

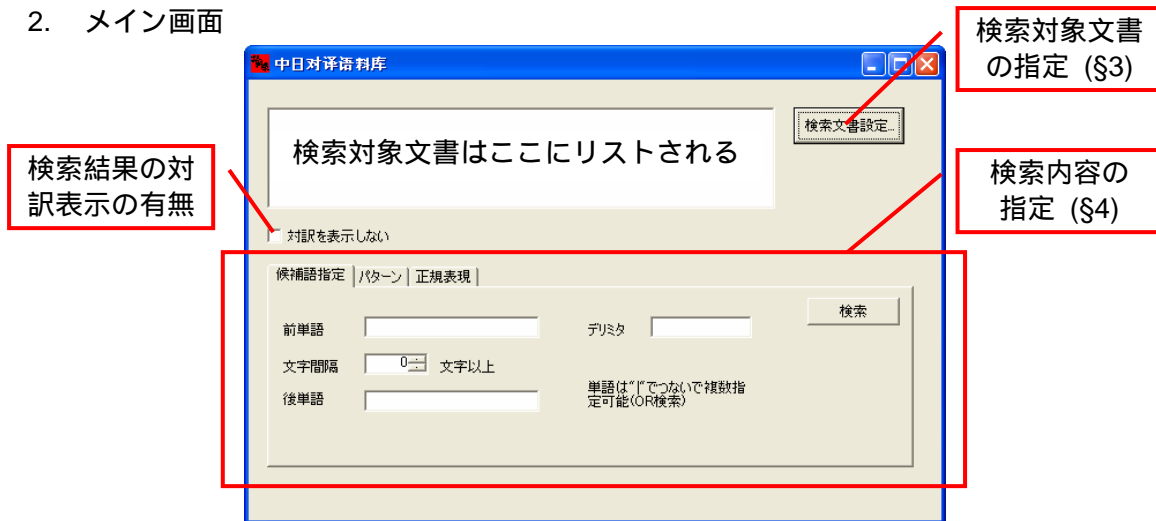
## 中日対訳コーパス（「中日対訳語料庫」）第一版の利用方法

曹 大峰(北京日本学研究中心, 国立国語研究所), 千葉 庄寿(麗澤大学)

### 1. コーパスの概略

- 企画・開発：北京日本学研究中心
  - 「中日対訳コーパスの構築とその応用研究」(1998年9月～) 中国社会科学基金研究プロジェクト, 研究代表者：徐一平(北京日本学研究センター)
  - 第一版公開：2003年7月
- 内容：小説, エッセイ, 伝記, 政治評論・白書, 法律関連文書・条約文書, 詩など各ジャンルの中日対訳テキスト。「中日対訳コーパスの構築とその応用研究」プロジェクトで開発された2千万字余りの対訳データのうち, 第一版では以下を収録。
  - 文学作品：中国23篇, 日本22篇とその訳本(合計105件, 約1130.3万字)
  - 文学以外：中国14篇, 日本14篇, 日中共同2篇とその訳本(合計45件, 約574.6万字)
- 動作環境：
  - ハードウェア：PC/AT互換機であること, メモリ128MB以上, HDDに70MB以上の空き容量
  - ソフトウェア：Microsoft Windows NT/2000/XP (Windows 98/MEでの動作は制限あり), Microsoft .NET Framework 1.0, MDAC 2.6以上, Microsoft Internet Explorer 5.5以上
- 特徴：
  - 日本語原文, 日本語訳文, 中国語原文, 中国語訳文の4つのカテゴリーから検索。
  - 日本語, 中国語の両言語の検索が可能(中国語データはGB2312, 日本語データはShift JISで構築)。
  - 検索結果を訳文とともに表示(対訳を非表示にすることも可能)。
  - 文字を指定せずに, パターンのみでの検索が可能。
  - 正規表現のサポート (Microsoft .NET Frameworkの正規表現を使用)。
- 利用条件：
  - 1ライセンスにつき1台のコンピュータにインストールして利用する。
  - 知的所有権などに関する「覚書」を提出する。
  - 指定の書式によりバグレポートをおこなう。
  - コーパスを利用した研究成果物はコピーを提出する。

