シソーラス作成支援ツールに関する基礎的考察

1P-1

堺　和宏，徳永健伸，田中穂積
東京工業大学

1．はじめに

現在，著者らは，自然言語の意味解釈に用いることを前提として，上位／下位関係を中心とした大規模なシソーラスの作成を進めている．この作成過程において，概念の追加・削除にともなう冗長性の排除や一貫性の保持が重要になる，現在，これらの作業は人手によっておこなっているが，シソーラスの規模が大きくなるにつれて，シソーラス作成を支援するツールが必要な状況になってきている．本稿では，ツール作成にあたり，特に冗長性の排除に関して，基礎的な考察を加える．

2．上位／下位関係シソーラス

上位概念の持つ性質が，下位概念にも伝搬されるとき，上位概念と下位概念との間に上位／下位関係があるという．このように，概念が上位／下位関係で結ばれた知識の体系を，上位／下位関係シソーラス（以下，簡単のためシソーラスと呼ぶ）という．

ここで，概念および上位／下位関係を，それぞれ，ノードおよびリンクに対応させると，シソーラスはループのない有向グラフ（N，L）として考えることができる．ここでNはノード（概念）の集合，Lはリンク（上位／下位関係）の集合である．

リンクには，ノード間を，直接結ぶ「直接リンク」と一つ以上のノードを経て間接的に結ぶ「間接リンク」がある．ノードaとその上位のノードbが直接リンクまたは間接リンクで結ばれるとき，それぞれ $a < b$，$a \ll b$ と書くことにすると，次の性質が成り立つ．

1. $a < a$
2. $a < b \rightarrow a \ll b$
3. $a \ll b,\ b \ll c \rightarrow a \ll c$

つまり，$\ll$ は反射律，推移律のなりたつ二項関係となっている．

3．多重継承における冗長性

筆者らのシソーラスでは多重継承が許すため，作成の過程で冗長なリンクを生ずる可能性がある．現在，下位概念は上位／下位関係のある上位概念の性質を無条件にすべて継承する．つまり，どのような経路を通ってきても継承する性質は変わらない．このように，継承の経路の意味を考慮しない場合，冗長性を次のように定義する．

<u>定義：</u>

あるリンク$(a \ll b)_1$に対し，それ以外のリンク$(a \ll b)_2$が存在するとき，リンク $a \ll b$ は冗長である．

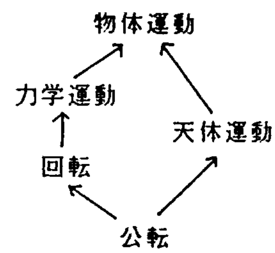

図1　冗長なリンクの例

例えば，　図1において、「公転」と「物体運動」の間には

1）公転＜回転＜力学運動＜物体運動
2）公転＜天体運動＜物体運動

という二つの間接リンクがある．このとき，「公転 $\ll$ 物体運動」は冗長である，という．

この定義によって，「最小シソーラス」を次のように定義する．

<u>定義：</u>

冗長なリンクがなく，かつ，入力された情報を失わないようなシソーラスを最小シソーラスという．

4．基本操作

筆者らの目的のひとつは，シソーラスを最小に保ちながら作成していくことである．以下，そのための基本操作のアルゴリズムを与える．シソーラス作成の過程で基本となる操作は

1. ノードの追加
2. ノードの削除
3. 直接リンクの追加
4. 直接リンク削除

Tools for the construction of a thesaurus : Basic consideration
Kazuhiro SAKAI, Takenobu TOKUNAGA, Hozumi TANAKA
Tokyo Institute of Technology

の四つである．

まず，準備として新しく次のような性質を持つ二項関係 $<_a$ を導入する．

$x << y,\ x \not<< a,\ y <_a z \ \rightarrow\ x << z$

（ただし，/<< は << が成立しないことを表す）

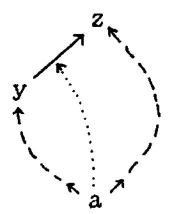

図２　二項関係 $<_a$ の例

例えば，図２においてｙ，ｚ間の矢印とａからこの矢印へ向かう点線の矢印の組がこの関係を表しており，ｙ，ｚ間の矢印を経て上位ノードの性質の継承をできるのは，継承の経路にａを含まない場合に限る，ということを意味している．この関係を利用して，上記した冗長性を排除することができる．図１の例に適用すれば，

