

領域性を考慮した結合による辞書の拡大

山口 智治 市山 俊治

日本電気(株) 関西C&C研究所

1 はじめに

自然言語処理システムにおいて、語の多義の解消は一つの基本課題である。この課題を解決するための有効な手段として、語意が一意に決定でき、多義が発生しないように、対象領域を極めて狭く限定するということがおこなわれてきた。

一方、対象領域の拡大はユーザの自然な要求である。[藤井88]などにみられるように、領域毎に別個の領域辞書を用意しておき、ユーザが選択する領域に応じて、領域辞書の入替えを行なって複数の領域にそれぞれ別個に対応する試みもなされている。これは領域毎に別々のシステムが用意されているのと同質的であり、互いに関連する次の2つの原因による操作-反応系間不整合が発生する。

- (1) 領域境界の不明確性
- (2) 領域辞書間の実装語彙の差

本稿では、語の意味決定にかかわる操作-反応系間不整合の減少のために、複数の領域辞書の結合による辞書の拡大を図る。

語彙の増加と多義の発生回避のために、結合前の領域を利用する。この結合辞書は、領域保存性と領域連続性を備え持つ。

また、領域境界の不明確性のために各領域辞書で未実装となっている語彙の存在と領域辞書間の実装語彙の差に着目して語彙を分類する。領域連続性に基づいて、この分類による潜在実装共有語を顕在化して結合辞書内での領域拡大をおこなう。

2 複数領域対象による操作-反応系間不整合

領域境界の不明確性とは、システムが対象とする領域の境界がその領域辞書に登録される語彙によって明確に定義されるのに対し、実世界では特定の領域を対象として想定しても、そこで使用される語彙を明確に制限し、領域の境界を明確な一線で定義することが困難なことである。

領域境界の不明確性により、対象とする領域の領域辞書に収録されている語彙と、ユーザがその領域で使用する語彙に差が現われ、ユーザの期待したシステムの挙動とシステムの実際の挙動が食い違うという問題が発生する。

また、複数の領域辞書を個々に用いる場合も同様に、それら辞書間の実装語彙の差によって領域毎にシステムの挙動が異なり、ユーザの期待する挙動と整合しないという問題が生じる。

このような挙動の不整合を“操作-反応系間不整合”と呼ぶことにする。

Extending Applicable Domain through Combining Separately Developed Domain Dictionaries  
Tomoharu YAMAGUCHI and Shunji ICHIYAMA  
Kansai C&C Research Lab., NEC Corp.

3 領域保存性、領域連続性を備えた結合辞書

**領域保存性:** 結合された辞書において、各辞書項目がどの領域辞書を起源とするかという情報を利用可能である場合、その結合辞書は領域保存性があるとする。

**領域連続性:** 複数の領域を対象とする辞書において、特定の領域を対象とした場合にも他の領域に属する項目を参照可能である場合、その辞書は領域連続性があるとする。

我々人間の言語理解過程をふりかえっても、知っている全ての事柄の内から、現在対象とする領域を想定し、その領域内だけで理解できなければ、近傍領域の情報を利用することがある[Car 83]。全ての領域の内からある領域を選択し、対象領域として想定すること領域保存性の利用で、近傍領域の利用を可能にするのが領域連続性である。

これらの性質を備えるための辞書結合は、複数ある領域辞書をひとつにまとめるだけではない。結合辞書は、もとの領域辞書が対象としていた領域と辞書項目の相互参照を可能にする情報をもつことによって、結合辞書の中に複数の領域が存在するという形態をとる。各辞書項目は、見出し語とその語が使用される領域のリストおよび各領域における意味(辞書内容)からなる(図1)。

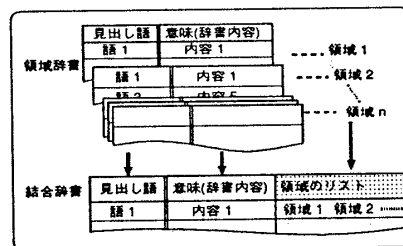


図1. 辞書項目の構成

4 領域と意味による語分類

複数の領域辞書の結合による辞書拡大の効果に貢献する語が、結合辞書と領域に関してどのような位置付けにあるかを明らかにするために、語彙を複数の領域間の関係において分類する。

まず、ある領域を対象とした時にそこで使用される語彙について考える。領域境界の不明確性のために、その語彙は、領域辞書に実装されたもの以外に未実装のものが存在すると考えられ、実装語と未実装語に大別される。

次に、複数の領域について考える。ある領域の実装語のうち、他のどの領域にも実装されていない語を実装領域固有語と呼ぶ。また、複数の領域で使用される語をそれらの領域の共有語とし、いずれかの領域で実装されているものを実装共有語、共有する領域のいずれでも未実装のものを未実装共有語と呼ぶ。実装共有語はさらに、複数の領域で実装されている実装顕在共有語と、ある

領域では実装されていないが、他の領域では実装されている実装潜在共有語にわけられる(図2)。

実装潜在共有語の存在が複数領域の操作-反応系間不整合の原因となる。よって実装潜在共有語を顕在化し、実装顕在共有語として扱えば、操作-反応系間不整合は減少する。実装潜在共有語を顕在化することは、共有語を実装していない領域の辞書をその実装潜在共有語を含むように拡大することである。

