

The comparative study of textstructure in Chinese and Japanese

メタデータ	言語: jpn 出版者: 公開日: 2017-10-05 キーワード (Ja): キーワード (En): 作成者: Otaki, Sachiko メールアドレス: 所属:
URL	https://doi.org/10.24517/00034785

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 International License.



第3部

検索ソフト「GPSV1」の解説

大瀧 幸子

(金沢大学文学部)

目 次

1. ソフトウェアの概要	77
2. ライブラリの構築・管理	79
2.1. ライブラリの選択	79
2.2. ライブラリの構築	79
2.3. ライブラリの閲覧	81
2.4. その他の事項	83
3. ライブラリからの検索	84
3.1. 一般的なライブラリの検索	84
3.2. 日中対照モードでの検索	87
3.3. レポートへの出力	88
4. ライブラリ構築の実態	90
5. アプリケーション仕様	90

1. ソフトウェアの概要

GPSV1は、中国語及び日本語の電子テキストをライブラリとして蓄積・管理し、電子テキストの閲覧・印刷・全文検索を行うアプリケーションである。

MSLUに対応しているため、日本語・簡体字中国語・繁体字中国語など必要な言語のフォントさえインストールしてあれば、Windows98/NT4.0以降の各国語版上で動作することができる。

テキストライブラリはユーザが自由に構築することができる。また、ライブラリはあえて通常のフォルダにテキストファイルとして保存してあるため、自由に移動が可能である。これにより、目的・カテゴリ別の複数のライブラリ構築が可能になり、またCD-ROMやリムーバブルメディア上にライブラリをコピーできるので可搬性にも優れている。

全てのテキストデータは各文字コードからUnicodeに変換されて保存されるので文字コードの問題は発生しない。また、文書情報を保持している為、作家・作品・版本によってライブラリを整理する事ができ、作家・作品を限定したテキスト検索や、同じ作品の日本語版と中国語版を対照検索が可能である。

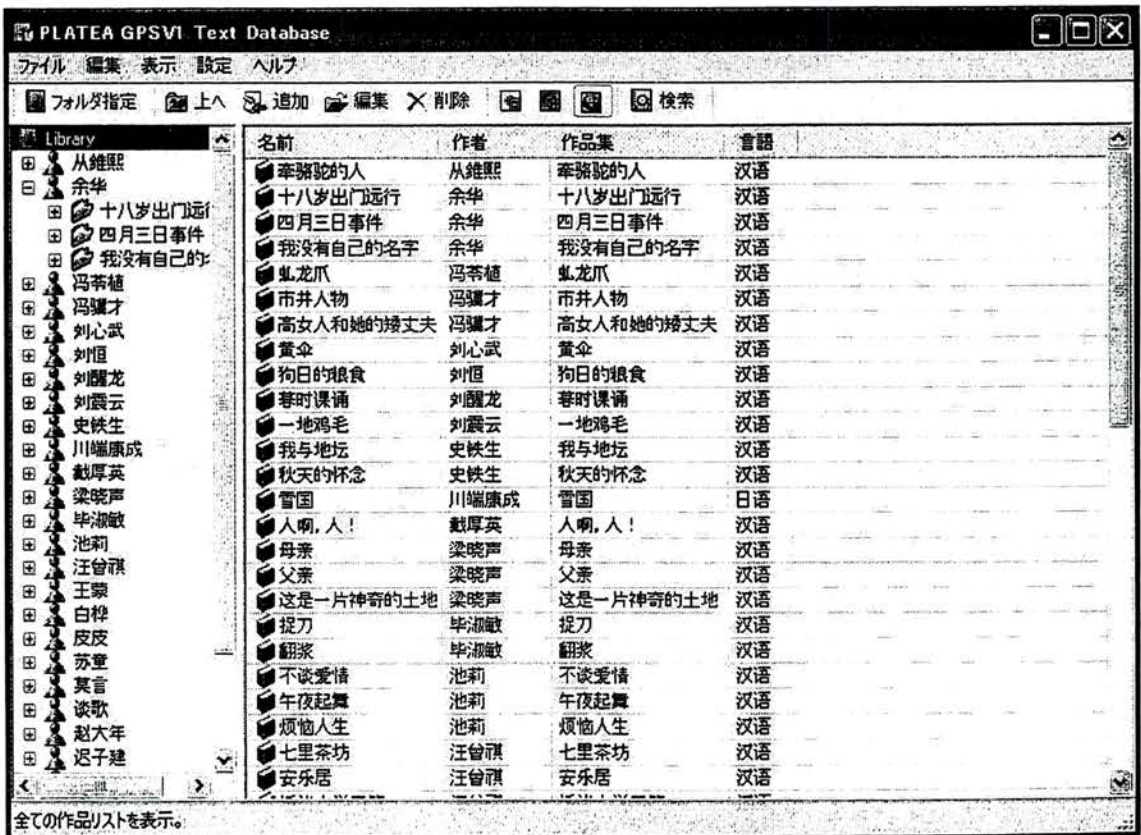


図 1-1. ライブラリ管理画面

構築済みのライブラリからの語句検索は検索結果リストに表示され、必要な例文を取捨選択した後、レポートファイルに保存できる。検索語句に指定された制御文字を使う事で、複数の語句の係り結びなどを検索する事も可能になっている。

