E_034

概念ベースを用いた場所連想システムの構築
Construction of the place association system using the concept-base

手原信太朗† 渡部広一† 河岡司†
Shintaro Tehara Hirokazu Watabe Tukasa Kawaoka

1. はじめに

近年，コンピュータは人間の道具として非常に便利な存在であるが，今後，人間のパートナーとしての役割がこれまで以上に期待されると考えられている．そこで，人間とコンピュータがより円滑にコミュニケーションを取れる手法が必要とされている．人間とコンピュータがコミュニケーションをとるには自然な会話が求められるが，その実現のためにはコンピュータに人間と同様の常識的判断のできる能力を持たせる必要がある．本稿では，その常識的判断の一つとして，場所に関しての常識的判断を行う「場所連想システム」を構築する．

人間は通常の会話から何について，また，どこで話をしているかなどの状況を理解することにより，自然な会話をしている．例えば，以下のような会話があった場合，

A：「運転免許証を見せていただけますか？」
B：「はい，どうぞ．スピード出し過ぎでしたか？」
A：「ええと，およそ時速140km出ていましたね」

これらの会話から場所は「高速道路」での会話ということが人間はわかる．

「場所連想システム」はそのような会話中の語から場所を連想することで，人間らしい自然な応答文を返すことを目的としている．

2. 場所知識

本研究における場所知識とは，場所語・主体語・目的語である．「場所語」とは，名詞単独で人が一般的に場所と考える語と定義する．また，「主体語」はその場所に何が存在するかを示した語，「目的語」はその場所が人間にとって何をする場所なのかを示した語と定義する．

例）「図書館」（場所語）には「本」（主体語）が
存在し「借りる」（目的語）を目的とする．

3. 場所連想システム

場所連想システムとは，ユーザから入力された語からそれに関係する場所語を出力するシステムである．つまり，主体語・目的語から適切な場所語を連想する．例を示すと，「本，借りる」から「図書館」が出力される．

本システムでは人手で作成した「場所語知識ベース」という場所に関する知識ベースを用いる．しかし，すべての場所語と主体語・目的語を知識として持たせることは困難であり，効率が悪い．そこで代表的な語のみを登録し，シソーラス[1]や概念ベース[2]により構築した連想システムを用いて知識の連想を行い，場所語知識ベースに登録されていない語に対しても対応できるようにする．

† 同志社大学大学院工学研究科
Graduate School of Engineering Doshisha University

シソーラスとは，一般名詞の意味的用法を表す約2700語の意味属性の上位下位関係，全体部分の関係を木構造で示したものであり，約13万語が登録されている．

概念ベース（以下「CB」と呼ぶ）とは，国語辞書から自動構築された汎用データベースである．本稿で用いるCBは，語（概念）と意味（属性）のセットが約9万語蓄積されているものである[2]．図1に概念ベースの一部を示す．また，概念と概念の関連の深さを定量的に表す手法を関連度計算[3]といい，その値を関連度という．関連度は0から1までの連続値で表され，値が高いほど二つの概念の関係が深いということを示す．概念A，Bの関連度$ChainW(A, B)$は以下のアルゴリズムにより計算する．

（1）まず，2つの概念A，Bを一次属性a_i，b_iと重みu_i，v_iを用いて，

$$A=\{(a_i,u_i)\mid i=1\sim L\}$$
$$B=\{(b_j,v_j)\mid j=1\sim M\}$$

と定義する．ここで，属性個数は重みの大きいものから30個を上限として展開するものとする．

（2）一次属性数が少ない方の概念を概念Aとし（$L\leqq M$），概念Aの一次属性の並びを固定する．

$$A=\{(a_1,u_1),(a_2,u_2),\cdots,(a_L,u_L)\}$$

（3）概念Bの各一次属性を対応する概念Aの各一次属性との一致度$(MatchW)$の合計が最大になるように並べ替える．ただし，対応にあふれた概念Bの一次属性$\{(b_{xj}, v_{xj})\mid j=L+1,\cdots, M\}$は無視する．

$$B_x=\{(b_{x1},v_{x1}),(b_{x2},v_{x2}),\cdots,(b_{xL},v_{xL})\}$$

（4）概念AとBとの関連度$ChainW(A, B)$は以下とする．

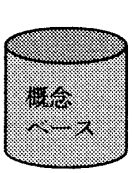
$$ChainW(A,B)=(s_A/n_A+s_B/n_B)/2$$
$$s_A=\sum_{i=1}^{L}u_i MatchW(a_i,b_{x_i})$$
$$s_B=\sum_{i=1}^{L}v_{xi} MatchW(a_i,b_{x_i})$$
$$n_A=\sum_{i=1}^{L}u_i$$
$$n_B=\sum_{j=1}^{M}v_j$$

