分類語彙表からの韓国語シソーラスの作成

黄 道三[†] 長尾 眞[†]

シソーラスは自然言語処理の研究において重要なデータあるいは知識情報としての役割を果たしてきた．国立国語研究所の『分類語彙表』は数多くの言語研究に利用され，特に意味解析の研究の基礎資料として種々の貢献をしてきている．しかしシソーラスの作成はこれまで人手による以外に方法がなく，長期間を要し，コストの高いものであった．われわれは，韓国語の言語処理研究に韓国語のシソーラスが欲しいが，これをわれわれが人手で作ることはできない．そこで，われわれは日本語シソーラス，特に『分類語彙語』を基として韓国語シソーラスを半自動的に作ることを考えた．本稿では，あらかじめ作成されている日韓対訳辞書と韓国語意味辞書と『分類語彙表』とを用いて，韓国語シソーラスを半自動的に作成した結果とその評価を示す．ここでは，韓国語と日本語の類似性に基づいた分類番号の付与方法，分類番号と意味属性との対応による分類番号の付与方法，意味属性の類推を用いた分類番号の付与方法を示した．

Construction of a Thesaurus for Korean from a Thesaurus for Japanese

DOSAM HWANG [†] and MAKOTO NAGAO [†]

Thesaurus has traditionally played an important role as a source of knowledge in Natural Language Processing research. However, the cost of manual construction of a thesaurus is very high. This fact has urged attempts to automate the process. A thesaurus for Japanese, Bunruigoihyou, has played an important role in linguistics research. We present in this paper a method to produce a thesaurus for Korean from a machine translation dictionary and a thesaurus for Japanese. More specifically, we transferred the semantic markers to the classification number of the thesaurus, not only by direct corrspondence, but also by the similarity among the semantic markers. We constructed a thesaurus for Korean by using a Japanese to Korean machine translation dictionary and Bunruigoihyou, a thesaurus for Japanese, and evaluated the result.

1. はじめに

19世紀中ごろにRogetによって作成されたシソーラスは最近の自然言語処理の研究になくてはならない重要なデータあるいは知識情報として利用されるようになってきている．1964年作成された国立国語研究所の『分類語彙表』[1]も日本語の言語処理研究にいろいろと利用されている，今日電子版で利用できる数少ないシソーラスの一つであり，1980年代からの日本語関連の機械翻訳研究の発展に多くの寄与をしてきた．本論文はわれわれが計算機の上で利用できるこのシソーラスを用いて行った他言語，特に韓国語のシソーラスを半自動的に作成する研究の結果について述べたものである．

今日までシソーラスは人手によって作られてきたが，そのためには膨大な開発費とマンパワーを投入しなければならないし，かなりの時間がかかっている．しかし，シソーラスは自然言語処理の研究のための基礎データであるから，電子版シソーラスをもたない韓国語などにおいて何らかの形でシソーラスを作る必要がある．

そこで，本稿では，あらかじめ開発されている日韓対訳辞書[2]，韓国語意味辞書[3]と増補版『分類語彙表』*（以後『』なしで呼ぶことにする）とを用いて，韓国語シソーラスを半自動的に作成することを考えた．ここでは，韓国語と日本語の類似性に基づいて分類番号を韓国語に付与する方法を基本とし，分類番号を意味属性に対応させて分類番号を韓国語に付与する方法と意味属性を類推することによって韓国語意味辞書の意味属性を分類語彙表の分類番号に変換する方法を試みた．そして具体的に約12,000語の韓国語シソーラスを作成し，その結果を評価した．

† 京都大学工学部電気工学第二教室
Department of Electrical Engineering, Faculty of Engineering, Kyoto University

* 本論文では国立国語研究所が『分類語彙表』を増補してオンラインデータの形式で作成したものを使用した．

Jkey	*SynA*	$Ktran_1$	$Ktran_2$	$Ktran_3$

Jkey: 見出し語，*SynA*: 構文属性，*Ktran*: 対訳語

図1 日韓対訳辞書のレコード構成
Fig. 1 A record of a Japanese to Korean dictionary.

Kkey	*CW*	$SemA_1$	…	$SemA_n$	*CI*	*AI*

Kkey: 見出し語，*CW*: 概念記号
SemA: 意味属性，*CI*: 格情報・機能属性，*AI*: 隣接情報

図2 韓国語意味辞書のレコード構成
Fig. 2 A record of a Korean semantic dictionary.

表1 使用した日韓対訳辞書の見出し語の品詞別単語数
Table 1 Numbers of words by parts of speeches.

品詞	名詞類	動詞	形容詞	形容動詞	副詞	接続詞	合計
見出し語数	21,166	3,616	293	880	53	9	26,017
対訳語数	21,356	4,321	293	1,760	53	9	27,792

2. 機械辞書と分類語彙表

2.1 日韓対訳辞書

ここで用いた日韓対訳辞書は日韓機械翻訳システムのために開発されたもので，一つのレコードは**図1**のように日本語の見出し語，構文属性，韓国語の対訳語の三つの部分からなっている．見出し語には日本語の単語が書かれていて，その単語の品詞的役割に従って与えられている属性が構文属性として書かれている．構文属性は約100種類に分類されている．対訳語には見出し語に対しての韓国語の対訳語が書かれているが，見出し語が多訳性を持っている場合には，対訳語が最大三つまで記述できるようになっている．例えば，日本語の「本」に対しては「책(book)」，「권(volume)」，「본(this)」の三つの対訳語が記述されている．