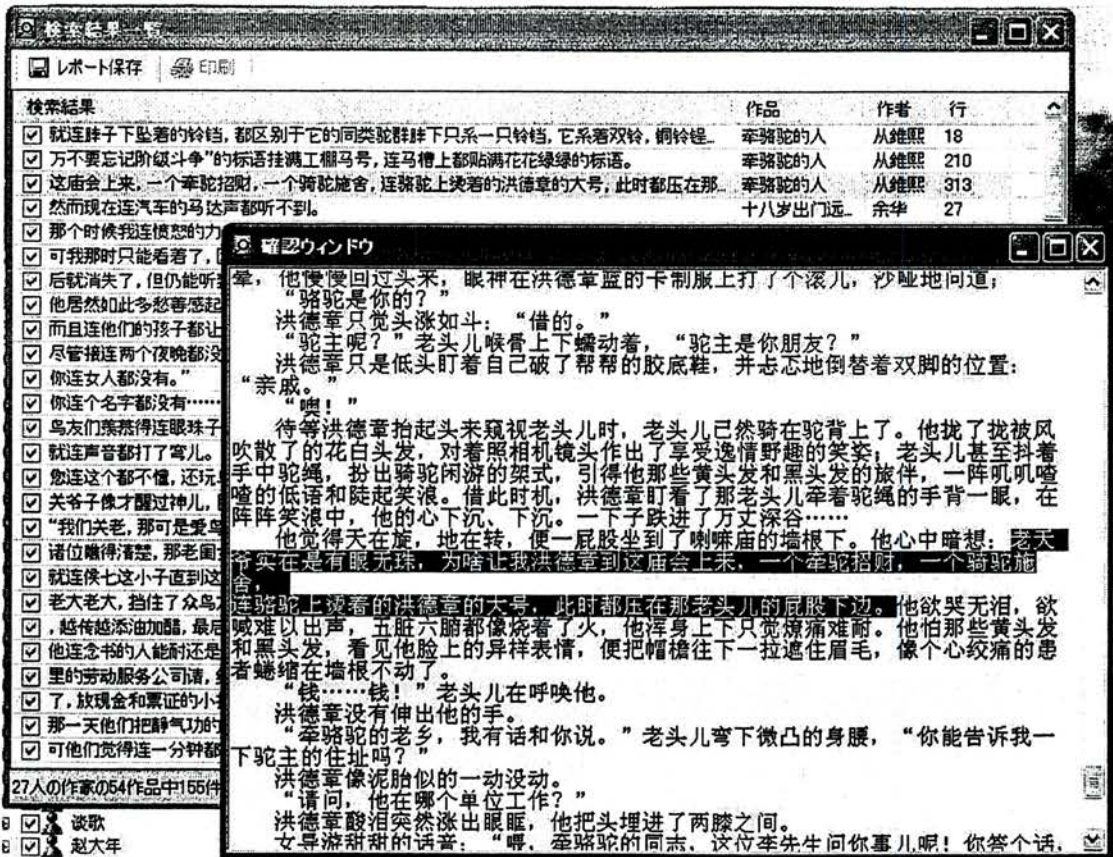


図1-2. 検索結果 (“连~都”を検索)

以上のように、GPSV1は自由に構築可能なライブラリの管理ソフトとテキスト検索ソフトの組み合わせにより、様々な電子テキストソースをライブラリ化し自由に閲覧・検索を行うことのできるアプリケーションである。

2. ライブラリの構築・管理

ライブラリの構築・管理は「GPSV1 Text Database」より行う。

2. 1 ライブラリの選択

任意のフォルダをライブラリの収納先として指定できるため、複数のライブラリの構築やライブラリの別のメディアへの移動などが容易である事は前に述べたとおりである。また、ネットワーク上の共有フォルダにライブラリを移行することにより、複数台のクライアントでライブラリを共有することも可能である。

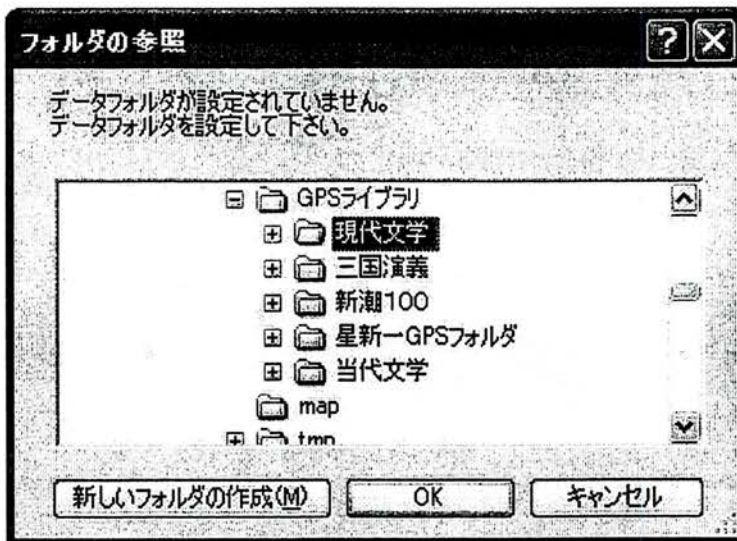


図 2-1. ライブラリの選択

2. 2 ライブラリの構築

ライブラリの構築は任意の電子テキストをライブラリに追加していく事によっておこなう。ライブラリ画面にファイルをドラッグ&ドロップし、そのテキストの情報を入力する事でライブラリへの追加が行われる。現状で対応しているファイル形式は、テキスト形式及び HTML 形式になっている。また、SHIFT-JIS・GB2312・UTF-8・UTF-16 の各文字コードに対応している。(上位版である「GPSXP」ではこれに加えて RTF・Word/XML 形式、BIG5・GB16030 コードに対応する)