また，概念Aと概念Bの一致度$MatchW(A, B)$は，一致する一次属性の重み（すなわち，a_i=b_jなるa_i，b_jの重み）の合計をそれぞれw_A，w_Bとするとき，次式で定義する．

$$MatchW(A,B)=(w_A/n_A+w_B/n_B)/2$$

3.1 場所語知識ベース

日常一般的によく使用される場所語を抽出し「代表語」として場所語知識ベース（以下「場所語 KB」と記す）に443語持たせておく．また，場所語KBにはシソーラスのノードから選ばれた代表語を分類した「分類語」が120語存在している．それらにも「主体語」「目的語」を与える（表1）．分類語と代表語は親子関係にあり，主体語と目的語の継承が可能となっている．



約９万語

概念	属性／重み
医者	医師／0.3, …, 治療／0.1
太陽	日光／0.2, …, 月／0.05
美術	芸術／0.4, …, 鑑賞／0.07
:	:

図１　概念ベース

表１　場所語知識ベースの一部

代表語	主体語	目的語	親分類語
検察庁	検察官 etc	捜査 etc	司法官庁

分類語	主体語	目的語
司法官庁	公務員，役人 etc.	司法 etc.

3.2 システムの流れ

場所連想システムの流れを図２に示す．

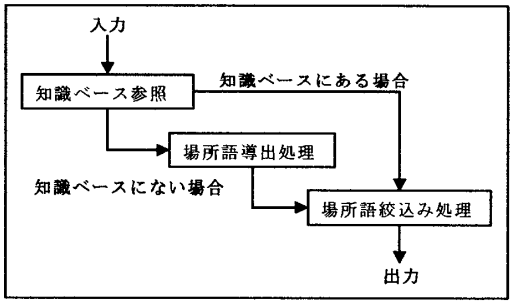


図２　場所連想システムの流れ

場所連想システムは，「場所語知識ベースの参照」「場所語導出処理」「場所語絞込み処理」の３つの処理を行う．入力語が主体語の場合は，場所語知識ベースの主体語を参照し，目的語の場合は場所語知識ベースの目的語を参照する．存在すればその場所語を取得する．知識ベースに存在しない場合は，場所語導出処理を行う．そうして得られた場所語すべてに対して，場所語絞込み処理を行い，最終的な出力をシステムの出力とする．

3.2.1 場所語導出処理

入力語は全て重要な情報であり，ひとつでも欠けると連想する場所は大きく変わってしまう．

例）入力「本・借りる」→「図書館」
　　入力「本」→「図書館・書店」
　　入力「借りる」→「図書館・銀行・レンタカー」

上記の例からわかるように，ユーザの意図する場所を連想するには全ての入力語から場所語を取得する必要がある．場所語 KB 参照のみでは，場所語 KB にない語は無視され，入力語に含まれていないのと同じ扱いになってしまう．これでは意図した場所が連想されにくくなる．

場所連想システムでは，入力語が場所語 KB 内の主体語・目的語に存在しない時，その入力語を未知語と呼ぶ．そこで，概念ベースや関連度計算などの連想メカニズムを用いて未知語と関連の高い場所語を導き出すことを考える．これが場所語導出処理である．

場所語導出処理の流れを図３に示す．入力語は主体語・目的語に区別するので，閾値は２種類用意した．

場所語導出処理の具体例を，未知語「乗員」を用いて説明する．まず，「乗員」と分類語「空港・駅・橋 etc」すべてと関連度計算を行う．閾値以上の候補ノード「空港・駅・鉄道」を取得する．取得したすべての候補ノードに属する候補リーフ「空港・管制塔・駅・駅舎 etc」をすべて取得する．「乗員」と取得したすべての候補リーフ「空港・管制塔・駅・駅舎 etc」で関連度計算を行い，閾値以上の候補リーフ「機関区，乗り場，都電，停留所 etc」を取得する．最終的に得られたそれらの候補リーフを関連の高い場所語として出力する．