この辞書には約39,220語の見出し語があり，これらに対して約42,280語の対訳語が登録されている．ここでは，連語，熟語，特殊文字などの翻訳のために登録されている単語とタイプミスの単語を取り除いた約26,020語とその対訳語の約27,800語を対象にした．対象単語の品詞別単語数は**表1**のようである．

2.2 韓国語意味辞書

韓国語意味辞書には**図2**のように，韓国語の見出し語，概念記号，意味属性，格情報および機能属性，隣接情報が書かれている．概念記号とは単語がもつ意味を表現するもので，例えば，「나(私)」，「간다(行く)」「학교(学校)」の場合，おのおの[I]，[go]，[school]のように英語で表されている．また，意味属性は大きく二つのサブ属性から構成されていて，第1番目の文字は名詞，動詞，形容詞，副詞などの構文属性を，残りの文字列は意味属性を表し，名詞99個，動詞53個，形容詞31個，副詞28個の意味属性が分類されていて，約86,220語が登録されている．

図3に示すように，意味属性の体系としては，レベル1に品詞分類項目を，レベル2に意味分類項目をもち，レベル3から意味属性が細分類されていく階層構造になっている．一つの単語にはほとんど複数の意味属性が付与されている．例えば，「나(私)」には「NANI(動物)」，「NCON(具体物)」，「NCRE(生物)」，「NHUM(人間)」などの四つの意味属性が記述されている．構文情報としては格情報と機能属性があり，格情報は用言に対してどんな格要素でどのようなパターンとして表されるかが格属性で表現されている．一つの格要素は「AGENT. GA」* のように格と格助詞のペアで表現されている．機能属性の部分には単語の文法機能によって分類されている属性が与えられている．

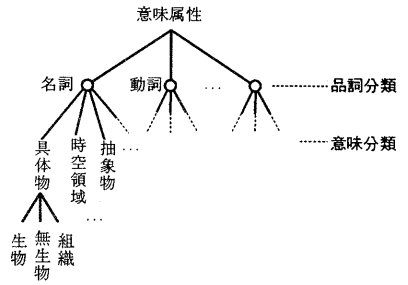


図3 意味属性の体系
Fig. 3 The categories of semantic attributes.

IN_1	…	IN_5	*TN*	*SN*	*Jkey*

IN: 分類番号，*TN*: 段落番号，*SN*: 段落内番号，*Jkey*: 見出し語

図4 分類語彙表のレコード構成
Fig. 4 A record of the Japanese thesaurus: Bunruigoi-hyou.

2.3 分類語彙表

分類語彙表の分類項目は**図4**のように，5桁の分類番号，段落番号，段落内番号の三つに分けられている．分類番号の左から第1桁は大分類として主に品詞に従って体の類，用の類，相の類，その他などの四つの類

* 「GA」は韓国語の主格助詞「가(が)」を表す．

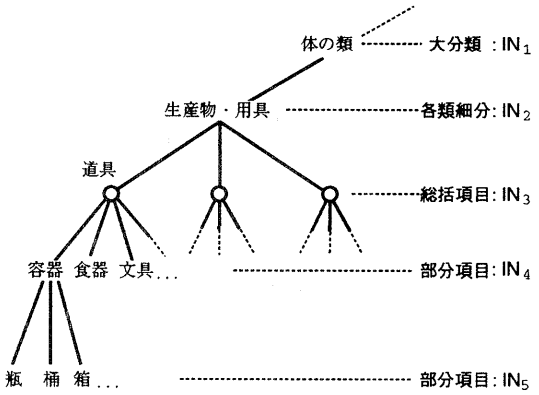


図5 分類番号の体系
Fig. 5 The components of a thesaurus classification number.

表2 分類語彙表の大分類別単語数
Table 2 Numbers of words by the classification numbers.

大分類	体の類	用の類	相の類	その他	合計
収録語数	44,632	8,683	6,987	482	60,784

に分けられている．次の第2桁はおのおのの類に対して，意味的に細分されていて，第3桁から第5桁まではもっと細かく意味分類された細分意味番号が付与されている．そして，分類番号の体系は**図5**のように深さ5の階層構造で示すことができる．現在，収録語数は60,784語に至っており，分類項目の大分類別に**表2**のように収録されている．分類番号は総計798項になっている．

3. 韓国語シソーラスの作成方法

3.1 単純検索による方法(方法1)

日本語と韓国語には意味上の使い方がほとんど完全に一致すると思われる単語が多い．このような現象は，漢字が含まれている日本語単語の場合に多く見られる．そこで，われわれは，

- 2文字以上の日本語単語で1文字以上の漢字を含み，
- 日韓対訳辞書で唯一の韓国語単語が対訳語として与えられている

単語については意味上の使い方が日韓でほとんど同じであると仮定し，分類語彙表でその日本語単語が何カ所に現れ，複数個の分類番号をもつ場合もそれらすべての分類番号を対応する韓国語単語に与えるようにした．例えば，「学生」という日本語単語は上記の条件に合っているので，「学生」の分類番号である「12410 2 60」と「12410 3 50」をその対訳語である「학생」という韓国語単語に付与する．これを方法1と呼ぶ．4章の表6に示すように，この対応関係は非常に精度の高いものであることがわかり，われわれの仮定が妥当なものであることが明らかとなった．

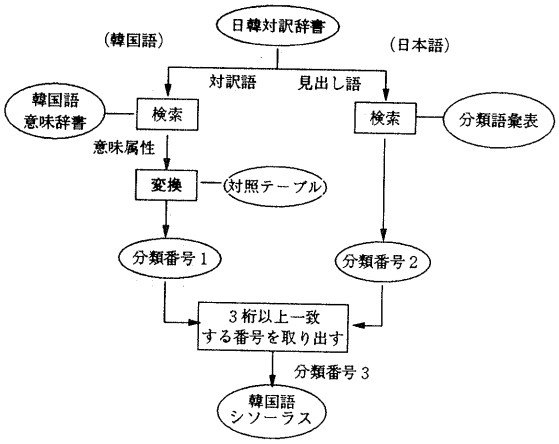


図6 韓国語シソーラスの作成の流れ
Fig. 6 Flowchart for construction of Korean thesaurus.