この際、作品が複数のテキストファイルに分かれていても、一括してドラッグ&ドロップしたり、追加画面に次々にドラッグ&ドロップしていくことで、一つの作品の中にテキストをまとめていく事ができる。その上、追加された複数のテキストを一つのファイルに

まとめあげることも行わないので、章立ての構造を保存したまま一つの作品としてライブラリ化する事が可能である。

作品情報は、作家・作品集・作品名の3つの情報を保持する。通常のテキストライブラリであれば、作家「鲁迅」の作品集「呐喊」の作品「狂人日記」という風に入力していく事で、テキストの閲覧・検索時に「呐喊」に収められている作品の一覧を取得したり、その中に限定してテキスト検索を行うことができる。また、日中作品を対照させる場合であれば、作家「星新一」の作品「きまぐれロボット」の中に日本語の版本「きまぐれロボット」と中国語の版本「反复无常的机器人」を含めて一括管理する事ができる。

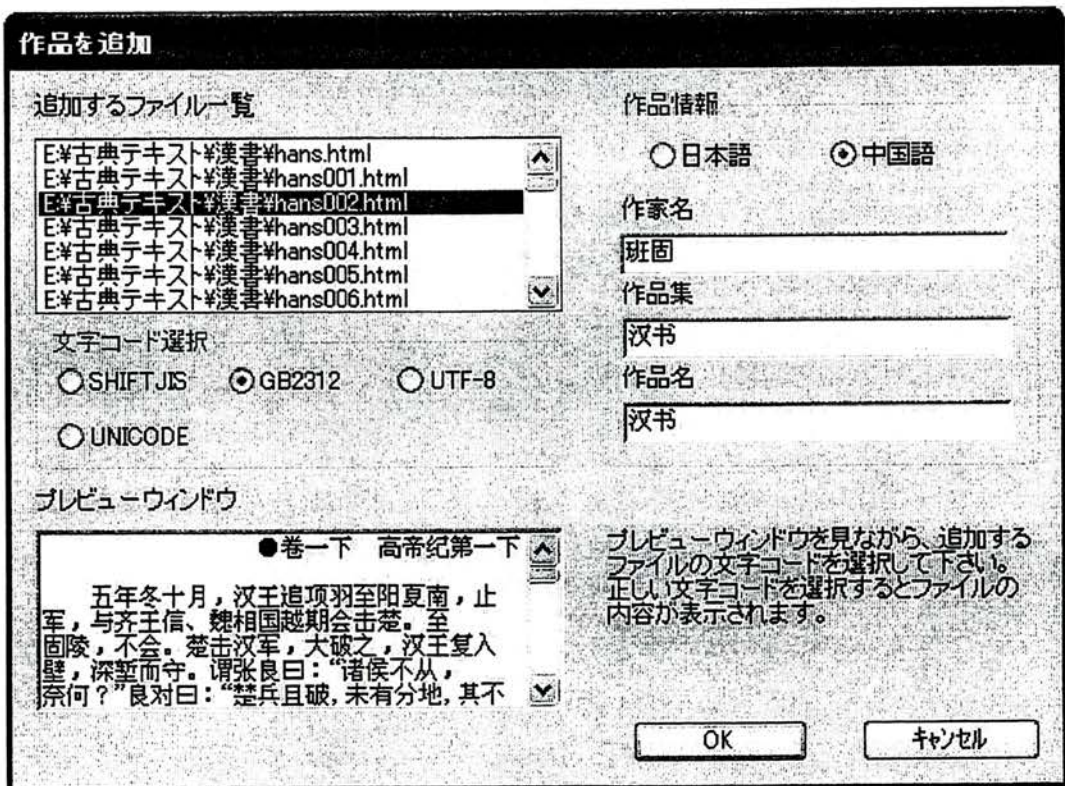


図 2 - 2. ライブラリへの作品の追加

2. 3 ライブラリの閲覧

「Text Database」では構築されたライブラリに含まれるテキストを閲覧・修正・印刷することができる。

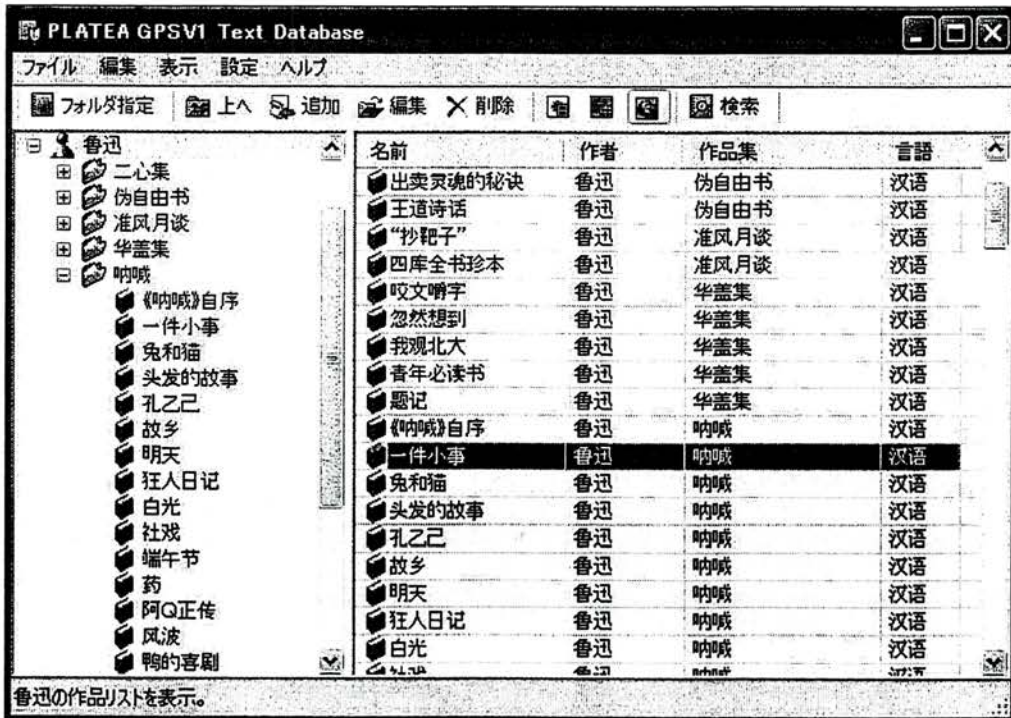


図 2-3. ライブラリ画面

