HOKUGA 北海学園学術情報リポジトリ
学校法人北海学園
北海学園大学
北海商科大学

タイトル	層指定検索ツールの開発
著者	安曇，恭徳；桃内，佳雄
引用	工学研究 ： 北海学園大学大学院工学研究科紀要，8: 53-62
発行日	2008-09-28

Hokkai-Gakuen Organization of Knowledge Ubiquitous through Gaining Archives.

研究論文

層指定検索ツールの開発

安 曇 恭 徳* ・ 桃 内 佳 雄*

A Tool for Retrieving in Specified Layers

Yasunori Azumi* and Yoshio Momouchi*

1．はじめに

機械翻訳システムの開発においては，異なる言語間の対照的な言語学的考察や翻訳ルールの解析と構成が必須であり，そのために，対訳コーパスと呼ばれる電子化された対訳テキストデータを構築して利用することが行われている．

アイヌ語と日本語の機械翻訳システムの開発や言語学的考察においても，アイヌ語と日本語の翻訳対を基礎とするアイヌ語・日本語対訳データを利用することは有用である．しかし，対訳データを用いて特定の言語情報に関する考察や解析を行う際に，人手では，その言語情報が含まれている対訳データ要素を探すために時間がかかる．そのため，効率的に情報検索を行うツールを開発することは重要である．

本報告では，アイヌ語・日本語機械翻訳システムの開発を支援するツールとして開発した，アイヌ語・日本語対訳データ要素の階層的な構成を考慮に入れた情報検索ツール「層指定検索ツール」の構成と，「層指定検索ツール」を利用した解析例について報告する．

また，層指定検索ツールの開発にUMLを用いてモデリングを行った．これにより，システムの構造や処理の流れをわかりやすく表現することが可能となり，システムの修正や機能の追加が容易になると考えられる．本報告では，層指定検索ツールにおけるUMLの利用についても述べる．

2．対訳データ要素の構成と層指定検索ツール

層指定検索ツールは，一定の階層の対訳データ要素の集合で構成されている対訳データを対象として，検索の条件や検索する層を指定して検索を行い，検索の条件に合致した対訳データ要素を表示，出力するツールである．

2.1　対訳データ要素の構成

アイヌ語・日本語対訳データ要素の基本的な構成について，まず，文献[1,2]に対応する対訳データを構成する中で検討を進めた．その結果，アイヌ語・日本語対訳データ要素の基本的な構成は，付加コードを付与した，アイヌ語文，アイヌ語品詞列，日本語逐語訳，日本語品詞列，日本語文の5層の情報から構成される単位を基本構成要素として，次のように設定することとした．

〈5層の構成〉
付加コード01：アイヌ語文
付加コード02：アイヌ語品詞列
付加コード03：日本語逐語訳
付加コード04：日本語品詞列
付加コード05：日本語文（自然な）

各層のデータの〝:〞より右側の部分を「文」と呼ぶことにする．また，付加コードは次のようになっており，テキストの類別，章や節などの番号，文の番号，層に対応している．

exp 01001xx：(文)
exp　　テキストの類別
　01　　章，節

* 北海学園大学大学院工学研究科電子情報工学専攻
Graduate School of Engineering (Electronics and Information Eng.), Hokkai-Gakuen University

001　文番号
xx　層
※(exp：エクスプレス　アイヌ語[1])

文献[1,2] に対応する対訳データ要素の構成では，次のような 6 層の情報による構成が設定される．

〈6 層の構成〉
付加コード 01：アイヌ語文
付加コード 02：アイヌ語品詞列
付加コード 03：日本語逐語訳
付加コード 04：日本語品詞列
付加コード 05：日本語文（自然な）
付加コード 06：日本語文（原著の）

これは，基本構成に「原著の日本語文」を追加したものである．対訳データ要素の例を次に示す．

〈6 層の構成の例[1]〉
exp 0100521：sonno　keraan　.
exp 0100502：副詞　自動詞　.
exp 0100503：本当に　おいしい　。
exp 0100504：副詞　形容詞　。
exp 0100505：本当においしい　。
exp 0100506：とってもおいしいよ　。

次に，文献[3] に対応する対訳データ要素の構成では，次のような 7 層の情報による構成が設定される．

〈7 層の構成〉
付加コード 01：アイヌ語文（原著）
付加コード 02：アイヌ語文（切替版）
付加コード 03：アイヌ語品詞列
付加コード 04：日本語逐語訳
付加コード 05：日本語品詞列
付加コード 06：日本語文（自然な）
付加コード 07：日本語文（原著の）

6 層から新たに加わったのは，第 2 層の「アイヌ語文（切替版）」である．知里幸恵による原著アイヌ語文は，区切り単位が必ずしも品詞単位ではなく，切替版は，これを品詞単位に区切ったものである．対訳データ要素の例を次に示す．