3.2 意味属性の分類番号への変換による方法(方法2)

3.2.1 韓国語シソーラスの作成の流れ

日韓対訳辞書の見出し語(日本語)をキーとして，分類語彙表からその単語の分類番号を検索する．ところで，多くの単語は複数の意味を含んでいるため，分類語彙表には一つの見出し語に対して複数の分類番号が付与されている．また，両言語の意味上の使い方が一対一に完全には一致しないため，日韓対訳辞書にも一つの日本語単語が複数の韓国語単語に対訳されているものがある．したがって，これらの分類番号を韓国語の対訳語に単純に付与してはいけない．例えば，日韓対訳辞書では日本語の「本」が韓国語の「책(book)」，「권(volume)」，「본(this)」などに対訳されており，分類語彙表では「本」に「11960(単位)」，「13160(文献)」，「14590(帳)」，「31000(こそあど)」など複数の分類番号が付与されている．

この分類番号をおのおの意味的に一致する韓国語に付与するために**図6**に示す方法をとった．まず，韓国語に対する意味属性が必要となるので，韓国語意味辞書を用いた．そして日韓対訳辞書の対訳語をキーとして，韓国語意味辞書からその意味属性を抽出する．と

ころが，この意味属性と分類番号との分類体系が違っているので，それらを対照した対照テーブルを作って，図6に示すように対応する分類番号1を生成する．次に，この分類番号1と分類語彙表より検索された分類番号2とを比較して類似性の高い分類番号3だけを抽出して韓国語シソーラスを作成する．この比較では多くの場合分類番号全体で一致がとれず，上位3桁以上の一致が得られた分類番号だけを取り出した．

3.2.2 意味属性の分類番号への変換

意味属性を分類番号に変換するためには，両分類項目の対照テーブルが必要になる．しかし，韓国語意味辞書の意味属性と分類語彙表の分類番号とは分類体系が違っているし，おのおのの分類の数が多いので，人手でそれらの対応づけをするのは大変であり，しかも間違う可能性が非常に高い．そこで，意味属性と分類番号の対照テーブルを半自動的に作成することにした．このために，ある文字列を入力すると分類番号の一致度に基づいた類似度を計算して入力文字列と似ている例文の文字列が検索されて出てくるECTM(用例検索による韓日・日韓翻訳支援システム)[4]を用いた．

韓国語意味辞書の意味属性一覧と分類語彙表の分類項目一覧には，おのおの分類の基準となる見出し語が日本語で**表3**と**表4**のように書かれている．分類語彙表の分類項目一覧を例文ファイルとして用意した後，**図7**のように韓国語意味属性の分類のための基準語(以下意味基準語)を入力キーとし，分類語彙表の分類のための基準語(以下分類基準語)と分類番号の一致度に基づいて類似度を計算して，最も似ている分類基準語を取り出した．この仕組みは図7のようである．こうして検索された分類基準語を人手で検討して意味的に一番近い分類基準語の分類番号だけを選んで，**表5**のような意味属性と分類番号との対照テーブルを作った．

例えば，**図8**に示すように意味基準語の「方向」を入力すると，それと意味的に近い単語が類似点数の大きい順にその点数とともに「方向」，「上下」，「方面」，「左右」などの分類基準語とその分類番号が検索されて出てくるので，効率的に意味属性に対応する分類番号を探すことができる．類似点数の計算の仕方は3.2.3項で述べるが，分類語彙表の分類番号の一致度を用いて求めた．

3.2.3 一致度による分類番号の選択

3.2.1項で説明したように分類語彙表より検索された分類番号2の中で対照テーブルを用いて意味辞書よ

表3 韓国語意味辞書の意味属性一覧
Table 3 A list of some semantic attributes.

意味属性	意味基準語
N. ABI	能力，性向
N. ARCH	建築物，施設
N. ARR	方向
N. CON	具体物
N. CRE	生物
N. HUM	人間
V. ASPHE	社会現象，発生
V. DIF	無意志，抽象的状態，優劣
V. DSPHE	無意志，社会現象，衰退
V. EXI	無意志，存在状態

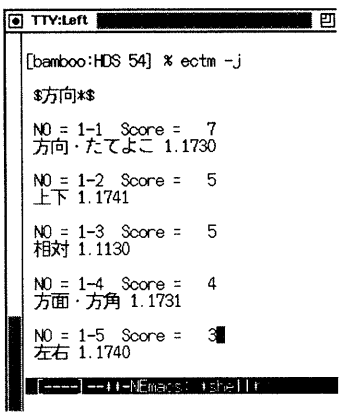


図8 ECTMの検索結果の例
Fig. 8 An example of ECTM's retrieval results.

