて，副詞のシソーラス上での位置を決定する方法である．コーパスあり推定は，最上位概念から順番に，もっとも類似度が高い下位概念を検索することによって行う．途中で，要素ベクトルが直交した時点で探索を終了する．

コーパスなし推定を実験することで，シソーラスの一般性を検証する．つまり，分類要素の頻度のシソーラスに与える影響を測定する．

コーパスあり推定の結果は，分類要素の妥当性を示すことになる．もし，最上位概念から副詞に近い下位

概念を推定できるのであれば，その分類要素は妥当であるといえる．

4．実験

実験は，EDR 日本語コーパス[6]から抽出した，副詞が係る文または節を用いて行った．用いた EDR 日本語コーパスは，全文数が 208,156 文あるが，そのうち副詞の延べ出現回数は，50,359 回である．副詞の出現回数のばらつきは大きく，もっとも多頻出語は「さらに」で 1109 回である．1 回しか出現しなかった副詞は 765 単語ある．低頻度の副詞から分類要素付きシソーラスを構築するのは困難であると考え，今回の実験では，101 回以上出現する語を用いた．EDR 日本語コーパスでは，各単語に概念識別子がつけられている．これは，人間によって主観的に定められており，意味的なシソーラスを構成する．頻度が 101 以上の単語を構成するシソーラスについては，概念識別子の平均分枝数は，1.7 程度である．また，副詞から最上位概念までの距離は，最低が 2（つまり，副詞の概念識別子がそのまま最上位概念になっている），最大が 14，平均 8.9 程度である．

分類要素付きシソーラスを作成するに当たって，分類要素には一定の制限を付けた．これは，基本的に頻度が高い分類要素によって，その副詞の特徴が表されやすいと考えたからである．今回は，出現頻度が全体の３０％以上，１０％以上の2 種類の単語について，分類要素付きシソーラスを作成し，評価する．

分類要素は，係り先の文または節に属する次の6 種類の品詞に場合分けした．

動詞，助動詞，形容詞，形容動詞，名詞，助詞

多くの副詞は，基本的に述句に係り，述句の構成要素である動詞，形容詞，形容動詞，名詞，助動詞，助詞などの組み合わせに影響を与える場合が多いと考えられるので，実際にはこれらの組み合わせを用いるべきである．しかし，今回の実験では，収集された副詞の量が少ないので，単純に単一の品詞だけを扱うこととした．

評価用に用いる副詞は，10～19，20～40，40～60，60～80，80～100，101 回以上の頻度にグループ分けする．

5．実験結果・考察

グラフ1 ～4 に，実験結果を示す．縦軸は，真の副詞の分類階層長に対する推定された分類の深さの平均割合を表す．例えば，分類階層長 14 の副詞に対して，最上位概念から 10 階層を推定できたならば，71%ということになる．

コーパスなし推定で 101 回出現する副詞の場合の推定深度が1 にならないのは，分類要素の数を限定しているからである．分類要素を 10%以上にした方が 30%以上の場合よりも良い結果が得られるのは，分類要素の落ちが少ないためである．

コーパスあり推定はコーパスなし推定の結果を越えることはあり得ない．全体の傾向として，副詞の頻度が下がると推定深度も下がる．グラフ２，４ から形容詞による推定の精度が割合高いが，副詞の出現頻度が低くなると，助詞や動詞，助動詞の推定に比べて，下がる割合が高い．グラフ１，３ と２，４ の比較から，分類要素に関して頻度によって制限しないほうが良い結果が得られることがわかる．しかし，低頻度語が分類要素で支配的な役割を果たしてしまうと，特定の単語で副詞の分類が決定されることになる．グラフ２，４ で，４ の方が副詞の頻度に比べて，推定深度の低下が著しいことから，更に大規模なコーパスを用いた実験が望まれる．

これらの結果から，副詞の分類を行う要素としては，助動詞，動詞，助詞を用いることがよいと推測される．次に，頻度が８０～１００ の副詞について，どのような分類要素によって最も良い予測ができたかを表1 に示す．コーパスありに分類されている副詞は，コーパスなしの場合と同等の結果が得られたものである．この表から，副詞を特徴づける品詞に関する考察が得られる．例えば，「次第に」は形容動詞と伴ってでてくることが多く，それによって，他の副詞との違いが明確になる．

表1　副詞の意味を決定する品詞（頻度80～100）

コーパスなし 形容詞 動詞	いかにも 多く
コーパスあり 形容動詞 助詞 形容詞 助動詞	次第に いっそう 大変　相次いで しっかり ずいぶん 仮に 年々 だんだん どのように 当然　一挙に これまで もっぱら いまだに 結局　今や　おそらく はじめて 早く　いよいよ 最初に しゅとして なんとか 急に すでに なんとか

6．おわりに

既存のシソーラスを使って，コーパスから副詞を分類する手法について述べた．そして，副詞を分類する要素について考察した．副詞の分類には，助動詞，動詞，助詞が有効であることを示した．ただ，副詞の係り先にある単語そのものを使って分類しているので，スパース性が内在していることを留意するべきである．

今後の課題として，品詞の組み合わせによる分類精度の向上，副詞の分類の観点から見た，動詞や名詞の分類，それらの言語学的な意味づけ，副詞の持つ論理構造[2]の抽出を行う予定である．

謝辞　本研究の一部は，文部省科研費奨励研究(A)09780315の支援で行われている．関係各位に感謝する．

参考文献

[1] N.Inui, Y.Kotani, H.Nisimura : Expressing Knowledge of State Transition for Disambiguation of Japanese Time Adverb, MOU and MADA, Proc. of NLPRS'95 pp.652-657 (1995)

[2] N.Inui, S.Shimada, Y.Kotani, H.Nisimura : Handling Time Recognition for Friendly Natural Language Interface : A Case Study about using Temporal Adverbs in Japanese, Proc. of IEEE SMC, pp.696-701 (1995)

[3] 春野雅彦：最小汎化を用いたコーパスからの動詞格フレーム学習，「自然言語処理における学習」シンポジウム，pp.9-16 (1994)

[4] 大石，松本：共起する副詞を用いた動詞分類について，言語処理学会第2 回年次大会, pp.229-232 (1996)

[5] 森田良行：基礎日本語辞典，角川書店(1988)

[6] 日本電子化辞書研究所：ＥＤＲ電子化辞書1 ．５版仕様説明書 (1996)

[7] Y.C.Park, K.Choi : Automatic Thesaurus Construction using Bayesian Networks, Proc. of NLPRS'95, pp.228-233, (1995)