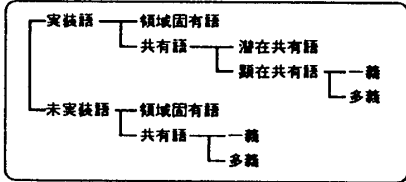


図2. 領域と意味による語分類

実装顕在共有語はそれぞれの領域における意味が共通であるか否かにより、さらは一義と多義にわけられる。図3では2つの領域について、使用される語彙の重なりを図化してある。実装顕在共有語は、実装顕在共有一義語と実装顕在共有多義語にわけられる。

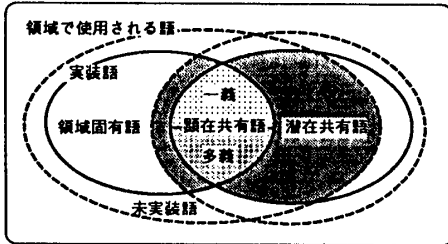


図3. 語彙の重なりと語分類

## 5 領域の更新

### 5.1 意味決定

結合辞書を用いた語の意味決定は、一つの主対象領域に属する辞書項目のみでなく、いくつかの副対象領域の辞書項目も参照することが許される。ここでは、事前に主対象領域は指定されず、入力文から適切な主対象領域を決定しなければならないものとする。

簡単な例として、文を語群としてとらえ、語群の属する領域から主対象領域を決定する場合について考える。一領域内では語の意味は一意に決定可能であるとする。このとき語群が属する領域について以下の分類ができる。

- (a) 文中の全ての語が共通に属する唯一の領域がある。
- (b) 文中の全ての語が共通に属する複数の領域がある。
- (c) 文中の全ての語が共通に属する領域が無い。

(a) の場合は、共通に属する唯一の領域を主対象領域として、各語の意味が決定される。(b) のとき、文は曖昧である。文中の全ての語が顕在共有一義語である場合は全ての語が共通に属するいずれの領域を主対象領域としても問題なく意味が決定できる。しかし、文中の語のいくつかは顕在共有多義語であったなら、どの領域を主対象領域にするかによって文の解釈結果に差が生じる。このときはユーザの介入を要するが、主対象領域の候補を提示することができる。(c) の場合は、どの領域を主対象領域としても主対象領域に属さない語が文中に存在する。この主対象領域に属さない語が属する領域を副対象領域と呼ぶ。

### 5.2 領域更新

前項の(c)の場合、文中に主対象領域に属さない語が存在する。これらの語は、主対象領域でも使用される語でありながら主対象領域の辞書に不足した潜在共有語であると推測できる。

潜在共有語の意味は副対象領域におけるものが候補となる。その意味が主対象領域においても正しいければ、その語は主対象領域にも属するものとして顕在共有一義語となる。語の意味が誤っていれば、主対象領域における正しい意味とあわせて辞書を更新し、顕在共有多義語となる(図4)。

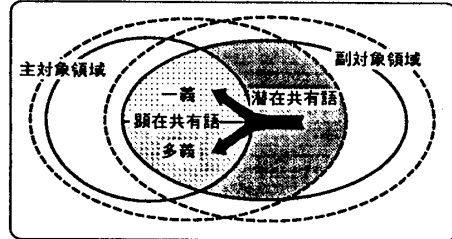


図4. 潜在共有語の顕在化

この場合には、主対象領域では未知語である語に対して、副対象領域における意味を提示した上で、違いを明らかにする定義を求めることができ、ユーザインタラクションにおける可視性と可制御性[谷91]の向上にも効果がある。

このような更新によって、各領域で本来共通に用いられるはずの情報特定の領域において不足していた場合に、それを補う効果が発生し、このとき結合辞書中で主対象領域に相当する辞書が拡張される。

## 6 おわりに

制限された領域を対象として別個に作成された複数の領域辞書の結合による自然言語処理システム用辞書の拡大について述べた。領域辞書が対象としていた元の領域と辞書項目の相互参照を可能にする情報を結合辞書に記し、語の意味決定に際して参照することによって、辞書の拡大による曖昧性の増加を回避可能にする。また、領域辞書の未実装語の差に着目して、結合辞書中の複数領域を相補的に用いる。入力表現に応じて主対象領域と副対象領域から検索を行ない、潜在実装共有語を顕在化することによって、対象領域拡大が可能になる。

本稿では、文を語群として捉えた時に各語の属する領域のみから主対象領域を決定する例をあげたが、語の文内での重要度などに応じて、重要な語の属する領域を優先的に主対象領域にするような枠組についても検討する必要がある。

### [参考文献]

- [藤井 88] 藤井則夫、銀林純、村上憲稔：会話によるコンサルテーションモデル構築ツール, FUJITSU, Vol.39, No.3, pp.223-228, 1988.
- [Car 83] J.G.Carbonell: "Learning by Analogy:Formulating and Generalizing", Plans from Past Experience", in Machine Learning, Vol.1, Chap.3, R.S.Michalski et al.(eds.), Morgan Kaufmann, 1983.
- [谷 91] 谷幹也, 市山俊治, 村木一至: 自然言語インタフェースにおけるユーザインタラクション, 情処 43 回全大論文集 1H-1, 1991.