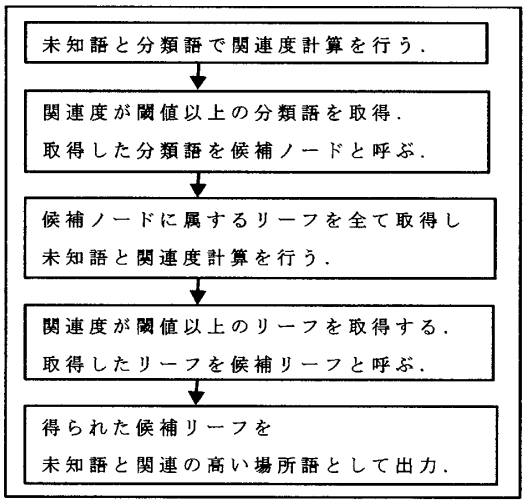


図３　場所語導出処理の流れ

3.2.2 場所語絞込み処理

場所語知識ベースの参照・場所語導出処理では各単語について高関連の場所語が導き出される．しかし，全てを見ると人間が想定する場所とは関連の弱い場所語やまったく関係のないと思われる雑音も含まれる場合が多い．

例）入力「らくだ・灼熱・オアシス・砂・砂嵐」
　　→出力「砂丘，砂州，砂場，砂漠」

上記の例で，「砂丘・砂漠」は関連が高いと思われるが，「砂州・砂場」といった語は関連が低い．場所語絞込み処理では，関連が低いと思われる語を除き，入力語のどれからも妥当と思われる場所語を選び出す．

場所語絞込み処理の流れを図４に示す．

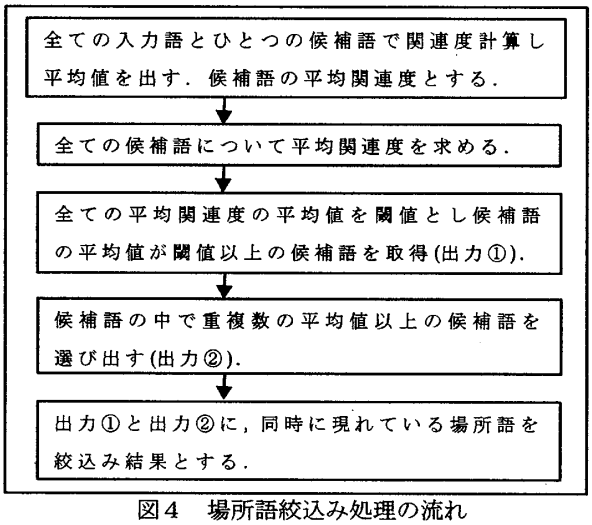


図４　場所語絞込み処理の流れ

場所語絞込み処理の具体例を以下に示す．

入力「シャンプー・石鹸・桶・湯船」から，場所語知識ベースの参照，場所語導出処理によって候補語「浴場・風呂場・浴場・風呂屋・銭湯・浴場・湯屋・洗い場・岩風呂・砂風呂・サウナ」が得られる．

候補語「浴場」とすべての入力語で関連度計算を行い，その平均（以下「平均関連度」と呼ぶ）を求める．この作業を，その他の候補語「風呂場・サウナ etc」で行いすべての語において平均関連度を求める．得られたすべての平均関連度の平均値を閾値として，閾値以上の候補語「浴場・銭湯・洗い場・風呂場」を出力①とする．

候補語の中には，違う入力語から出力されて重複した場所語が存在する．その場所語の重複した回数を重複数と定義し，すべての候補語の長複数から平均値を求め，重複数が平均値以上の候補語「浴場」を出力②とする．

出力①と出力②に同時に現れている場所語「浴場」を最終的な結果として出力する．

4. システム評価

場所連想システムの評価は想定場所とその場所の主体語と目的語をセットにして，人手によって作成した 183 セットのデータを使う．このデータの想定場所は重複するものが存在するが，それとセットになる主体語・目的語が完璧に重複するものはない．例を表２に示す．

表２　想定場所の重複例

想定場所	入力語
映画館	スクリーン, 座席,チケット, 上映
映画館	映画 ,ポップコーン, 椅子 ,チケット
映画館	鑑賞する,映画,楽しむ, 泣く,感動する,ムービー