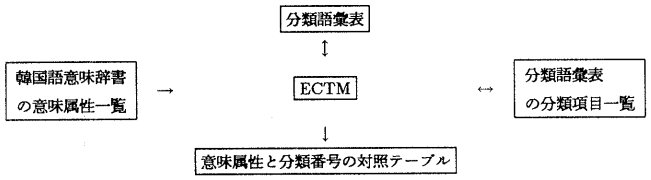


図7 対照テーブルの作成
Fig. 7 Construction of conversion table.

表4 分類語彙表の分類項目一覧
Table 4 A list of some of the classification numbers.

分類番号	分類基準語	分類番号	分類基準語
⋮	⋮	1.4323	さかな・鰹節・肉
1.1330	性質	⋮	⋮
⋮	⋮	1.4720	その他の土木施設
1.1404	能力	⋮	⋮
⋮	⋮	1.5500	生物
1.1730	方向・たてよこ	1.5510	植物
1.1731	方向・方角	⋮	⋮
1.1740	左右	1.5620	鳥
1.1741	上下	⋮	⋮
⋮	⋮	2.1200	存在
1.2000	われ・なれ・かれ・だれ	2.1210	出現
1.2010	自他	2.1220	成立・発生
1.2020	人間	2.1230	仕上げ
1.2040	男女	2.1240	残存・消滅
1.2050	老少	⋮	⋮
⋮	⋮	2.1526	進退
1.2340	人物	⋮	⋮
⋮	⋮	2.1583	強め・衰えなど
1.3823	建築	⋮	⋮
⋮	⋮	2.1900	過不足・優劣など
1.4000	人間活動の生産物	⋮	⋮
⋮	⋮	⋮	⋮

表5 意味属性と分類番号の対照テーブル
Table 5 A part of the conversion table.

韓国語意味辞書		分類語彙表
意味基準語	意味属性	分類番号（分類基準語）
能力	N. ABI	1.1404(能力) 1.1330(性質)
建築物，施設	N. ARCH	1.3823(建築) 1.4720(その他の土木施設)
方向	N. ARR	1.1730(方向・たてよこ) 1.1741(上下)
		1.1731(方向・方角) 1.1740(左右)
具体物	N. CON	1.4000(人間活動の生産物) 1.5000(自然)
生物	N. CRE	1.5500(生物) 1.2040(男女) 1.5620(鳥)
		1.4323(さかな・鰹節・肉) 1.5510(植物)
人間	N. HUM	1.2020(人間) 1.2000(人間活動の主体) 1.2340(人物)
		1.2050(老少) 1.2040(男女) 1.2000(われ・なれ・かれ・だれ)
社会現象，発生	V. ASPHE	2.1210(出現) 2.1220(成立・発生)
無意志，抽象的状態，優劣	V. DIF	2.1900(過不足・優劣など)
無意志，社会現象，衰退	V. DSPHE	2.1583(強め・衰えなど) 2.1526(進退)
無意志，存在状態	V. EXI	2.1200(存在) 2.1240(残存・消滅)

り生成された分類番号1と一致する番号だけを選ぶ必要がある．だが，両番号には完全に一致するものもあれば，部分的にしか一致しないものもある．特に，部分一致するものは複数個存在することが多い．ところで図5をみると，レベル3まで一致すれば，相当に意味的に近いことがわかる．そこで，少なくともレベル3まで一致する分類番号はすべて選ぶことにした．すなわち分類番号の上位から少なくとも3桁以上一致するものを選ぶのである．

例えば，**図9**のように「나(私)」の場合には，意味辞書より「NCON」，「NCRE」，「NHUM」などの意味属性が検索され，対照テーブルによる分類番号への変換に

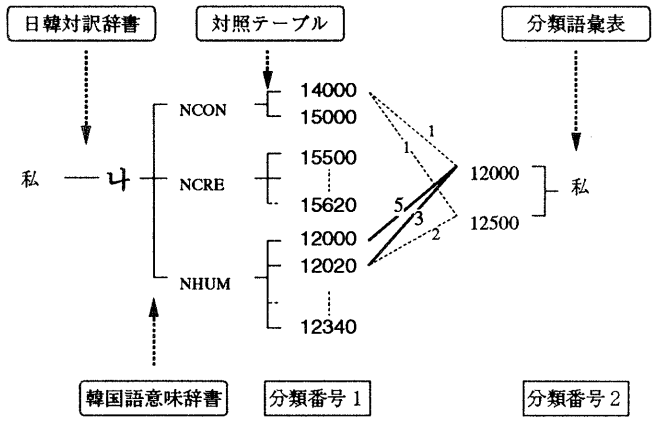


図 9 一致度による分類番号の選択
Fig. 9 Thesaurus number selection by matching degree.

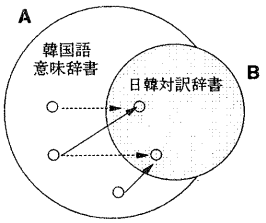


図 10 類推による分類番号の付与の概念
Fig. 10 Concept of thesaurus numbering by analogy.