〈7 層の構成の例[3]〉
yuk 0100101：〃 Shirokanipe ranran pishkan , konkanipe
yuk 0100102：〃 sirokani pe ran ran piskan , konkani pe
yuk 0100103：〃 名詞 名詞 自動詞 自動詞 位置名詞 , 名詞 名詞
yuk 0100104：「 銀 滴 降る 降る のまわり , 金 滴
yuk 0100105：「 名詞 名詞 自動詞 自動詞 名詞 , 名詞 名詞
yuk 0100106：「銀の滴降る降るまわり，金の滴
yuk 0100107：「銀の滴降る降るまはりに，金の滴

2.2 層指定検索

層指定検索は，N層（N＞0）の階層で構成されている対訳データ要素中の階層M番目(M≦N)を指定し，その階層でキーワード検索を行い，キーワードが含まれていれば，そのデータ要素を出力するという検索である．

また，層指定検索を行う条件として，検索する対訳データ要素の階層はすべて同じ階層で構成されていなければならない．

2.3 層指定検索ツールの処理概要

層指定検索ツールの処理は，(1)対訳データファイルからデータを読み込む，(2)分かち書き文を単語に分割する，(3)層指定検索を行う，(4)表示・出力の判定をする，(5)文中のキーワードに印をつける，(6)表示・出力をする，これら 6 つの段階に大きく分けられる．

層指定検索ツールでは，対訳データファイルからの『読み込み』は一行ずつとなる．そのため，一行読み込むごとに『単語分割』の処理を行い，検索層に到達したところで『検索』を行う．そして，1 つのデータ要素分のデータを読み込んだところで，『判定，印付け，出力』の処理を行う．

以上のことを踏まえて，層指定検索ツールの処理概要をまとめると次のようになる．

①入出力ファイル名や検索キーワード，検索条件，出力条件を入力する．
②入力ファイルから，データを一行読み込む．
③分かち書き文を単語に分割する．
（分かち書き文については後に説明する．）
④検索する層に到達したところで，設定した検索条件に合わせてキーワード検索を行い，その結果を返す．
⑤1 つのデータ要素のデータをすべて読み込むまで，②～④の処理を繰り返す．
⑥検索した結果を出力条件に照らし合わせて判定を行う．
⑦判定結果が可であり，かつ，〝(キーワードに)

印をつける〞場合には，キーワードの前後に印をつける．

⑧判定結果が可であれば出力する．

⑨入力ファイル終端まで②～⑧の処理を行う．

2.4　層指定検索ツールの検索機能

層指定検索ツールの検索機能として，検索条件と出力条件を設定できるようになっている．検索条件は 3 つ存在し，検索の種類とその意味は以下の通りである．

- 完全一致検索：検索キーワードと完全に一致する語が存在するデータ要素を検索する．
- 部分一致検索：検索キーワードを部分的に含んでいる語が存在するデータ要素を検索する．
- 否定検索：検索キーワードが存在しないデータ要素を検索する．

また，一度に 2 つの階層を検索する 2 層検索が可能となっており，それぞれを階層Ａにおける検索（Ａ検索）と階層Ｂにおける検索（Ｂ検索）と呼び，以下のような 2 層検索における出力条件の設定が可能である．

AND　：2 つの検索の両方が成り立つデータ要素を出力する．

OR　：2 つの検索のどちらか一方が成り立つデータ要素を出力する．

同一位置　：AND 検索において，2 つの検索が同一の位置において成り立つ場合にデータ要素を出力する．

これらの検索条件を設定することにより，様々な場合における検索を行うことが可能となる．

3．層指定検索ツールの UML モデリング

UML とは「Unified Modeling Language（統一モデリング言語）」の略で，オブジェクト指向におけるシステムの分析や設計，実装を円滑に進めるために作成する〝モデル〞を記述するための表記法である[9]．現在，UML には，13 種類の表記法が存在し，使用する場面によって使い分けを行う．層指定検索ツールの開発には，クラス間の関連を示すクラス図と処理の流れを示すシーケンス図を利用して，モデリングを行った．本章では，クラス図についてのみ説明する．図 1 のクラス図には，

