海外のテキスト・アーカイヴにおける管理・運営上の問題点について：アンケート調査報告

国立国語研究所報告107 研究報告集15 (1994)

海外のテキスト・アーカイヴにおける管理・運営上の問題点について

— アンケート調査報告 —

伊 藤 雅 光

ITŌ Masamitsu: Problems of the Administration and Maintenance on Overseas Machine-Readable Text Archives — Report from a Questionnaire Survey

要旨：このアンケート調査は1991年現在における海外のテキスト・アーカイヴの管理・運営状況を明らかにするとともに，その問題点を抽出して，今後のテキスト・アーカイヴ開設の可能性を討議する際の資料を提供する目的で行われた。

主な問題点としては次の諸点が浮び上がってきた。

(1) 著作権をめぐる時間と労力の浪費
(2) 運営資金の不足
(3) 職員不足
(4) 困難さを増してきたテキスト・ファイルの収集と作成
(5) テキスト・ファイルの入力ミスの排除の困難さ

キーワード：テキスト・アーカイヴ，フルテキスト・データベース，テキスト・ファイル，電子化テキスト，著作権，アンケート

Abstract: The National Language Research Institute has instituted a plan to establish a repository or archive of machine-readable texts. For this purpose, the department of data orientation started in April 1991 a preparatory study to survey repositories and archives of machine-readable texts throughout the world and to analyse problems of maintaining them.

The following problems came to light.

(1) Obtaining copyrights from so many different sources has been extremely timeconsuming
(2) Shortage of funds
(3) Lack of staff
(4) Obtaining and preparing texts
(5) Quality of texts

Key words: machine-readable text archives, full-text database, text files, copyright, written questionnaire

目　　次

1. 本調査の概要

1.1 本調査の目的

この調査は平成3年度から開始された国立国語研究所の通常研究「日本語情報資料データベース構築のための準備的研究」（3年計画）の一環として行われたものである。「日本語情報資料データベース」とは日本語に関する研究と教育のために必要な文字情報・音声情報・画像情報を含む資料を収集対象とする総合的なデータベースのことである（文献［1］～［5］）。このデータベースはまだ研究所で構築するという決定には至っていない。この研究はこのデータベースを将来，研究所で構築できるかどうか，あるいは構築できるとしてもどの程度までのものが可能であるかという見通しを立てるための資料を提供することを目的として行われたものである。

日本語情報資料データベースの構成と研究年次との関係は次のようになる。

第1年次（平成3年度）
（1）文字情報資料データベース（文字データ）
第2年次（平成4年度）
（2）音声情報資料データベース（音声データ）
（3）画像情報資料データベース（画像データ）
a．静止画像情報資料データベース（静止画像データ）
第3年次（平成5年度）
b．動画像情報資料データベース（動画像データ）
（4）複合情報資料データベース
（文字データ＋音声データ＋画像データ）

この調査は第1年次の「文字情報資料データベース」に関する調査として行われたものである。

なお，アンケート用紙の発送時期は平成3年（1991年）11月～4年2月である。

1.2 「テキスト・アーカイヴ（Text Archive）」という用語について

本報告の標題中にある「テキスト・アーカイヴ」という用語はテキスト・データベースを活用している研究者以外の方にはまだ耳新しい用語かと思われる。「テキスト・アーカイヴ」とは簡単にいうと「電子化テキストを収集・管理しているところ」のことである。世界的に有名なテキスト・アーカイヴとしては「オックスフォード・テキスト・アーカイヴ（O.T.A.; Oxford Text Archive）」が挙げられる（文献［6］）。O.T.Aの主な活動は「電子化テキストの収集・作成・交換・提供（有料）」である。日本語で「テキスト・アーカイヴ」を「文書局」と訳す向きもあるがまだ定着していない。また，「文書局」では誤解を生む可能性が高いと思われるので，この報告ではそのまま「テキスト・アーカイヴ」という用語を使うことにした。

なお，パソコン通信の世界では「アーカイバ（archiver）」という用語がよく使われる．この「アーカイバ」というのは，プログラムの名前で，大容量のファイルや複数のファイルを半分くらいの容量に圧縮して，一つのファイルにまとめる機能を持っている．このようにするとより多くのファイルを保管できるし，また，電話回線を使ってファイルを転送する場合の時間も少なくてすむため，パソコン通信では不可欠の道具（ツール）としてよく普及している．そして，このプログラムで圧縮されたファイルのことを「アーカイバル・ファイル（arhcival file）」という．「アーカイバ」，「アーカイバル」，どちらも「アーカイヴ」という単語の派生語なので，本質的には通じるものがあるのだが，実際に指しているものは大分異なっている．蛇足かも知れないが，誤解を恐れてひとこと触れておく．

1.3 アンケートの内容と本報告の記述順序について

アンケート調査用紙はこの報告の最後に資料2として添付した。内容は大きく2つにわかれ，前半（質問項目G01－G09）は送付先の組織自体に関する一般的な情報に関するもので（General information），後半（質問項目T01－T13）はその機関のテキスト・アーカイヴの管理・運営に関する質問となっている（Text Archive information）。その概要は以下に示した

とおりであるが，本報告の記述の順序は下記のアンケート用紙の順序とは異なっている。その理由はアンケートの最後の項目である「テキスト・アーカイヴの管理・運営上の問題点とその解決策」（T12, T13）の分析の大要をまず示し，それに基づいて重要な問題から記述するようにしたためである。これによりアンケートと本報告の記述との対応関係を示す必要が出てきたため，各章ごとにアンケート調査項目の番号を付すとともに，下記の概要の各項目に本報告の章番号を付した。