よって「14000(人間活動の生産物)」,「15500(生物)」,「12020(人間)」などの分類番号 1 が生成される．一方，分類語彙表からは「12000(われ)」,「12500(公私)」などの分類番号 2 が検索されて出てくる．これらの番号を比較して一致度が 5 である「12000」だけが選ばれることになる．このとき，一致度は次のような数式(1)を作って求めている．

$$Md=k|BN1[1,\ldots,k]\equiv BN2[1,\ldots,k]\wedge BN1[k+1]\neq BN2[k+1]\} \quad (1)$$

Md：一致度，$BN1$：分類番号 1，$BN2$：分類番号 2

3.3 類推による分類番号の生成(方法 3)

上記の二つの方法で 12,000 個の分類番号を韓国語の単語に付与することができた．しかし，それは日韓対訳辞書の対訳語だけで，対訳語として登録されていない単語については分類番号を付与することができない．そこで日韓対訳辞書に載っていないけれども，意味辞書には載っている単語については以下のように意味属性を類推することによって分類番号を付与する方法を用いた．この処理概念を**図 10** に示す．

意味辞書の中には上記の処理で分類番号の付与された単語(図 10 の B の部分：意味辞書 1 と呼ぶ)と，分類番号の付与されていない単語(図 10 の A の実線部分：意味辞書 2 と呼ぶ)がある．ところが，両方とも概念記号，意味属性，格情報および機能属性，隣接情報をもっている．そこで，これらの構文および意味情報の一致度を用いて B 部分の単語から A 部分に意味的に似ている単語を探して，方法 1 と方法 2 によってあらかじめ構築されている韓国語シソーラスからその単語の分類番号を取り出し，A 部分の単語に写す．このようにして，日韓対訳辞書に載っていない単語にも分類番号を与えることができる．

しかし，このときにも多義語の場合があるので，検索された分類番号を 3.2.3 項に述べた方法で意味的に一致する番号だけを選ぶ．図 10 で実線は意味的に強い一致をすることを，点線は意味的に弱い一致しかしないことを表す．この場合，意味的一致度は経験的に次のような数式(2)を作って求めた．

$$Sd=w_1\times M_{cw}+w_2\times M_{sa}+w_3\times M_{ci}+w_4\times M_{ai} \quad (2)$$

$$\begin{cases}\text{if } I_{cw}=R_{cw} \text{ then } M_{cw}=1 \text{ else } M_{cw}=0\\ \text{if } I_{sa}=R_{sa} \text{ then } M_{sa}=1 \text{ else } M_{sa}=0\\ \text{if } I_{ci}=R_{ci} \text{ then } M_{ci}=1 \text{ else } M_{ci}=0\\ \text{If } I_{ai}=R_{ai} \text{ then } M_{ai}=1 \text{ else } M_{ai}=0\end{cases}$$

Sd：一致度の合計，M_n：属性別一致度，w_n：加重値，I：A 部分の単語，R：B 部分の単語，cw：概念記号，sa：意味属性，

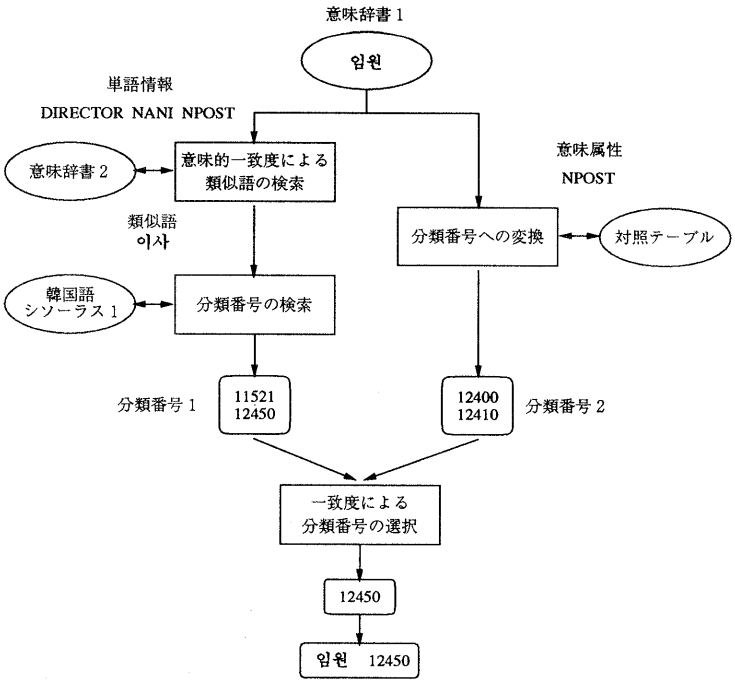


図 11 類推による分類番号の付与の処理過程
Fig. 11 Process of thesaurus numbering by analogy.

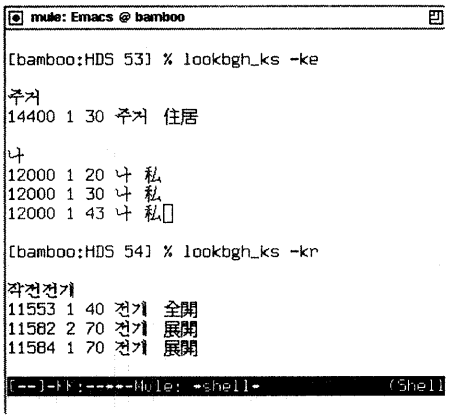


図 12 検索結果の例
Fig. 12 Some examples of retrievals.

ci：格情報・構文属性，*ai*：隣接情報

意味辞書 1 の「임원(役員)」という単語を例にして説明すると，

- まず，意味的一致度を用いて意味辞書 2 より「임원(役員)」の単語情報* と似ている類似語を探し出す．この場合，「이사」という単語が検索される．
- 次に，この「이사」の分類番号を韓国語シソーラス 1 から引き出す．この場合，「이사」には「引越し」と「理事」との二つの意味が含まれているので，「11521」と「12450」などの複数の分類番号 1 が出てくる．そして，この中から意味的に「임원(役員)」と一致する分類番号を取り出すために，
- 対照テーブルを用いて「임원(役員)」の意味属性を「12400」「12410」などの分類番号 2 に変換する．
- 最後に，分類番号 1 と分類番号 2 との一致度を求めて，一致度が 3 である「12450」(臨時的地位)だけを取り出して，「임원(役員)」に付与する．

この処理過程を**図 11** に示す．

表 6 作成方法別の評価結果
Table 6 Evaluation results for the three methods.