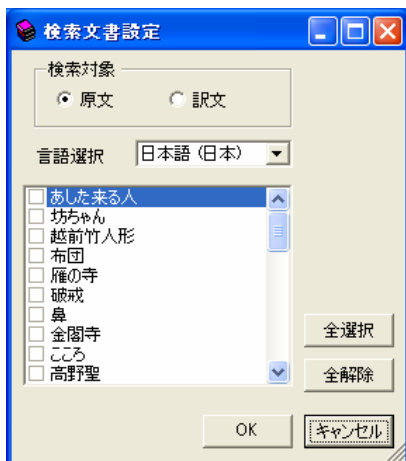
2. メイン画面



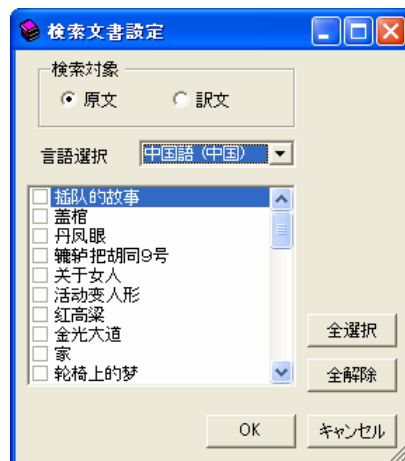
3. 検索文書の指定

- 日本語原文・日本語訳文，中国語原文，中国語訳文の4つのカテゴリーで，文書の個別選択・全選択が可能。なお，複数のカテゴリーを同時に横断検索することはできない。

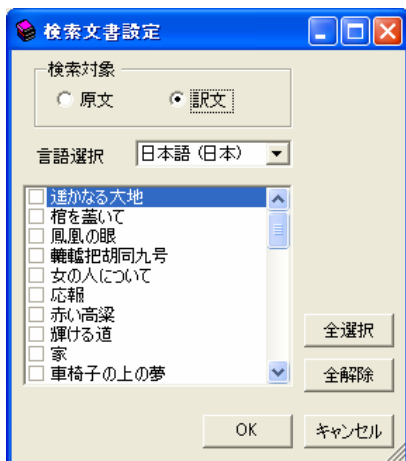
3.1. 日本語原文



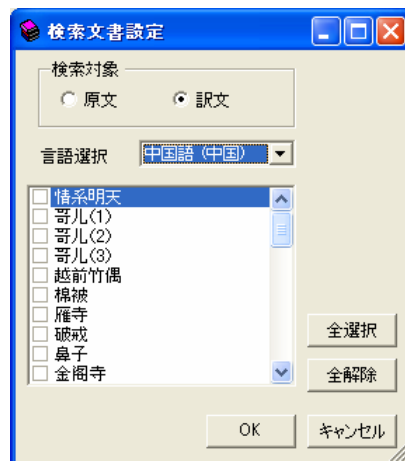
3.3. 中国語原文



3.2. 日本語訳文



3.4. 中国語訳文



## 4. 検索語の指定

- 検索文字列は Unicode で処理され、全角・半角を区別する。

## 4.1. 候補語指定

- 前後 2 つのパターンを指定した「文型指定」が可能。その際、「文字間隔」(最低値)が指定できる(文字間隔の最大値を指定することはできない)。
- 1 語だけ指定する場合には「前単語」に記述する。
- 検索前単語と後単語の間に含めない文字を「デリミタ」として指定する。例：。。。(文をまたがないで検索する場合。「中日対訳コーパス」第一版では、句点が通常日本語で使われる「。」(U+3002)ではなく「。」(U+FF61)で入力されている箇所があるので両方を並べて指定する。)

- パイプ | で区切ることで複数のパターンを指定できる (OR 検索)。例：中国|日本|韓国
- 文字列の一部について、複数のパターンを指定したい場合は丸カッコ () 内に | で区切って記述する。例：(工夫|くふう)する
- 複数の文字を四角カッコ [] 内にまとめて指定できる。例：中[心央間]
- [^ と ] の中に含めたくない文字をまとめて指定できる。例：中[^国華文]