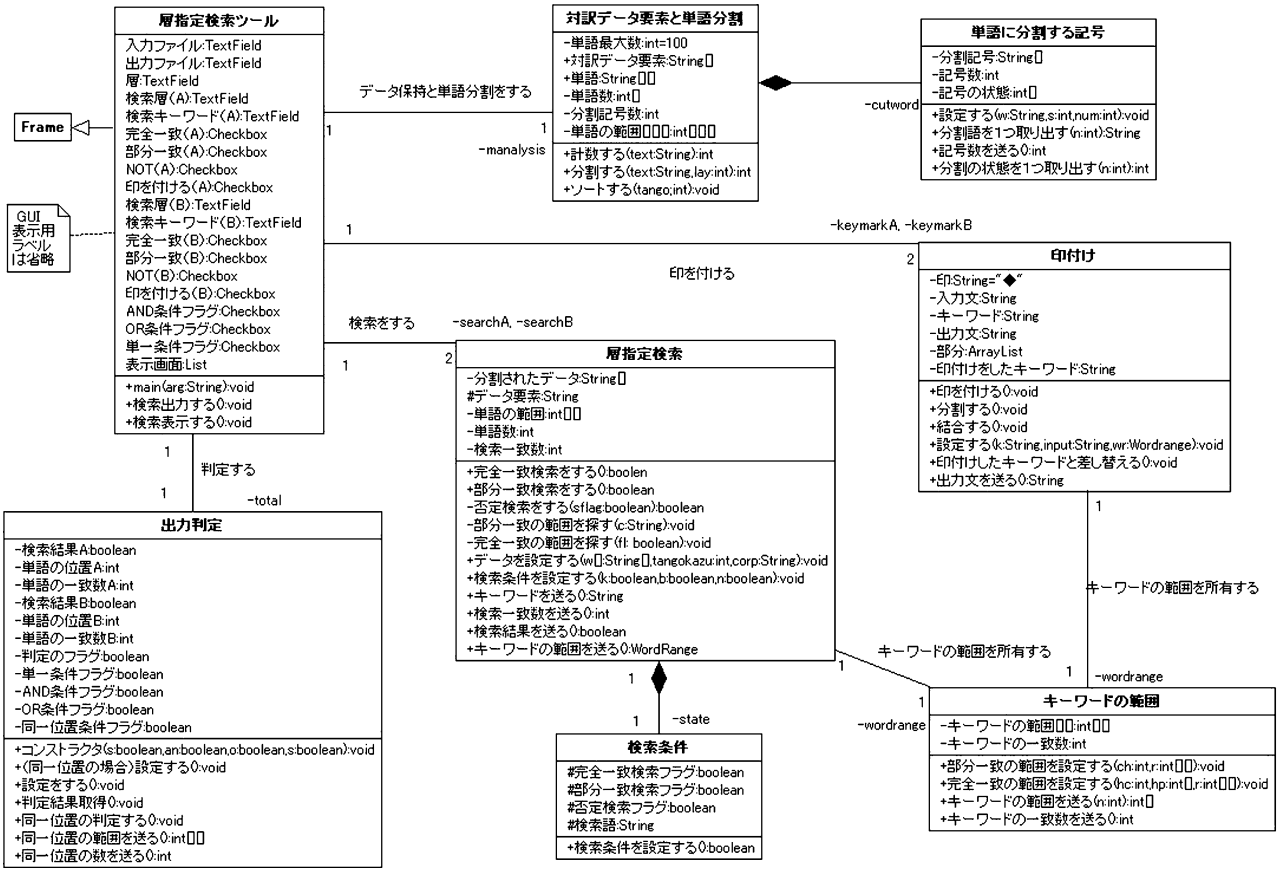

図 1　層指定検索ツールのクラス図

８つのクラスとそれぞれの関連が記述されている．関連については，関連名と多重度，インスタンス名が記述されている．なお，インスタンスの先頭部やクラスの中にある記号（＋，－，＃）は可視性を示す．

層指定検索ツールにおけるクラスとクラスの役割，クラス間の関連は次のようになっている．

・層指定検索ツール
　層指定検索ツールのメインメソッドが記述されている．そのほかにも，GUI や各クラスのインスタンスが記述されており，層指定検索ツールの本体を示している．図 1 では，GUI を記述する API の『Frame』やその他のクラスとの関連が示されている．なお，GUI の構成要素は省略する．
・対訳データ要素と単語分割
　読み込んだデータ要素を保持し，データ要素を単語単位に分割するためのクラスである．
・単語に分割する記号
　単語を分割するための記号データを格納するためのクラスである．図 1 では，『対訳データ要素と単語分割』のクラスとコンポジションの関連を持ち，そのクラスの部分であることを意味する．なお，コンポジションは黒く塗りつぶしたひし形で表現する．
・層指定検索
　層指定検索を行うためのクラスである．このクラスでは，検索処理を行うときに，印付けの処理のために，文中に存在するキーワードの位置データを作成する．
・検索条件
　層指定検索を実行するための検索条件を格納するためのクラスで，層指定検索のクラスとは，コンポジションの関連をもつ．
・印付け
　対訳データ要素にあるキーワードの前後に印を付けるためのクラスである．
・キーワードの範囲
　印付け処理を行うために必要なデータを保有するクラスである．
・出力判定
　検索処理と出力条件に合わせて，出力判定を行うためのクラスである．