ウィンドウの左側にライブラリに含まれる作品がツリー状に表示され、任意の作家、作品集、作品名を選択する事により、右のウィンドウにそのカテゴリーに含まれる作品・テキストの一覧が表示される。例えば、「鲁迅」を選択すれば鲁迅の作品全てが表示され、「呐喊」を選択すれば「呐喊」に含まれる作品の一覧が表示される。

右に表示された作品をダブルクリックする事で、テキストの内容を表示するエディットウィンドウが開く。

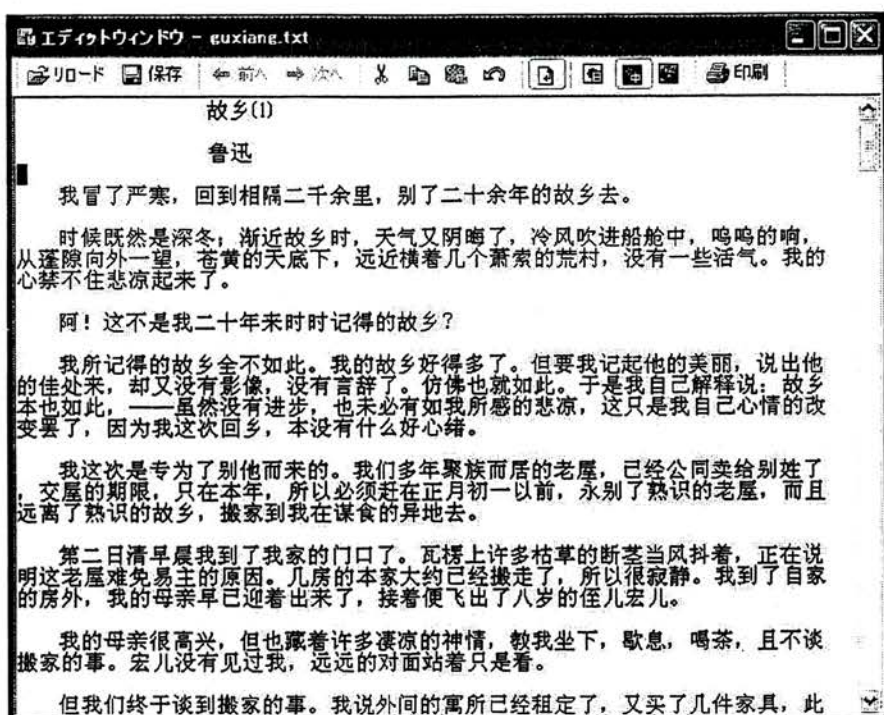


図 2-4. エディットウィンドウ

このウィンドウから、テキストの閲覧・印刷・修正を行うことができる。

また、作品を右クリックするとポップアップメニューが開き、テキスト・作品情報の修正やライブラリからの削除を行うことができる。

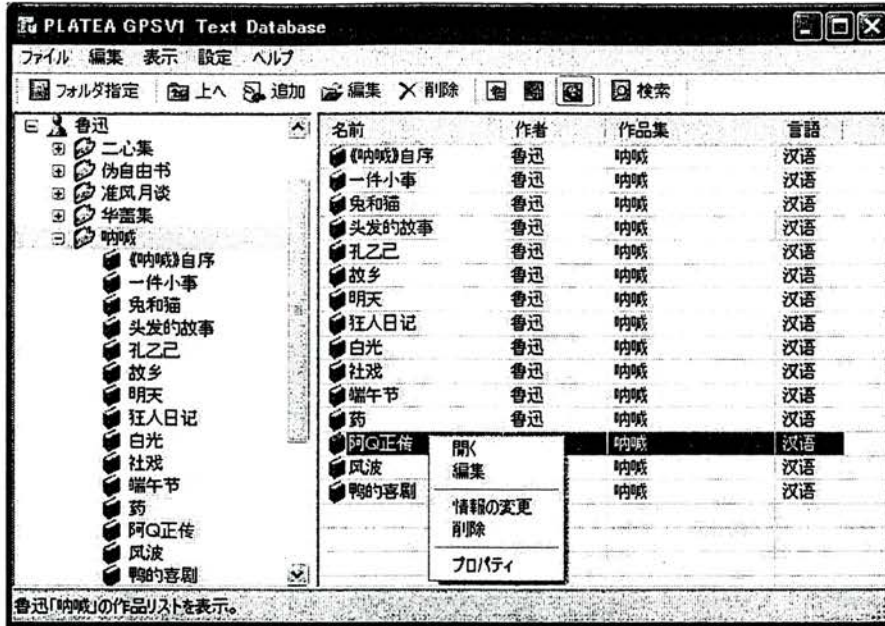


図 2 - 5. 作品の管理

2. 4 その他の事項

「Text Database」の基本的な機能は以上の通りである。

この他に、ライブラリに含まれる作品一覧の印刷や、日本語・中国語・ユニコードフォントを各自任意に設定し表示に用いるフォントを切り替えることによりその場に応じたフォントをフォントマッピング無しにフレキシブルに行うことも出来る。(Win98系ではフォントマッピングが効かない為この機能が必要となる。Win2000/XPでは自動的に行われるため単純なフォント指定機能として働く)