ランク	方法 1	方法 2	方法 3	合計
A	11,261(98 %)	455(98 %)	242(92 %)	11,958(98 %)
E	236(2 %)	10(2 %)	21(8 %)	267(2 %)
合計	11,497	465	263	12,225

* 概念記号「director」，意味属性「NANI」「NPOST」…，機能属性「NOUN」，隣接情報が単語情報として書かれている．

表7 品詞別の評価結果
Table 7 Evaluation results according to part of speech.

ランク	体の類	用の類	相の類	その他	合計
A	10,170(99%)	796(92%)	959(92%)	33(87%)	11,958(98%)
E	142(1%)	73(8%)	47(8%)	5(13%)	267(2%)
合計	10,312	869	1,006	38	12,225

表8 分類番号付与失敗の原因別単語数
Table 8 Cause errors and numbers of words given incorrect classification number.

失敗原因	方法1	方法2	方法3	合計
辞書情報記述の誤り	111	3	13	127(48 %)
多義語処理の不処理	125	7	8	140(52 %)
合計	236	10	21	267(100 %)

表9 別の日韓対訳辞書からの韓国語シソーラスの作成
Table 9 Evaluation results using another Japanese to Korean translation dictionary.

方法1	方法2	方法3	合計
31,280	1,963	96	33,339

4. 韓国語シソーラスの検索および評価

4.1 韓国語シソーラスの検索

以上の三つの方法を用いて韓国語シソーラスを作成することができたので，韓国語の分類番号を検索するシステムを作成した．検索方法としては完全一致法と末尾最長一致法* を用いることができる．検索例を**図12**に示す．

4.2 評　　価

上記の三つの方法によって9,274語の韓国語単語に対しておのおの11,497個，465個，263個の分類番号を与えることができた．ここで作成された結果に対しては一つ一つの分類番号の正確性を検討しなければ，シソーラスとしての良さを判断することが難しいということを考慮して，すべての単語に対して人手によってその妥当性を調べた．

評価のランクは次のように二つに分けた．

A：分類番号がINとTNで一致していて，それが適切である．

E：与えている分類番号(IN・TN)が違っている．

* 文字列の末尾から韓国語シソーラスの見出し語と照合をとり，複数個の見出し語と照合がとれた場合に，最も長い見出し語を優先して選択する方法．

評価の結果，韓国語の単語に与えられた分類番号の中で約98%が正しいことがわかった．**表6**は第3章に述べた方法1から方法3までの作成方法別に分けて評価した結果で，**表7**は分類番号を品詞別に分けて評価した結果である．分類番号の付与に失敗した原因を分析した結果，方法1と方法2では日韓対訳辞書の対訳語が間違って記述されて分類番号の付与に失敗した単語がおのおの111語，3語であって，方法3では韓国語意味辞書の概念記号が間違って記述されて分類番号の付与に失敗した単語が13語であった．つまり，約127語が辞書情報の記述の誤りによって分類番号の付与に失敗した．残りの267語は多義性をもつ日本語として本研究の方法では処理ができなかった部分である．失敗原因の分析結果を**表8**に示した．

以上の評価の結果，この方法はかなりの正確さをもつことがわかったので，大規模な単語辞書を用いて，本格的な韓国語シソーラスを作成することを試みた．約61,183語の見出し語に対して約67,187語の対訳語が登録されている別の日韓対訳辞書に対して上記の三つの方法を適用して，韓国語単語約2万3千語に対して**表9**のように約3万3千個の分類番号を与えることができた．こうしてある程度大規模な韓国語シソーラスを作ることができた．しかし，このシソーラスについてはまだ質の評価を行っていない．

5. おわりに

本稿では，既存の日韓機械辞書と分類語彙表を利用して，韓国語シソーラスを半自動的に作成する方法を示した．特に，韓日語間の対照関係，また日韓機械辞書と分類語彙表の対応に基づいて行った単純検索による方法，意味属性の分類番号への変換による方法，意味属性の類推による方法を用いて非常に良質の結果が得られることがわかった．付加的に本研究での評価を通して，韓国語と日本語が単語のレベルにおいても非常に似ていることが確められた．