## 4.2. パターン

- 文字を問わず AABB, ABAB, ABB, ABA, AAB, AA の 6 パターンで指定。
- A, B にはそれぞれ特定の 1 文字を指定することも可能。

## 4.3. 正規表現

- .NET Framework の正規表現が使用できる(プログラミング言語 Perl の正規表現に近いが、一部表記方法や実装内容が異なる)。
- 詳細は『.NET Framework 開発者ガイド』の「.NET Framework の正規表現」([http://msdn2.microsoft.com/ja-jp/library/hs600312\(VS.80\).aspx](http://msdn2.microsoft.com/ja-jp/library/hs600312(VS.80).aspx))などを参照。

5. 検索結果

- 検索結果は Internet Explorer 上に表示される。検索ごとに個別のウィンドウが開く。
- 例文番号，行番号とともに検索されたテキストが表示される。
- 対訳が並んで表示される（検索時にオプションで表示・非表示を選択できる）。
- 検索結果は表形式で順次表示されていき，最後に全ヒット数が表示される。

検索終了 - Microsoft Internet Explorer

検索条件 = 候補語指定: 前単語(考え)

No	行番	あした来る人 (原文)	情系明天 (訳文)
1	10	浜名湖の湖面には春の陽が散っているが、湖面の色も、湖の流れ方も寒々としている。まだ冬を感じである。湖面の海への切り口に、白い波濤が、これも冷たく砕けている。曾根二郎はちょっと考える風にしていたが、網棚の方へ手を伸ばして、重そうなりリュックサックを降ろすと、それを肩にかけて歩き出そうとした。	浜名湖上，尽管春日阳光晃晃闪烁，但无论湖面的色调还是水的流姿，都显得冷气森森，给人以寒冬未尽之感。湖水入海口处，雪浪叠起，也是冷清清地栽下海去。曾根二郎沉吟片刻，把手伸向行李架，取下似乎重重的旅行背囊，挎在肩上。刚要迈步，忽听得有人叫他：
2	47	「(ま)くは酒三本とトンカツを食べて三百円払いました。その時は廉いなと思って払ったんですが、あとで考えてみると、どうも、貴女の伝票とつつかえたのではないかと思うんです」	“我喝三瓶酒吃一个炸猪排，才花三百元。当时以为很便宜，但后来一想，才意识到可能同您弄错了传票。”
=中略=			
123	5934	ホテルの玄関口まで杏子は行先を <b>考え</b> ていたが、玄関を出たところで、	杏子边往饭店大门口边走边琢磨去哪里合适，一出大门，杏子提议：
124	6005	梶は久しぶりで気持のいい散歩をしたと思った。こうした場所を歩くのは何年ぶりだろう。何十年ぶりかも知れない。そののんびりした気持の中で、梶は杏子の背後姿をながめながら、若い羚羊を自由にしてやらなければなるまいと <b>考え</b> ていた。克平にしろ、曾根にしろ、八千代にしろ、みんな欠点はあるが、どこかに自分などの若い時持っていなかった純粹なものがある。その純粹なもののために、みんな傷ついたり回りの道をしたたりしている。それでいいのかも知れない。	梶觉得自己好久好久没有如此痛快地散步了。在这种场所漫步该是多少年以前的事呢？恐怕已相隔几十度春秋了。梶是那样地怡然自得，一边望着杏子的背影一边暗自思忖，无论如何也要将这年轻的羚羊放回自由的天地。+++克平也好，曾根也好，八千代也好，尽管各有缺点，但又都具有某种自己这代人年轻时所没有的純粹的东西。为那純粹的东西，他们或者受到创伤，或者行走弯路，而这未必不是好事。