また、デフォルトで搭載しているエディットウィンドウは基本的なエディタ機能しか実装していないので、テキスト閲覧・修正の時に外部エディタを用いるように設定する事もできるようになっている。

3. ライブラリからの検索

3. 1 一般的なライブラリの検索

ライブラリのテキストに含まれる語句検索は、「Search Engine」より行う。

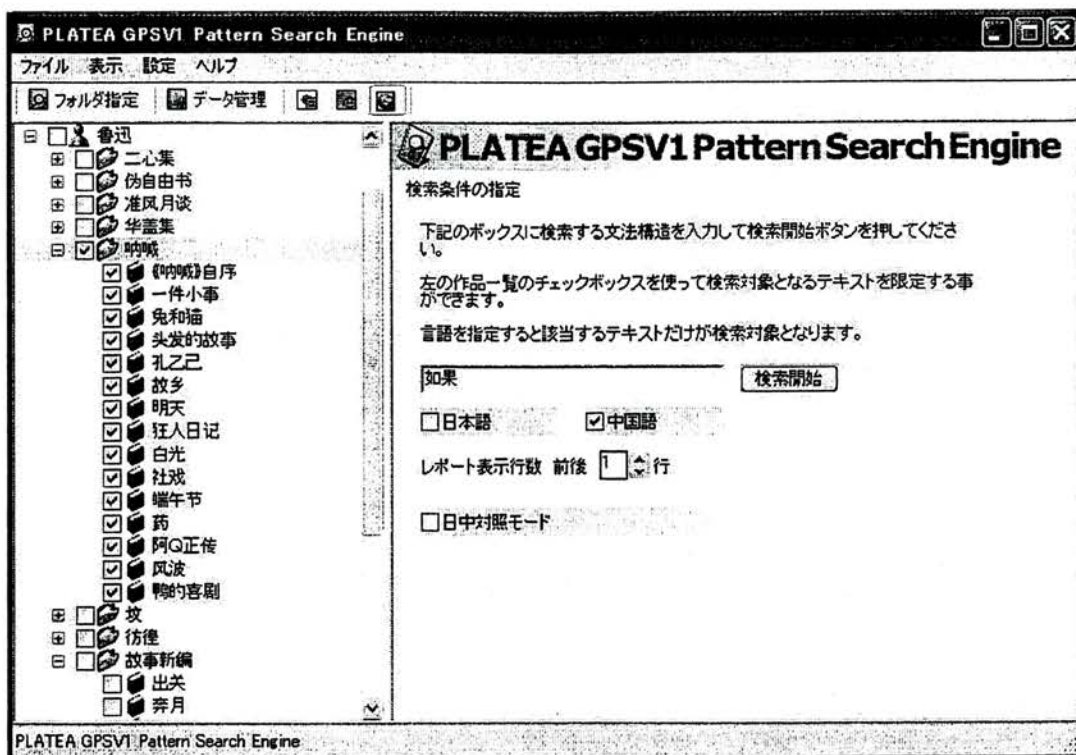


図 3-1. 検索ソフト

検索対象のライブラリを選択すると、左のツリービューに作品一覧が表示される。ここから検索対象とする作品を選択し検索条件を入力して検索開始ボタンを押すと、検索結果の一覧ウィンドウが表示される。

検索条件では、どの言語の版本を対象とするか（片方の言語のテキストしかないライブラリの場合は設定不要）、レポートに表示する行数、日中の版本を対照させるか（両方の版本が無いと無意味）を選択できる。

検索の実行と同時に検索結果一覧に次々とヒットした例文が表示されていく。検索の終了を待たずに結果一覧が表示されるため、ストレス無く検索結果を閲覧する事が可能になっている。この際、例文中の検索語句がリストビューのおよそ中心に表示される。