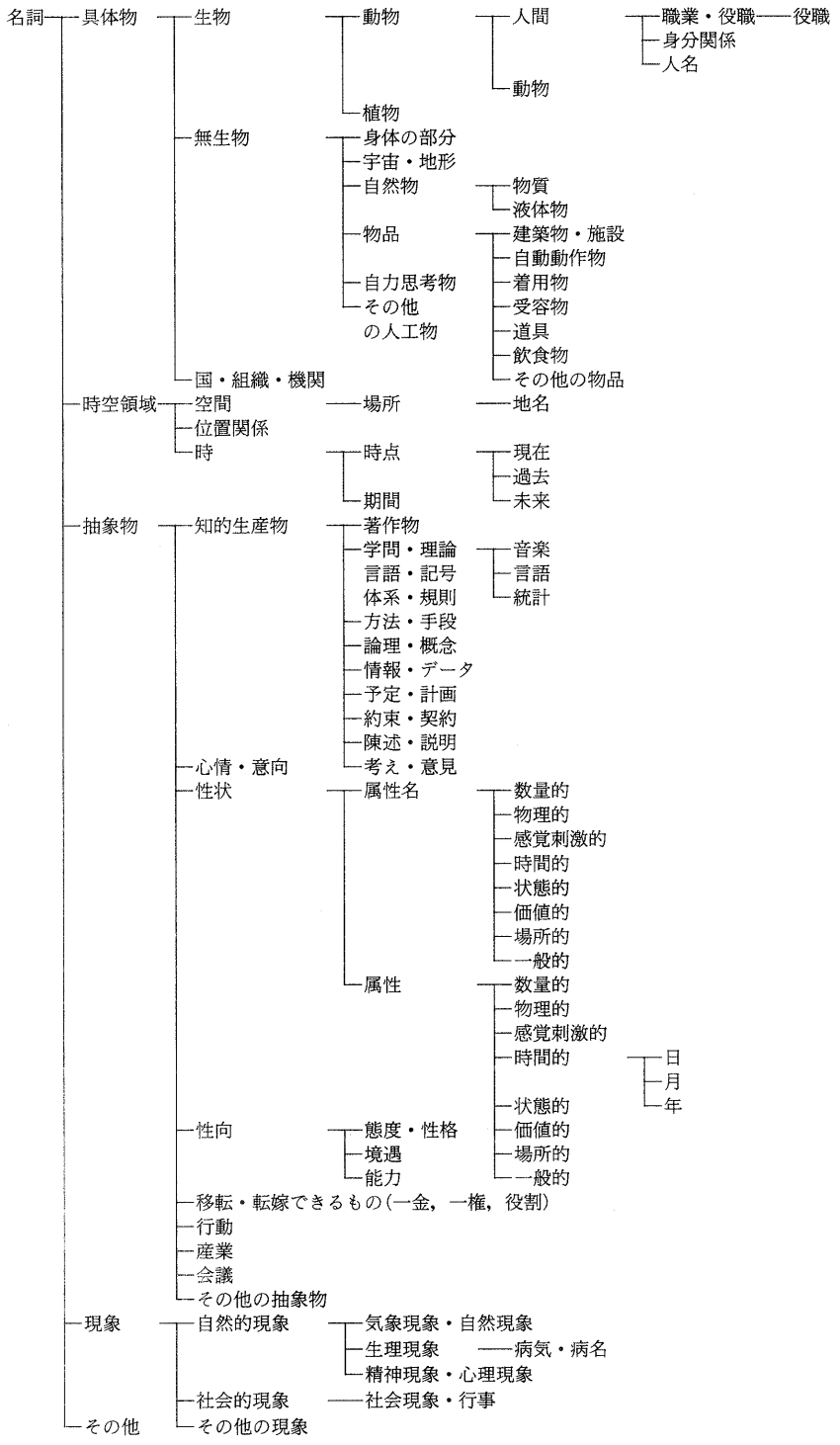
しかし，一般の辞書ではなくて，機械翻訳用の辞書を利用したので，見出し語が単語ではなくて，句また

は節であるものが相当数あり，うまく変換された単語数が少ないという問題点は残っている．今後さらに，分類番号の正確さを低下させずに，単語の数を増やす方法についての研究を行う必要がある．そのほかに，分類語彙表の分類体系に基づいて韓国語に分類番号を与えたので，もし韓国語だけに存在する固有な分類項目があれば，分類番号が付与されていないこともありうる．ところが，分類語彙表の現在の分類項目は韓国語にも適用できるので，本研究で韓国語に付与した分類番号は有効であると考えられる．最後に，本研究で使われた方法を他言語の間，例えば英語と日本語にも活用できるのだろうということが考えられる．この場合，方法2と方法3を用いると，付与される分類番号は少なくなるが，良い結果が得られると予想している．また，ここで作成した単語を元にして，人手で単語を追加してゆくことは比較的に容易であると思われる．