４．層指定検索ツールの処理

層指定検索ツールの処理は，2.3 で 6 つの段階に分けられると述べた．本章では，これらの処理の詳細について述べる．

⑴　ファイルからデータを読み込む

ユーザによって指定された対訳データファイルからデータを読み込む処理を行う．ファイルの先頭から 1 行ずつ読み込み，格納する．

⑵　分かち書き文を単語に分割する

分かち書き文を単語の集合に変換する．分かち書き文とは，次の例のように単語と単語の間にスペースが入った文を指す．

〈例〉
単語 1　単語 2　単語 3　単語 4　．
ahup　wa　sini　yan　．（入って休みなさい。）

アイヌ語には，人称接辞というものがあり，人称接辞を示すための補助的な記号として〝＝〞を用い，人称接辞を含んだ文は次の例のようになる．

〈例〉
単語 1　単語 2 ＝単語 3　単語 4　．
usey　ku＝ku　rusuy　．（お湯を飲みたい。）
　※（単語 2 に対応する〝ku〞が人称接辞である．）

層指定検索ツールにおいては，人称接辞を示すための〝＝〞は，検索を可能とするために単語の 1 つとして登録できるようになっている．

また，次のような文は分かち書き文ではないので単語を分割することはできない．

〈例〉
単語 1 単語 2 単語 3 単語 4 。
お湯を飲みたい。

分かち書き文の単語分割をするために，層指定検索ツールでは，単語に分割する前に単語ごとに区切る記号をプログラム中で登録し，読み込んだ文の中から登録された記号を探し，その位置に合わせて単語に区切るようにしている．その例を以下に示す．

〈例〉
〝exp 0100611：usey　e＝ku　rusuy　ya　？〞
区切り記号＝　{‘　’, ‘：’, ‘＝’}
　※（‘＝’ は単語として残す．）
文字＝{［０］：‘ｅ’, ［１］：‘ｘ’, ・・・, ［10］：‘：’, •••, [15]：‘　’, •••, [20]：‘　’, •••, [26]：‘　’, ・・・, [29]：‘　’, ［30］：‘？’}

記号の位置（数値）＝{10，15，17，20，26，29}
単語の範囲＝{（0 ～ 9），（11～14），（16～16），
（17～17），（18～19），（21～25），
（27～28），（30～30）}
※（（17～17）は‘＝’の部分.）
結果＝{〝exp 0100611〞，〝usey〞，〝e〞，〝＝〞，〝ku〞，
〝rusuy〞，〝ya〞，〝？〞}

⑶　層指定検索を行う

この処理では，指定した層において設定した検索条件に従って，入力したキーワードを探す処理を行う．ここで設定可能な検索条件は，完全一致検索，部分一致検索，否定検索の３つであり，それぞれの場合において以下の処理をする．

(i) 完全一致検索

データ要素の指定した層の分かち書き文の単語の中にキーワードと一致する単語があるかどうか調べる．ある場合には検索結果を True にし，そうでなければ False とする．また，一致する場合にその単語の位置をキーワードの出現位置として記録する．

(ii) 部分一致検索

データ要素の指定した層の文がキーワードを含むかどうか調べる．含んでいる場合には検索結果を True にし，そうでなければ False とする．

(iii) 否定検索

完全一致検索または，部分一致検索を行い，その結果を True から False，False から True へ反転させる．

次にそれぞれの検索の例を示す．

〈例〉
〝exp 0100611：usey　e＝ku　rusuy　ya　？〞
単語＝{〝exp 0100611〞，〝usey〞，〝e〞，〝＝〞，〝ku〞，
〝rusuy〞，〝ya〞，〝？〞}
キーワード＝〝e〞
(i) 完全一致検索で検索した場合
結果＝True
単語の位置＝３
(ii) 部分一致検索で検索した場合
結果＝True
(iii) 否定検索で検索した場合
結果＝False（完全一致検索の場合）
結果＝False（部分一致検索の場合）

〈例〉
〝exp 0100611：usey　e＝ku　rusuy　ya　？〞
単語＝{〝exp0100611〞，〝usey〞，〝e〞，〝＝〞，〝ku〞，
〝rusuy〞，〝ya〞，〝？〞}
キーワード＝〝suy〞
(i) 完全一致検索で検索した場合
結果＝False
(ii) 部分一致検索で検索した場合
結果＝True
(iii) 否定検索で検索した場合
結果＝True（完全一致検索の場合）
結果＝False（部分一致検索の場合）

⑷　表示・出力の判定をする

設定した出力条件と層指定検索の結果から対訳データ要素をファイルまたは，画面に出力するかどうかを判定する．ここで用いる出力条件は１層検索，２層検索（AND／OR），同一位置である．

・１層検索

A検索の結果のみを評価し，判定結果を返す．評価方法は，結果が True ならば，True を返し，False ならば，False を返す．

・２層検索

出力条件とA，B検索の結果を評価して，判定結果を返す．

出力条件が AND の場合には，A検索の結果が True，かつB検索の結果が True の場合において，True の判定結果を返し，そうでなければ，False の判定結果を返す．

出力条件が OR の場合には，A検索とB検索の結果のうちどちらか一方が True ならば，True の判定結果を返し，両方 False の場合は False の判定結果を返す．