全ヒット数 = 124 件

- 複数の対訳データがある場合には，並べて表示される。

検索終了 - Microsoft Internet Explorer

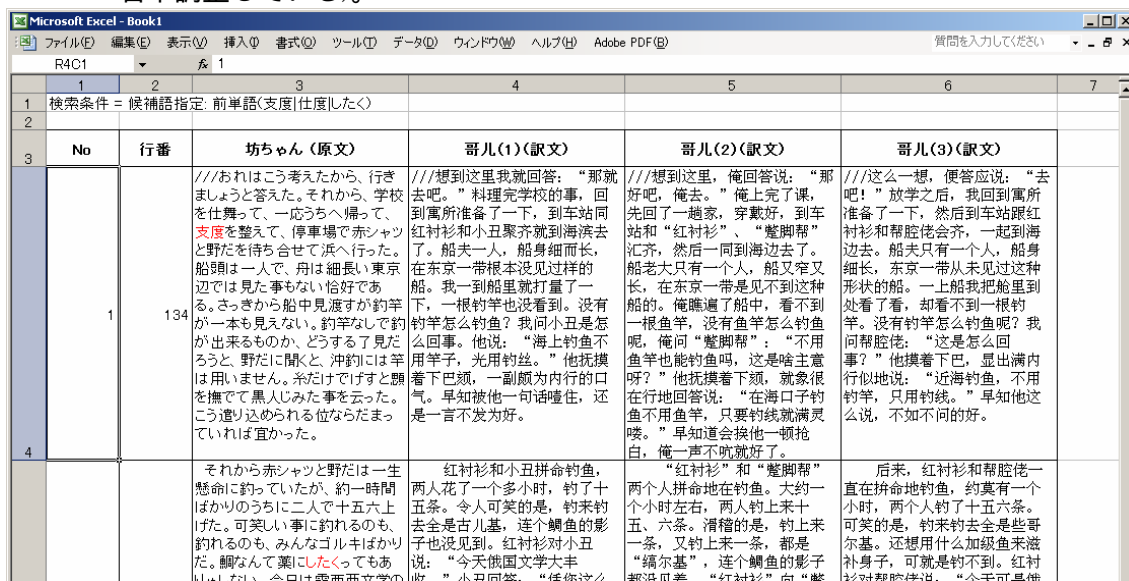
検索条件 = 候補語指定: 前単語(支度[仕度]したく)

No	行番	坊ちゃん (原文)	哥儿 (1) (訳文)	哥儿 (2) (訳文)	哥儿 (3) (訳文)
1	134	///おれはこう考えたから、行きましようと思えた。それから、学校を仕舞って、一応うちへ帰って、支度を整えて、停車場で赤シャツと野たを待ち合せて浜へ行った。船頭は一人で、舟は細長い東京辺では見た事もない恰好である。さっきから船中見渡すが釣竿が一本も見えない。釣竿なしで釣が出来るものか、どうする見だらうと、野たに聞くと、沖釣には竿は用いません。糸だけでせうと題を撫でて黒人じみた事を云った。こう道り込められる位ならたまっていれば良かった。	///想到这里我就回答：“那就去吧。”料理完学校的事，回到寓所准备了一下，到车站同红衬衫和小丑聚齐就到海滨去了。船夫一人，船身细而长，在东京一带根本没见过的船。我一到船里就打量了一下，一根钓竿也没看到。没有钓竿怎么钓鱼？我问小丑是怎么回事。他说：“海上钓鱼不用竿子，光用钓丝。”他抚摸着下巴颏，一副颇为内行的口气。早知被他一句话噎住，还是一言不发为好。	///想到这里，俺回答说：“那好吧，俺去。”俺上完了课，先回了一趟家，穿戴好，到车站和“红衬衫”、“蹩脚帮”汇齐，然后一同到海边去了。船老大只有一个人，船又窄又长，在东京一带是见不到这种船的。俺瞧遍了船中，看不到一根鱼竿，没有鱼竿怎么钓鱼呢，俺问“蹩脚帮”：“不用鱼竿也能钓鱼吗，这是啥主意呀？”他抚摸着下颏，就象很在行地回答说：“在海口子钓鱼不用鱼竿，只要钓线就满灵喽。”早知道会挨他一顿抢白，俺一声不吭就好了。	///这么一想，便答应说：“去吧！”放学之后，我回到寓所准备了一下，然后到车站跟红衬衫和帮腔佬会齐，一起到海边去。船夫只有一个人，船身细长，东京一带从未见过这种形状的船。一上船我把舱里到处看了看，却看不到一根钓竿。没有钓竿怎么钓鱼呢？我问帮腔佬：“这是怎么回事？”他摸着下巴，显出满内行似地说：“近海钓鱼，不用钓竿，只用钓线。”早知他这么说，不如不问的好。
		それから赤シャツと野たは一生懸命に釣っていたが、約一時間ばかりのうちに二人で十五六上げた。可	红衬衫和小丑拼命钓鱼，两人花了一个多小时	“红衬衫”和“蹩脚帮”两个人拼命地在钓鱼。大约一个小时左右，两人钓上来十五、六条。滑稽	后来，红衬衫和帮腔佬一直在拼命地钓鱼，约莫有一个小时，两个人钓了十五六