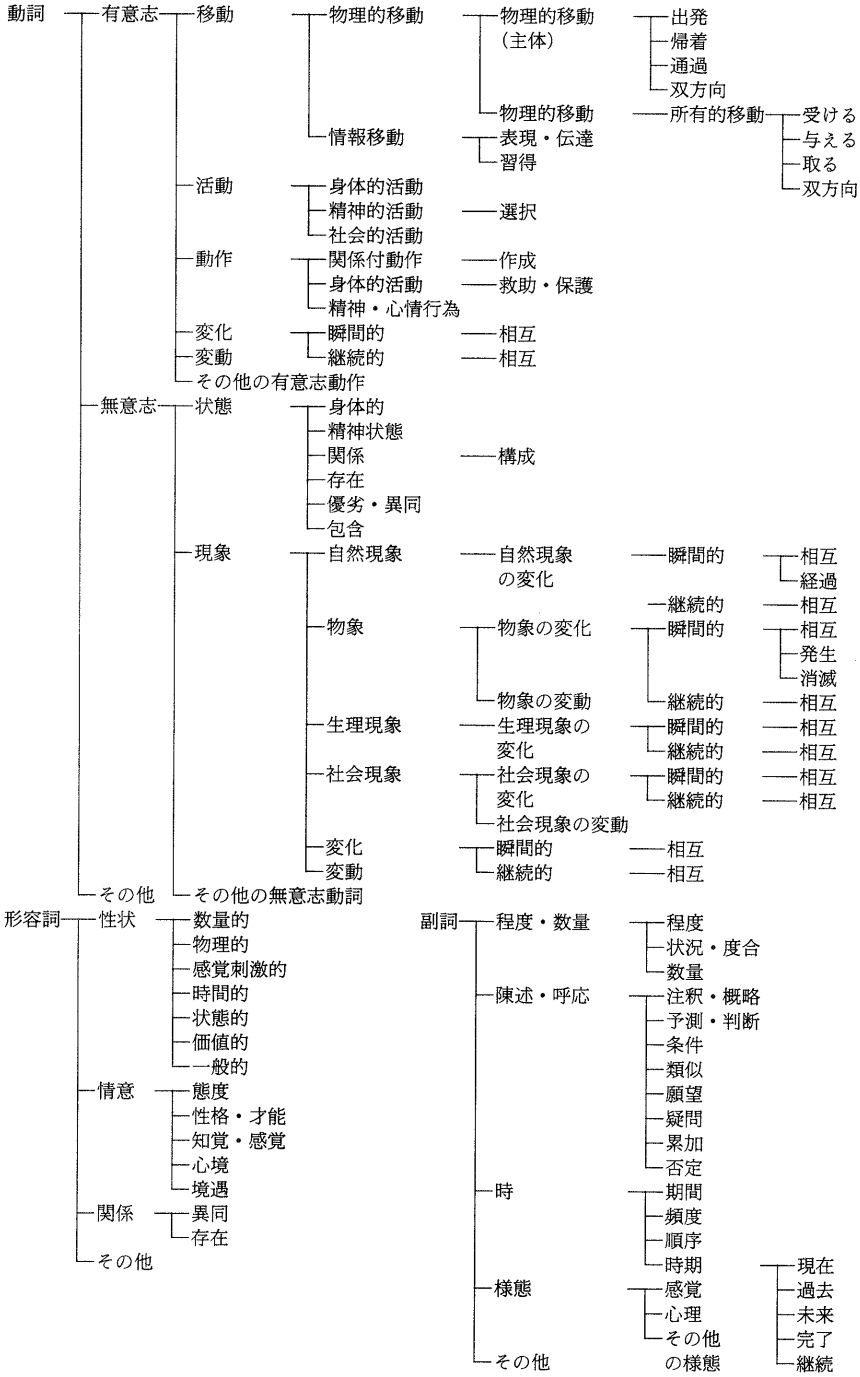
参考文献

1) 国立国語研究所：分類語彙表，p. 362，秀英出版(1964)．
2) 黄　道三，張　遠ほか：第4回日韓機械翻訳共同研究報告書，p. 34，韓国科学技術研究院システム工学研究所＋富士通株式会社(1986)．
3) 黄　道三，李　尚起，張　遠：韓日機械翻訳システム開発に関する研究，韓国富士通(株)の委託研究報告書，p. 230，韓国科学技術研究院システム工学研究所(1991)．
4) 黄　道三，長尾　真，佐藤理史：用例検索による韓日・日韓翻訳支援システム，情報処理学会自然言語処理研究会報告，Vol. 93, No. 61, pp. 49-56 (1993)．
5) Park, D., Hwang, D., Lee, S., Chang W. et al.: A Study on the Development of the Korean to Japanese Machine Translation System, *Proc. of PRICAI'90,* Vol. 4, pp. 221-226(1990).
6) 林　大：分類語彙表について，言語生活，No. 394, pp. 70-76(1984).
7) 長尾　真：言語工学，p. 246，昭晃堂(1983)．
8) 長尾　真ほか：岩波情報科学辞典，p. 1172, 岩波書店(1990)．

付録Ⅰ　名詞の意味分類


名詞
具体物
生物
動物
人間
職業・役職
役職
身分関係
人名
動物
植物
無生物
身体の部分
宇宙・地形
自然物
物質
液体物
物品
建築物・施設
自動動作物
着用物
受容物
道具
飲食物
その他の物品
自力思考物
その他の人工物
国・組織・機関
時空領域
空間
場所
地名
位置関係
時
時点
現在
過去
期間
未来
抽象物
知的生産物
著作物
学問・理論
言語・記号
体系・規則
音楽
言語
統計
方法・手段
論理・概念
情報・データ
予定・計画
約束・契約
陳述・説明
心情・意向
考え・意見
性状
属性名
数量的
物理的
感覚刺激的
時間的
状態的
価値的
場所的
一般的
属性
数量的
物理的
感覚刺激的
時間的
日
月
年
状態的
性向
態度・性格
境遇
能力
価値的
場所的
一般的
移転・転嫁できるもの（一金，一権，役割）
行動
産業
会議
その他の抽象物
現象
自然的現象
気象現象・自然現象
生理現象
病気・病名
精神現象・心理現象
社会的現象
社会現象・行事
その他
その他の現象

付録II　動詞・形容詞・副詞の意味分類



（平成5年6月14日受付）
（平成5年11月11日採録）

黄　道三（正会員）

1980年韓国弘益大学工学部電子計算学科卒業．1983年年韓国延世大学大作院修士課程修了．同年韓国科学技術研究院入所．現在，京都大学大学研究科博士後期課程在学中．自然言語処理，知識情報処理，機械翻訳の研究に従事．

長尾　眞（正会員）

1959年京都大学工学部電子工学科卒業．1961年同大学院修士課程修了．京都大学工学部助手，助教授を経て，1973年より同教授，現在に至る．1976年より国立民族学博物館併任教授．パターン認識，画像処理，自然言語処理，機械翻訳等の研究に従事．