・同一位置

同一位置は，A検索でのキーワード位置とB検索でのキーワード位置を比較し，一致している場合は True を返し，一致していなければ，False を返す．このときに比較するキーワードの位置は⑶の完全一致検索を行ったときの単語の位置のデータである．そのため，同一位置の判定は検索条件を完全一致検索としなければならない．また，同一位置は出力条件を AND とした場合においてのみ動作する処理である．

次に，同一位置における判定の例を示す．

〈例〉
A検索層の文：〝ku＝ku　rusuy　na　.〞
A検索キーワード：〝ku〞
B検索層の文：〝私＝飲む　たい　な　。〞
B検索キーワード：〝私〞

A検索の結果：True，B 検索の結果：True
A検索キーワードの位置：１， 3
B検索キーワードの位置：１
一致する位置：１
判定結果：True
〈例〉
A検索層の文：〝ku＝ku　rusuy　na　.〟
A検索キーワード：〝rusuy〟
B検索層の文：〝私＝飲む　たい　な　。〟
B検索キーワード：〝私〟
A検索の結果：True，B 検索の結果：True
A検索キーワードの位置：３
B検索キーワードの位置：２
一致する位置：なし
判定結果：False

⑸　文中のキーワードに印をつける

表示・出力を行うデータにおいて，キーワードの前後に印をつけることで，検索したキーワードがどこにあるかを示す.この処理を以降では「マーキング処理」と呼ぶ.

まず，基本として，入力文中にキーワードが１つだけ存在する場合におけるマーキング処理について考える．マーキング処理は，大きく３つの処理に分けられる.まず,①入力文中に含まれるキーワードの位置データを検索処理段階で求めておき，そのデータを元に，入力文を〔キーワード前の文字列〕と〔キーワード〕と〔キーワード後の文字列〕の３つの文字列に分ける．次に，②キーワードに印を付け，新しい文字列として置き換える．そして，③〔キーワード前の文字列〕と〔印を付けた新しい文字列〕と〔キーワード後の文字列〕を結果文として統合する．これら①～③の処理を行い文中のキーワードに印を付けていく.

マーキング処理の例を次に示す.

〈例〉
・入力文＝〝単語１__単語２__単語３__単語４__
　　単語５〟
　※（〝__〟は全角スペースを表す.）
・キーワード＝〝単語２〟,
・キーワードに付ける印＝〝◆〟
①文を解体する
文字列１（キーワードの前）＝〝単語１__〟
文字列２（キーワード）＝〝単語２〟
文字列３（キーワードの後）＝〝__単語３__単語４
　　__単語５〟
②キーワードに印をつける
文字列２′＝〝◆〟＋文字列２＋〝◆〟
　　＝〝◆単語２◆〟
③統合する
結果文＝文字列１＋文字列２′＋文字列３
　　＝〝単語１__◆単語２◆__単語３__
　　単語４__単語５〟

キーワードは１つの文字列に複数個存在する場合があるので，文字列の配列（ArrayList）を用いて，解体された文が格納できるようにした.

マーキング処理を行った後に，その結果の表示・出力を可能とするために，前述のファイル読み込みで得られたデータ要素から表示・出力するためのデータ要素を作成しなければならない．そのために，読み込みで得られたデータ要素中の指定した層のデータとマーキング処理を施したデータを次の例のように置き換えなければならない.

〈例〉
データ要素の構造
　データ要素
　　｜－層１
　　｜－層２
　　｜－層３
　　｜－層４
　　｜－層５
検索する層＝〝層３〟
処理後の層３のデータ＝〝層３′〟

データ要素	（新）データ要素
｜－層１	｜－層１
｜－層２	｜－層２
｜－層３←層３′	｜－層３′
｜－層４	｜－層４
｜－層５	｜－層５

⑹　表示・出力をする

ここでは⑷での結果が True の場合はデータ要素を出力し，False の場合はデータ要素を表示･出力しないようにする.

「GUI の画面に結果を表示する」のか，「結果を表示し，なおかつファイルに出力する」のかは，ユーザが次章で示す GUI 画面の〝検索表示〟ボタンを押すか〝結果出力〟ボタンを押すかによって決められる．また，ファイルに出力する場合において，入力ファイル名と出力ファイル名が一致した際には，処理を行わず注意文が出るようになっている.