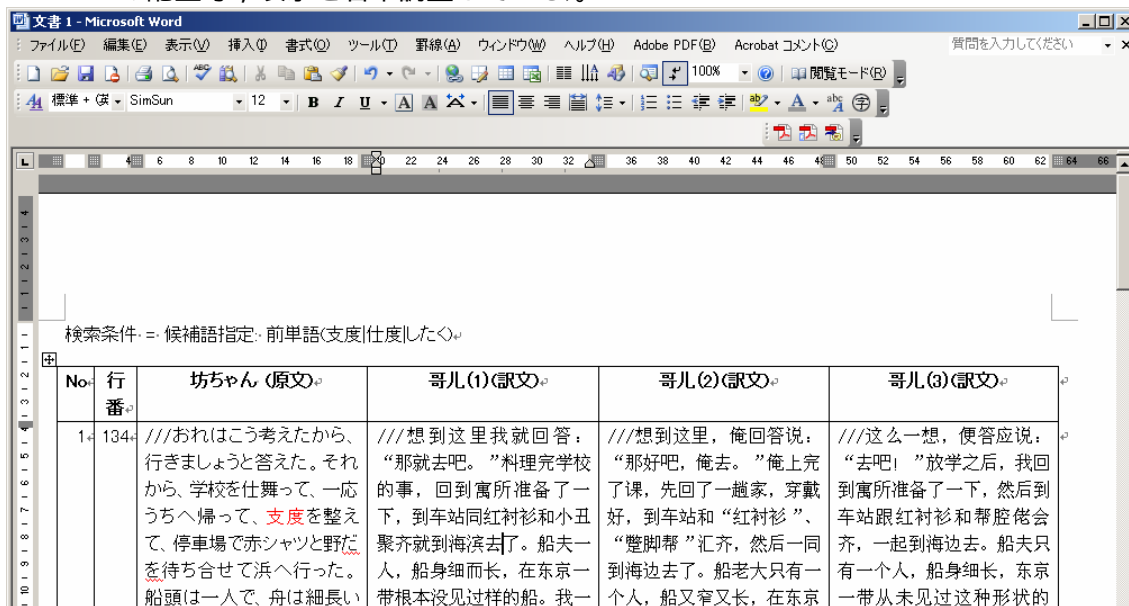
- 検索結果は印刷が可能なほか、Unicode テキストとして保存するか、マウス等で選択しコピーした結果を Unicode 対応テキストエディタや Excel などの表計算ソフト、Word などのワープロソフトなどに貼り付けて利用することができる。
  - Internet Explorer の「ファイル」「名前をつけて保存」で Unicode テキストファイルとして保存（この場合、テキストは BOM なしの UTF-16 として保存される。したがって、エンコードを正しく指定しない場合文字化けがおこることがある）。



- Excel に貼り付け（以下ではセルの大きさやフォント、テキストの配置等、表示を若干調整している）。



- Word に貼り付け（以下ではページを横にし、セルの大きさやフォント、テキストの配置等、表示を若干調整している）。



以上