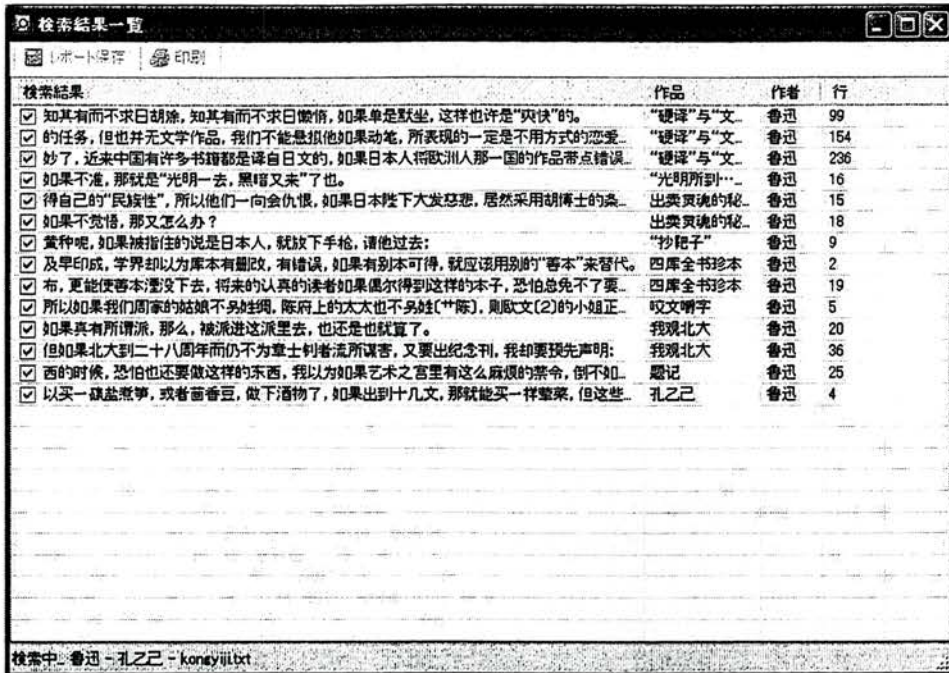


図 3-2. 検索中の検索結果一覧画面

検索結果の先頭についているチェックボックスによって必要の無い例文を排除することができる。また、ダブルクリックすると元のテキストの該当部分が確認ウインドウに表示される。



図 3-3. 確認ウインドウ

3. 2 日中対象モードでの検索

同じ作品の日本語版と中国語版を用意し、テキストにパラグラフ識別用の「@」という記号を挿入した場合は、日中の作品を対照しながら検索することができる。例えば、日本語で「だから」と検索し、検索結果ウィンドウから該当する例文をダブルクリックすると、確認ウィンドウが2つ開き両者の該当する部分が表示される。これにより、擬似的に日中対照コーパスとしてライブラリを用いることができる。

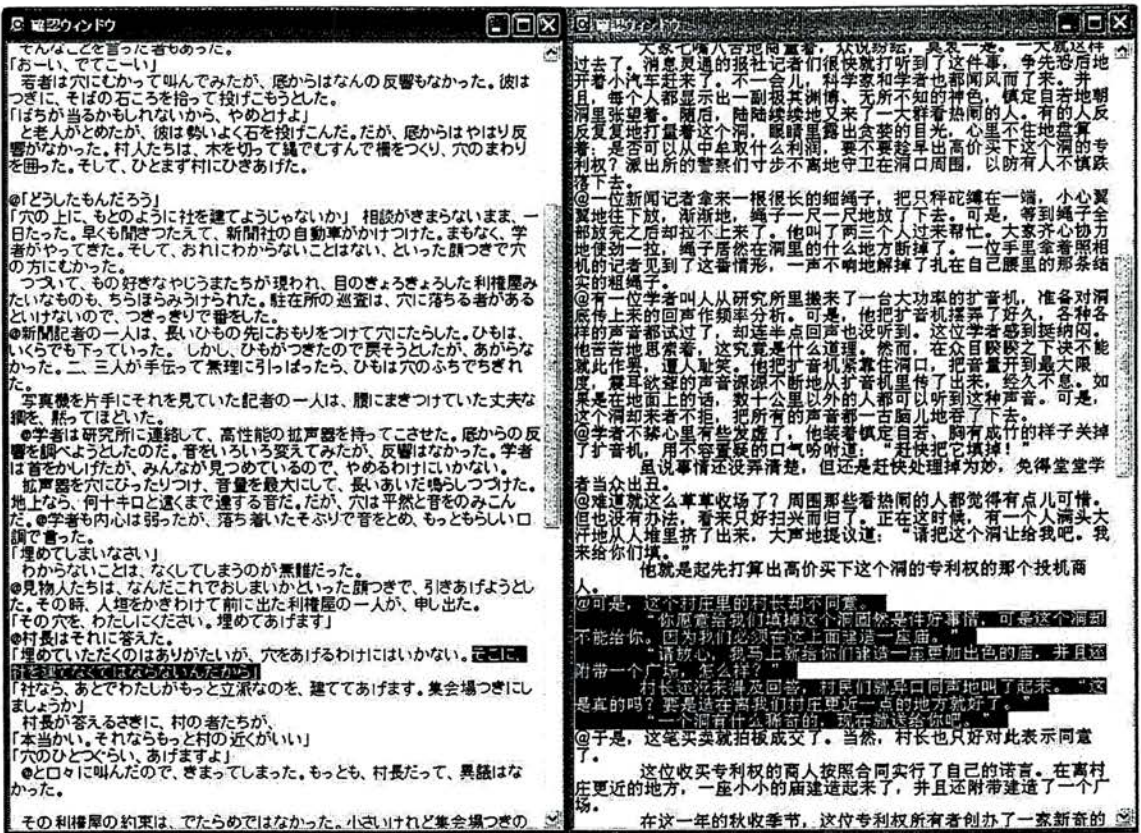


図3-4. 日中対照表示

3. 3 レポートへの出力

必要な例文の選択を終えた検索結果の一覧から、レポート保存ボタンをクリックすることでレポートをファイルとして出力できる。レポートは HTML 形式のファイルとして出力される。(GPSXP では Word に直接挿入したり、Word の Doc 形式に直接出力したりできる。また検索語へのマーキングを行うことも選択できる)