5．層指定検索ツールの利用

5.1　層指定検索ツールの GUI 画面

層指定検索ツールの GUI 画面を下に示す．

図 2　GUI 画面

・入力ファイル名：検索する対訳データファイル
〈GUI 画面での例〉Exp6Final.txt
・出力ファイル名：検索結果を出力するファイル
〈例〉out.txt
・データの階層：対訳データ要素の階層
〈例〉 6
・検索する階層：対訳データ要素の検索する階層
〈例〉A検索：１，B検索：３
・検索する語：検索キーワード
〈例〉A検索：kor，B 検索：持つ
・完全一致：完全一致検索を行うかどうか
〈例〉A検索：選択する，B検索：選択する
・部分一致：部分一致検索を行うかどうか
〈例〉A検索：選択しない，B検索：選択しない
・NOT：否定検索を行うかどうか
〈例〉A検索：選択しない，B検索：選択しない
・印を付ける：マーキング処理を行うかどうか
〈例〉A検索：選択する，B検索：選択する
・複数検索条件：
　A：階層Aのみ検索（一層検索）を行う
　A AND B：出力条件を AND にする
　A OR B：出力条件を OR にする
　同一位置：出力条件を同一位置にする
〈例〉A AND B

完全一致検索と部分一致検索は１つの検索において同時に使用することはないのでラジオボタンとしている．また，否定検索では完全一致検索でヒットしないデータ要素の出力と，部分一致検索でヒットしないデータ要素の出力を区別できるようにチェックボタンとなっている．

さらに，GUI 画面では，リストボックスには検索結果が出力され，その下に検索ヒット数が 12 であることも表示されている．

5.2　層指定検索ツールの利用例と考察

アイヌ語・日本語機械翻訳システムの作成を目的とし，アイヌ語・日本語対訳データを用いた言語学的考察を進めているが，そこでのデータ収集が層指定検索ツールを利用することで容易になった．本節では，層指定検索ツールの具体的な利用例とその出力結果についての考察を行う．

⑴　場所表現〝or〟の訳出について

アイヌ語〝or〟の訳出では，まず，「所」と訳出する場合と訳出しない場合がある．その場合分けは〝or〟の直前の名詞に場所性という特徴があるかどうかによって判断することができると考える．ない場合は訳出し，ある場合は訳出しない．場所性は，名詞が表現する対象の意味属性として，名詞の意味に付与されているものとする．場所性のある名詞として，「札幌」などの地名や「川」などがある．

また，〝or〟には「中」という訳もあり，この場合も，訳出する場合と訳出しなくてもよい場合がある．〝or〟の直前の名詞の内容性という特徴の有無を判断し，内容性がない場合は訳出し，ある場合には訳出してもしなくてもよいと考える．内容性は，「内」とか「中」を有する対象を表現する名詞の意味に付与されているものとする．内容性のある名詞として，「樽」や「おかゆ」などがある．場所表現についての考察は文献[13]でより詳しく論じている．

層指定検索ツールを段階的に利用して収集した〝or〟に関連する対訳データ要素の出力例を以下に示す．対訳データファイルは，文献[1,2]から作成した対訳データを蓄積した〝All6cp.txt〟という名前のテキストファイルである．

○段階的な検索
［1回目の検索］（位置名詞〝or〞の検索）
入力ファイル：All6cp.txt
入力ファイルのデータ要素数：306
A検索
検索層：1，　検索キーワード：or
検索条件：完全一致
B検索
検索層：2，　検索キーワード：位名(位置名詞)
検索条件：完全一致
出力条件：AND，同一位置
検索出力数：60
［2回目の検索］
入力ファイル：1回目の出力データファイル
［2回目その1］（〝or〞の訳が「所」）
A検索
検索層：3，　検索キーワード：所
検索条件：完全一致
出力条件：A検索のみ
検索出力数：40
［2回目その2］（〝or〞の訳が「中」）
A検索
検索層：3，　検索キーワード：中
検索条件：完全一致
出力条件：A検索のみ
検索出力数：21
○出力データ（一部）
・「所」の場合
〈1〉
exp 0300111：nisatta　Nupurpet　◆or◆　un e＝arpa　ya　？
exp 0300102：名詞　固名　◆位名◆　格助　人接＝自動　終助　？
exp 0300103：明日　登別　◆所◆　へ　あなた＝行く　か　？
exp 0300104：名詞　固名　名詞　格助　人接代＝自動　終助　？
exp 0300105：明日登別へ行くか？
exp 0300106：明日は登別へ行くのかい？
〈2〉
upa 0500301：aep　ramacihi　anakne　sinrit　◆or◆　un　arpa　sekor　a＝ye　．
upa 0500302：名詞　名詞・所　副助　名詞　◆位名◆　格助　自動　接助　人接＝他動　．
upa 0500303：食べ物　魂　は　先祖　◆所◆　へ　行く　と　人＝言う　。
upa 0500304：名詞　名詞　副助　名詞　名詞　格助　自動　接助　人接代＝他動　。
upa 0500305：食べ物の魂は先祖の所へ行くと言う。
upa 0500306：食べ物の魂は，先祖の所に行くと言われています。
・「中」の場合
〈3〉
exp 1400321：wakka　ku＝ta　wa　ontaro　◆or◆　k＝omare　wa　an　na　．
exp 1400302：名詞　人接＝他動　接助　名詞　◆位名◆　人接＝複他動　接助　自動　終助　．
exp 1400303：水　私＝汲む　て　樽　◆中◆　私＝入れる　て　ある　よ　。
exp 1400304：名詞　人接代＝他動　接助　名詞　名詞　人接代＝他動　接助　自動　終助　。
exp 1400305：水を汲んで樽の中にいれてあるよ。
exp 1400306：水は私が汲んで樽にいれてあるわよ。
〈4〉
upa 0100401：sayo　◆or◆　setanto　omare　wa　pon　kasup　ani　en＝ere　．
upa 0100402：名詞　◆位名◆　名詞　複他動　接助　自動　名詞　後副　人接＝複他動　．
upa 0100403：おかゆ　◆中◆　ナギナタコウジュ　入れる　て　小さい　しゃもじ　で　私＝食べさせる　。
upa 0100404：名詞　名詞　名詞　他動　接助　形容　名詞　格助　人接代＝他動　。
upa 0100405：おかゆの中にナギナタコウジュを入れて小さいしゃもじで私に食べさせた。
upa 0100406：おかゆの中にナギナタコウジュを入れて匙で食べさせてくれました。
○出力結果についての考察
〈1〉のデータでは，〝or〞の直前に地名を表す固有名詞の〝Nupurpet（登別）〞がある．地名は場所を示す語で場所性があると考えられるため，「登別の所へ」ではなく，「登別へ」と訳出されている．また，〈2〉のデータでは，〝or〞の直前に名詞の〝sinrit（先祖）〞があるが，〝先祖〞は人（物）を指す語であり，場所性がないと考えられる．そのため，「先祖の所」と訳出されている．
〈3〉，〈4〉は，〝or〞の直前に〝ontaro（樽）〞，〝sayo（おかゆ）〞がある．〈3〉の5層目の文では，「樽の中に」と訳出され，6層目では「樽に」と訳出されている．一般的な知識として〝樽〞の中に