この評価セットを場所連想システムにかけ，出力結果を入力語から正解（○），不正解（×），どちらともいえない（△）の三段階で評価した．出力された場所語が想定場所と異なっていても，人が見て入力語すべてから連想可能な場所であれば正解とした．まず，場所語導出処理を行う場合と行わない場合について評価・考察した．

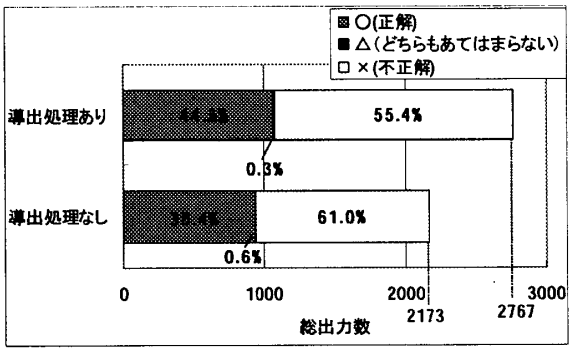


図５　場所語導出処理ありとなしの精度

図５から総出力を見ると，どちらも不正解が多く出力されている．それぞれの処理を見ると場所語導出処理なしでは正解の割合が総出力数の 38%であるが，場所語導出処理ありでは正解の割合が総出力の 44%であり，6%の向上が見られた．この 6%の精度向上は正解数が約 150 個増えていることを表しており，場所語導出処理が有用であることがわかる．

次に，場所語知識ベースの参照と場所語導出処理を行って得られた場所語に場所語絞込み処理を行う場合と行わないという場合について評価・考察した．

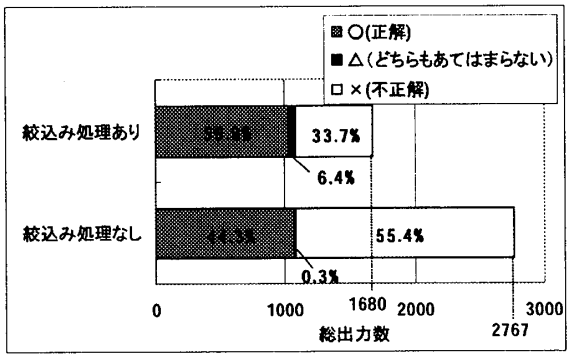


図６　場所語絞込み処理ありとなしの精度

図６から総出力数を見ると，場所語絞込み処理ありでは場所語絞込み処理なしと比べて約 2/3 に減少している．そして，正解の割合は約 15%増加し，不正解の割合は 22%減少していた．正解の割合が 15%増加したが，正解の数が約 40 個減少している．その原因は，不正解数の数が約 1100 個減少されたことにより不正解の割合が約 22%減少したことを表している．これにより，場所語絞込み処理は雑音に対して有用であることが確かめられた．

5. おわりに

本稿で提案した場所連想システムでは，主体語・目的語からの場所語の連想を実現できた．結果として，場所語導出処理によって場所語 KB にない未知語にも対応でき，場所語絞込み処理によって関連の低い場所語を削除する効果が見られた．

しかし，場所語導出処理では正解の割合が少なく，もっと多くの場所語を導出する必要がある．入力語の情報はどれも重要ですべての入力語から場所語を導き出すことが，より場所連想の幅を広げると考えられるため，今後更なる場所語導出処理の改良が必要である．場所語絞込み処理では不正解の数を大幅に削除させたが正解の数もやや減少する場合があるので，正解が削除されないように改良を加える必要がある．

6. 謝辞

本研究は文部科学省からの補助を受けた同志社大学の学術フロンティア研究プロジェクト「人間と生物の賢さの解明とその応用」における研究の一環として行った．

参考文献

[1] NTT コミュニケーション科学研究所監修，「日本語語彙大系」，岩波書店，東京，1997.

[2] 小島 一秀，渡部 広一，河岡 司，“ 連想システムのための概念ベース構成法－属性信頼度の考え方に基づく属性重みの決定”，自然言語処理，Vol. 8，No. 5, pp. 93-110，2002.

[3] 渡部 広一，河岡 司，“ 常識的判断のための概念間の関連度評価モデル”，自然言語処理，Vol. 8，No. 2, pp. 39-54，2001.