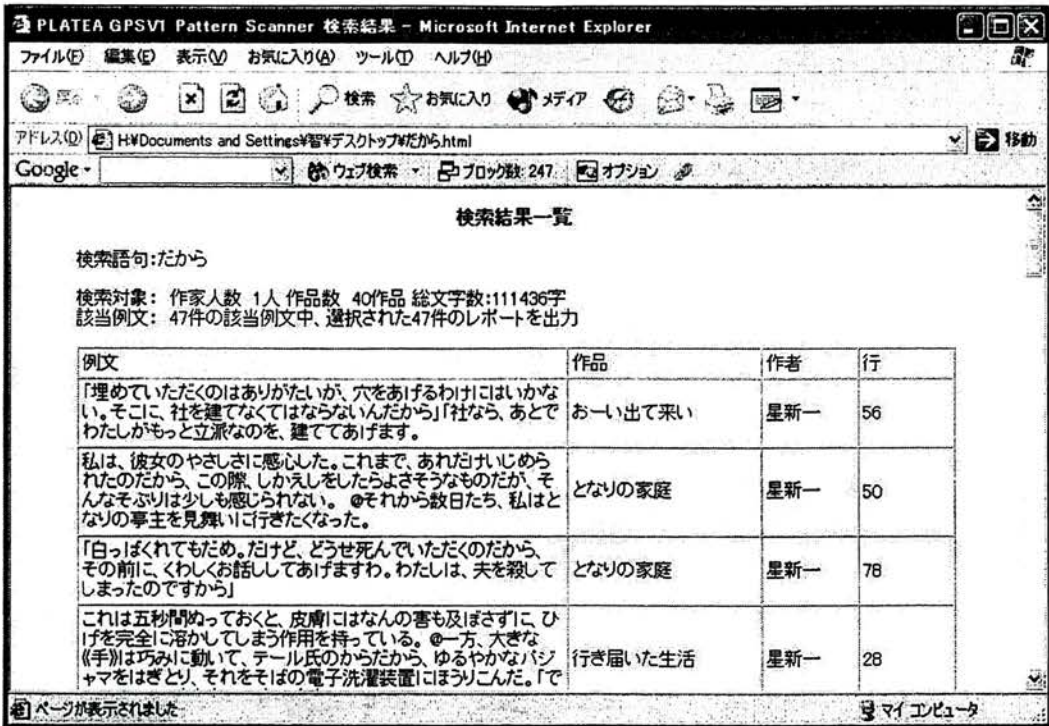


図3-5. レポート出力 (上下1行表示)

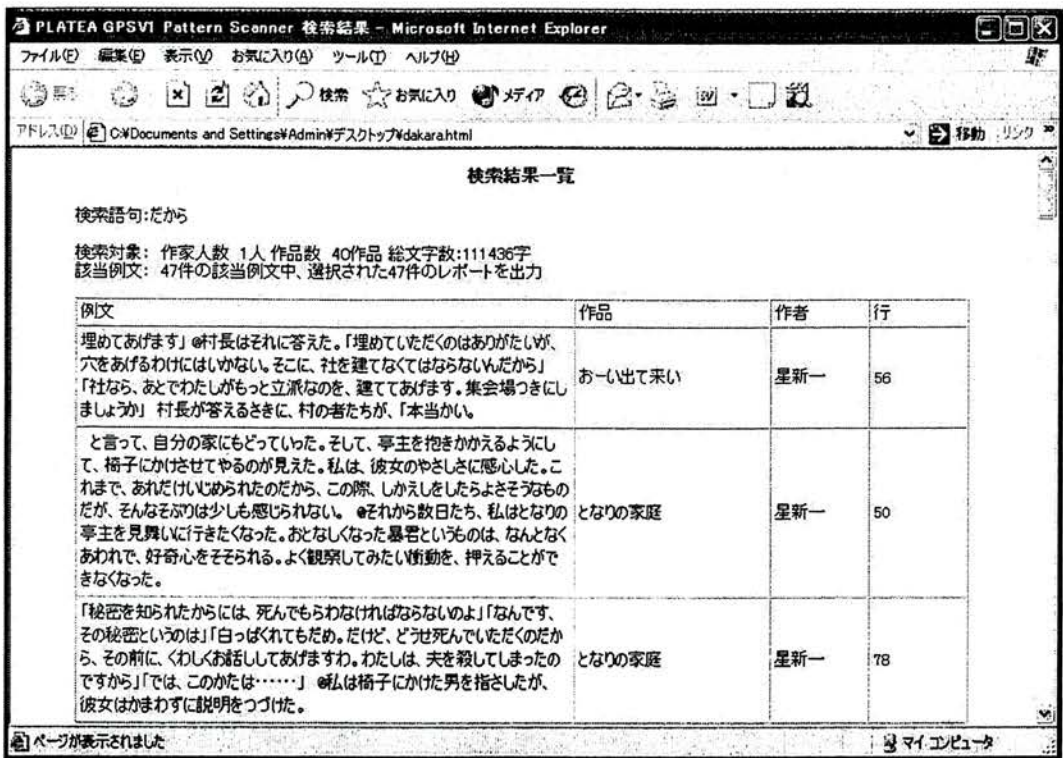


図3-6. レポート出力（上下3行表示）

4. ライブラリ構築の実態

上に述べてきたように、GPSV1 はユーザが任意に拡張できるテキストライブラリの構築とテキスト検索が行える事を特徴としたアプリケーションとして開発されている。各ユーザが自分に必要な作品群をライブラリ化し、それを相互に交換、或いは一ヶ所に集積していく事によりライブラリの容量を無制限に拡張できる。