『ものが入れられる，または入っている』と考えることができるため，〝樽〟には内容性があると考えることができ，〝中〟を訳出してもしなくてもよいと考えることができる．また，〈4〉の5層目，6層目では,「おかゆの中にナギナタコウジュを入れて」と訳出されているが,「おかゆにナギナタコウジュを入れて」と訳出しても日本語として問題はない．〝おかゆ〟には内容性があり，〝中〟を訳出してもしなくてもよいと考える．

⑵　動詞〝kor〟の訳出について

アイヌ語〝kor〟の訳出については，所有形式として「の」と訳出する場合と動詞として「持つ」と訳出する場合があり，〝kor〟の直後の単語に注目した結果，その場合分けをするための以下のようなルールを設定することができる．

①〝kor〟の次に来る単語が普通名詞の場合は所有形式で訳出する．

②〝kor〟の次に来る単語が普通名詞以外の品詞の場合は「持つ」と訳出する．なお，普通名詞以外の品詞とは形式名詞や接続助詞，助動詞などを指す．

層指定検索ツールを段階的に利用して得られた出力データを次に示す．

○段階的な検索

［1回目の検索］（〝kor〟の訳が「持つ」）

入力ファイル：All6cp.txt
入力ファイルのデータ要素数：306
A検索
検索層：1，　検索キーワード：kor
検索条件：完全一致
B検索
検索層：3，　検索キーワード：持つ
検索条件：完全一致
出力条件：AND，同一位置
検索出力数：48

［2回目の検索］

入力ファイル：1回目の出力データファイル

［2回目その1］（所有形式での訳出）

A検索
検索層：5，　検索キーワード：の
検索条件：部分一致
出力条件：A検索のみ
検索出力数：44

［2回目その2］（〝kor〟の訳が「持つ」）

A検索
検索層：5，　検索キーワード：持
検索条件：部分一致
出力条件：A検索のみ
検索出力数：15

○出力データ（一部）

・所有形式で訳出する場合

〈5〉

xp 0700421：e＝◆kor◆　seta　e＝tura ruwe　？
exp 0700402：人接＝他動　名詞　人接＝他動　形名　？
exp 0700403：あなた＝◆持つ◆　犬　あなた＝連れていく　の　？
exp 0700404：人接代＝他動　名詞　人接代＝他動　形名　？
exp 0700405：あなた◆の◆犬を連れて行った◆の◆？
exp 0700406：おまえの犬を連れて行ったのかい？

〈6〉

exp 1100231：e＝◆kor◆　totto　soyne　wa　isam wa　．
exp 1100302：人接＝他動　名詞　自動　接助　自動　終助　．
exp 1100303：あなた＝◆持つ◆　おかあさん　外出する　て　いない　よ　。
exp 1100304：人接代＝他動　名詞　自動　接助　連語　終助　。
exp 1100305：あなた◆の◆おかあさんは外出していないよ。
exp 1100306：おまえのおかあさんは外に出かけているよ。