（力学運動）$<_{(公転)}$（物体運動）

または

（天体運動）$<_{(公転)}$（物体運動）

のいずれか一方（どちらでもよい）を与えることで，公転と物体運動の間の間接リンクはひとつだけとなる．

最小シソーラスを保つ基本操作のアルゴリズムは以下のようになる．ただし，（Ｎ，Ｌ）を操作前，（Ｎ’，Ｌ’）を操作後の最小シソーラスとする．

(1)ノードの追加：(N’,L’)＝addnode(ａ,(N,L))

1.ａ∉Ｎならば，N’＝N＋{ａ}，L’＝L．
　そうでなければ，N’＝N，L’＝L．

(2)ノードの削除：(N’,L’)＝delnode(ａ,(N,L))

1.Lo＝{ｘ|ｘ<ａ}，Up＝{ｙ|ａ<ｙ}．
2.L’’＝L－{ｘ<ａ|ｘ∈Lo}-{ａ<ｙ|ｙ∈Up}．
3.N’’＝N－{ａ}
4.(N’,L’)＝for all (ｘ,ｙ) where ｘ∈Lo,ｙ∈Up
　addlink(ｘ<ｙ,(N’’,L’’))．

(3)リンクの追加：(N’,L’)＝addlink(ａ<ｂ,(N,L))

1.ａ,ｂ∈N ならば，2へ．そうでなければ，ノードを新しく追加する．
2.Ｃ＝{(ｘ,ｙ,ｔ)|ｘ<ｓ<<ｔ<ｙ,
　ｘ<<ａ,ｂ<<ｙ,ｓ/<<ａ,ｂ/<<ｔ},
　Ｘ＝{ｘ|(ｘ,_,_)∈Ｃ},Ｙ＝{ｙ|(_,ｙ,_)∈Ｃ}
　として，Ｃ＝∅ならば，L’＝L＋{ａ<ｂ}として終了．そうでなければ，3へ．
3.L’’=L；
　for all (x,y,t)∈C
　L’’＝L’’－{ｔ<ｙ}＋{ｔ$<_x$ｙ}+{ａ<ｂ}．
4.l∈Ｘ，ｕ∈Ｙについて，l/<<ｕ,ｓ$<_l$ｕ となるものがなければ，L’＝L．　あれば，
　L’＝L’’-{ｓ$<_l$ｕ}-{ｓ<ｕ}．

(4)リンクの除去：(N’,L’)＝dellink(ａ<ｂ,(N,L))

1.Ｃ＝{(ｘ,ｙ)|ｘ<<ａ,ｂ<<ｙ,ｔ$<_x$ｙ,ｂ/<<ｔ},
　Ｘ＝{ｘ|(ｘ,_)∈Ｃ}，Ｙ＝{ｙ|(_,ｙ)∈Ｃ}
　として，Ｃ＝∅ならば，N’＝N,L’＝L－{ａ<ｂ}として終了．そうでなければ，2へ．
2.Ｙの要素に対して，それぞれｔをひとつ定め，
　L’’＝L－{ｔ$<_x$ｙ}＋{ｔ<ｙ}－{ａ<ｂ}．
3.l∈Ｘ，ｕ∈Ｙなるについて，l<<ｕが冗長であるとき，Ｓ＝{ｓ|l<<ｓ<ｕ}として，Ｓの要素をひとつだけ残して，
　L’＝L’’－{ｓ<ｕ}＋{ｓ$<_l$ｕ}．
4.ａ,ｂ が孤立しているならば，除去．

５．おわりに

継承の経路が意味を持たない場合について冗長性および冗長性のない最小シソーラスを定義し，最小シソーラスを保ちながらシソーラスを成長させていくための基本操作について述べた．

継承の経路が意味を持つ例としては，継承の例外を取り扱う場合がある．この場合，例外のある関係のみを特別なものとして扱い，例外のない関係については上記の場合と同様に冗長性を排除することができると考えている．また，複数の種類の上位／下位関係を与え，各関係ごとに継承の制限を加えるということも考えられる．この場合には継承のパスがすべて意味を持っているため，冗長性を考える必要はなくなる．

多重継承の問題点には，冗長性のほかに，無矛盾性の維持があるが，これについても現在検討中である．

参考文献

Etherington,D.W. and Reiter,R.,
"On inheritance hierarchies with exceptions",in Proc. AAAI-83,Washington,DC,pp104-108, 1983.

Lenat,D.,Prakash,M. and Shpherd,M.,
"CYC :Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks",AI Magazine Vol.6,No.4,1986.

田中穂積，小山晴生，奥村学，
"知識表現形式ＤＣＫＲとその応用"，
コンピュータソフトウェア,Vol.3,No.4,1986.

Touretzky,D.S.,
"The Mathematics of Inheritance Systems",
Morgan Kaufmann Publishers,Inc.,Los Altos, California,1986.