これまでの運用により、鲁迅、茅盾、张爱玲の作品集などを集めた中国現代文学ライブラリや中国当代文学ライブラリを相当数構築し終えている。当代文学ライブラリについては日本語訳の電子化も同時に行い、最終的には日中対照コーパスとして利用できるライブラリを構築する。また、人民日報の95年以降の全記事ライブラリも構築中であり、近日中に稼動が可能になる予定である。現段階で即座に利用可能なライブラリの容量は、作家数約50人、作品数約500件、総文字数500万字弱となっており、準備段階のテキストを含めるとおよそ2～3倍になる事が予想される。

また、統計分析モジュールの組み込みにより語句の出現率などをグラフ化、Excelに統計情報を挿入する機能を持たせた「GPS」（別の単体ソフトとして開発済み）と統合し、ネットワークとの連携機能を新たに実装する事により、より総合的なアプリケーションとして進化させることも設計上は可能である。

5. アプリケーション仕様

OS	Windows 98/NT4.0 以降 各国語版に対応
メモリ容量	64MB 以上 (128MB 推奨)
記憶容量	アプリケーション本体は3MB 程度 他に必要な容量はライブラリの大きさによる
ライブラリの制限数	無制限 (メディアの空き記憶容量に従う)
ライブラリの制限容量	作家数・作品数・ファイル数は無制限
読み込み可能な文字コード	SHIFT-JIS GB2312 UTF-8 UTF-16
読み込み可能なファイル形式	テキスト形式 HTML 形式
レポート出力形式	HTML 形式

あとがき

本報告書を2年間でまとめる時期にあたり、研究代表者の大瀧幸子には交通事故という不慮の災難が生じ、研究協力者の王亜新には海外研究を実りあるものにするための幸運ではあるが非日常的な緊張が生じた。

研究の目的としたテキスト分析に関する立論へは、両者とも到達できたとはいいいがたい。しかしながら、根気よく基礎的な資料の作成にあたり、否応なしに豊富な言語事実を捉える機会を得ることができたのは、科研費を受領できた恩恵といえる。我々が作成したデータベースそのものをこの報告書で公開することは、労多くして益のないことであり、むしろ、ちかぢか独自のサーバーを備えて「用例検索」を希望する人々とともにデータを拡張していけることを希望している。我々の細かい要求に辛抱強く対応し、かつ必要な機材についても労を厭わずアドバイスしてくれた林智氏（金沢大学4年生）に、今後とも協力を仰ぐことができれば幸甚である。

本研究の発展をめざして、大瀧は金沢大学の同僚であるフランス語学、ドイツ語学の教員とともに、中国語のほかにフランス語とドイツ語のデータを加えていく予定がある。テキスト分析に用いることのできる、できる限り普遍性のある体系性を有した記述用概念を早急に獲得する必要に迫られているために、本報告書は理論的模索に重点をおいたものになった。また、王はイギリスにおいて主に認知言語学への研鑽を深める予定である。欧米言語を母語とする東洋文化の研究者との交流は、日本語学の分野で学位保持者である王にとっても、その研究者としての視点を大きく豊かなものにする契機になるはずである。

今回は個々に自分の研究テーマを迫及したかたちになり、資料収集や検索ソフト開発という技術的側面での協力体制が主となる共同研究であったが、1年後に再会したときには、名実伴った共同研究を行えるだけの学問的共通基盤を獲得したいと願っている。

最後に、インフォーマントとして協力してくださった方々に心から謝意を表します。

中国側協力者：12名

日本語研究者（8名）

- 翟東娜（北京師範大学日本語学部教授）例文分類と判定及び面接調査
- 宛金章（北京師範大学日本語学部教授）例文分類と判定
- 林 洪（北京師範大学日本語学部助教授）例文分類と判定
- 王志松（北京師範大学日本語学部助教授）例文分類と判定

冷麗敏（北京師範大学日本語学部助教授）例文分類と判定
林 涛（北京師範大学日本語学部助教授）例文分類と判定及び面接調査
姜 弘（北京師範大学日本語学部助教授）例文分類と判定及び面接調査
範曉萍（北京工業大学日本語専任講師）例文の分類と判定

大学院生（2名）

郭亜軍（北京師範大学大学院日本語学専攻修士課程1年）対訳例の統計
孫曉傑（北京師範大学大学院日本語学専攻修士課程1年）対訳例の統計

中国文学研究者（2名）

李偉東（北京師範大学漢語学院助教授）例文分類と判定及び面接調査
朱瑞平（北京師範大学漢語学院教授）例文分類と判定及び面接調査

日本側協力者：9名

橋本幸枝（東洋大学文学部中国語非常勤講師）例文判定とコメント
植松希久磨（東洋大学経営学部中国語非常勤講師）例文判定とコメント
寿老素子（東洋大学社会学部中国語非常勤講師）例文判定とコメント
荒見泰史（東洋大学経営学部中国語非常勤講師）例文判定とコメント
山寺未希子（東洋大学経済学部中国語非常勤講師）例文判定とコメント
飯泉彰裕（東洋大学経済学部中国語非常勤講師）例文判定とコメント

大学生・大学院生（3名）

東 真知子（金沢大学文学部4年生）例文コピーと日本語OCRファイル校正
山田 彩子（金沢大学大学院文学研究科2年）中国語OCRファイル校正
邹 景（金沢大学大学院文学研究科2年）中国語OCRファイル校正

(2004.3.15)