・動詞「持つ」と訳出する場合

〈7〉

upa 0300201：soy　ta　an　cikuni　ku＝◆kor◆　wa　cise　onnayke　k＝omare　．
upa 0300202：位名　格助　自動　名詞　人接＝他動　接助　名詞　位名・所　人接＝複他動　．
upa 0300203：外　に　ある　薪　私＝◆持つ◆て　家　中　私＝入れる　。
upa 0300204：名詞　格助　自動　名詞　人接代＝他動　接助　名詞　名詞　人接代＝他動　。
upa 0300205：外にある薪を◆持◆って家の中に入れる。
upa 0300206：外にある薪を持って家の中に入れます。

〈8〉
upa 0800801：SYOCU anak a＝◆ kor ◆ rusuy ka somo ki sekor yaynu korka ene a＝kar hi ka isam . onaha SORI SYOCU ore wa ninpa wa unihi ta hosipi .
upa 0800802：名詞 副助 人接＝他動 助動 副助 副詞 他動 接助 自動 接助 副詞 人接＝自動 形名 副助 自動 . 名詞・所 名詞 名詞 他動 接助 他動 接助 名詞・所格助 自動 .
upa 0800803：焼酎 は 私＝◆持つ◆ たいも ない する と 思う けれど どう 人＝する こと も ない 。 父 ソリ 焼酎 乗せる て 引きずる て 家 に 帰る 。
upa 0800804：名詞 副助 人接代＝他動 助動 副助 助動 他動 接助 自動 接助 副詞 人接代＝自動 形名 副助 助動 。 名詞 名詞 名詞 他動 接助 他動 接助 名詞 格助 自動 。
upa 0800805：焼酎は◆持◆ちたくもないと思うけれどもどうすることもできない。父はソリに焼酎を乗せて引きずって家に帰る。
upa 0800806：「焼酎なぞ欲しくもない」と父親は思ったけれど，どうすることも出来ません。焼酎をソリに乗せそれを引っぱって家に帰りました。
○出力結果についての考察

データ〈5〉，〈6〉では，〝kor〞の直後に普通名詞の〝seta（犬）〞や〝totto（おかあさん）〞が確認できる．また，データ〈7〉，〈8〉では，〝kor〞の直後に接続助詞の〝wa（て）〞や助動詞の〝rusuy（たい）〞が確認できる．

所有形式は，ものが誰かに所有されていることを示すための表現形式で，〝kor〞の直後の品詞は名詞だけであり，その場合は「の」と訳出し，その他の品詞の場合は「持つ」と訳出することができると考えられる．

6．おわりに

層指定検索ツールにおける現時点での改良点を以下のようにまとめることができる．

・出力条件〝同一位置〞で検索する場合，同一位置で対応するキーワードに印を付けられるようにする．
・対訳データごとに対応した分割対象の記号をユーザが任意に設定できるようにする．
・出力ファイルにヒット数や検索条件の項目を出力できるようなコメント機能をつける．

現在，層指定検索ツールを利用した結果についての考察は人間の手によるところが多く，その部分でまだ自動化に向けての改良の余地がある．解析の自動化の一つとして，対訳データから単語のつながりの頻度を解析するツールの開発などがあり，解析や考察における人間の労力をできる限り少なくしていく方法を検討していきたい．そして，アイヌ語と日本語の対照言語学的な考察を層指定検索ツール等の利用で効率的に行いながら，アイヌ語・日本語機械翻訳システムの構築を進めていくことが今後の課題である．

【参考文献】

1）中川裕・中本ムツ子：エクスプレスアイヌ語，白水社，1997.
2）中本ムツ子・片山龍峯：アイヌの知恵 ウパシクマ Ⅰ，Ⅱ，片山言語文化研究所，1999，2001.
3）切替英雄：アイヌ神謡集辞典，大学書林，2003.
4）久野禎子・久野靖：Java によるプログラミング入門，共立出版，2001.
5）岩谷宏：決定版ひとつ上を行く Java の教科書，技術評論社，2005.
6）井上智洋：新しい Java の教科書，ソフトバンククリエイティブ，2006.
7）山田隆太・村岡敦彦：Java・オブジェクト指向の壁を突破する抽象化プログラミング入門，技術評論社，2007.
8）浅海智晴：やさしい UML 入門，ピアソンエデュケーション，2001.
9）長瀬嘉秀・橋本大輔：独習 UML，テクノロジックアート，2005.
10）児玉公信：UML モデリングの本質，日経 BP 社，2004.
11）木村浩之：アイヌ語・日本語機械翻訳システムの開発，平成 14 年度卒業研究報告，2002.
12）小杉一峰：JavaScript を用いたキーワード検索，平成 19 年度卒業研究報告，2007.
13）桃内佳雄・安曇恭徳：アイヌ語・日本語機械翻訳のための場所表現に関する基礎的考察，工学研究，第 7 号，pp.35-42，2007.