G01　公的機関か私的機関か　（2.2）
G02　主な活動（資料篇1）
G03　組織全体の職員数　（6.2）
G04　テキスト・アーカイヴ担当部局の職員数　（6.3）
G05　テキスト・アーカイヴ担当部局内で文献学または書誌学の知識をもつ職員数　（6.4）
G06　テキスト・アーカイヴ担当部局に対する資金的援助の有無（5.3）
G07　援助金を受けている場合の出資元(政府・財団・その他）　（5.4）
G08　1991年度の援助額とその出資元　（5.5）
G09　1990年度までの援助額とその出資元　（5.5）

T01　テキスト・ファイルの収集法　（7.5）
T02　テキスト・ファイルの作成法　（7.7）
T03　ＯＣＲの機種名　（7.8）
T04　収集管理しているテキスト・ファイルの内容的な種類　（7.2）
T05　収集管理しているテキスト・ファイルの主な言語　（7.3）
T06　収集するテキスト・ファイルの受入れを決定するのは誰か（7.6）
T07　収集したテキスト・ファイルのどこをチェックするか　（8.2）
T08　収集したテキスト・ファイルの入力ミスを訂正するために，そのテキスト・ファイルの元となった底本を収集しているか。また，どの

くらい収集しているか。　(8.3)

T09　TEIを知っているか　(9.1)

T10　TEIを将来受入れるか　(9.2)

T11　もしTEIを受入れないとしたら，その理由は何か　(9.3)

T12　テキスト・アーカイヴの管理・運営上の問題点（3～8）

T13　テキスト・アーカイヴの管理・運営上の問題点の解決策（4～8）

2.　調査対象機関の概要

アンケートの対象機関の選定にあたってはつぎの2つの資料を参考にした。

a.『オックスフォード・テキスト・アーカイヴ目録（簡略版）』1988 オックスフォード大学コンピューティング・サービス編（文献［7］）

b.『言語工学関係機関一覧　1990年版』1990, 言語産業調査会編, 欧州共同体（EC）委員会発行, フロッピー版　（文献［8］）

この中からテキスト・アーカイヴを開設していると思われる122機関を選びアンケート調査用紙を送付した。

2.1　国別アンケート調査対象機関数

表1にあるように，地域別では欧州が89（72.95%）機関，北米地域が28（22.95%）機関と，この2地域で95%を占めているが，これは故意に地域を限定したわけではなく，それだけ他の地域の機関の情報が収集できなかったというだけのことである。

表1

（1）欧州　　89（72.95%）

（2）北米　　28（22.95%）

（3）中近東　　3（ 2.46%）

（4）アジア　　2（ 1.64%）

総数　　122機関

国別ではつぎの表2のようになる。多い国から並べるとイギリス（24），アメリカ合衆国（18），フランス（11），カナダ（10）というように英語圏とフランス語圏とが上位を占めている。

表2

（1）欧州 89			
アイスランド	1	スペイン	3
イギリス	24	デンマーク	4
イタリア	5	ノルウェー	4
オーストリア	2	ハンガリー	4
オランダ	7	フランス	11
ギリシャ	1	ブルガリア	1
スイス	1	ベルギー	5
スウェーデン	3	ロシア	1
（2）北米 28			
アメリカ合衆国	18	カナダ	10
（3）中近東 3			
イスラエル	3		
（4）アジア 2			
インド	1	韓国	1

以上がアンケート用紙を送付した国々の概要である。

2.2 アンケートの回答があった機関 （G01－2）

アンケートの回答があったのは122機関中，23機関（18.85%）であった。また，その中には返事はあったが，自機関が今回のアンケートの対象とはならない旨を断ってきたところも4機関あり，結局，有効回答は19機関ということになった。表3はその国別の内訳であるが，表2の結果と比べるとスウェーデンとカナダの回答率の高さが目立つ。また，設立母体別（表4）では「私立」が，所属機関別（表5）では「大学」が大半を占めている。

表3　国別回答状況

有効回答　19

（1）欧州　11　（イギリス　4, イタリア　1, オランダ　1,
スウェーデン　3, ドイツ　1, フランス　1）

（2）北米　6　（アメリカ　2, カナダ　4）

（3）中近東　1　（イスラエル　1）

（4）アジア　1　（韓国　1）

無効回答　4

（1）欧州　1　（イギリス　1）

（2）北米　3　（アメリカ　2, カナダ　1）

表4　設立母体別

公立　5

私立　13

その他　1（Educational Trust – メリーランド大学）

表5　所属機関別

大学	14（University 11, College 2, Academy 1）
研究所	2
政府直属機関	2（ケベック州政府フランス語局, ケベック州政府教育省）
私企業	1（ダゴスティーニ協会）

3.　テキスト・アーカイヴを管理・運営していく上での問題点の概要（T12）

「テキスト・アーカイヴを運営していく上での問題点」はアンケート用紙では最後の質問（T12, T13）であったが，これが最も重要なことなので，この問題から取上げることにする。アンケートではこの運営上の問題点を過去・現在・未来にわたって，それぞれで記述するようになっている。また，その解決策についても同様である。寄せられた回答の中で，指摘の多かった問

題点から並べたのが表6である。回答形式が記述式だった点やまた内容自体が答えにくい事柄だったためか,あまり回答率はよくないが,「著作権」「資金不足」「職員不足」が上位を占めている点は予想通りで，これらの点が問題になるのは洋の東西を問わないようである.以下，「著作権，資金，職員，テキストファイルの収集と作成，テキストファイルの質，TEI」について述べていく。

表6

	過去	現在	未来	合計
01 著作権	4	4	2	10
02 資金が足りない	3	2	3	8
03 職員が足りない	2	1	2	5
04 テキストの質	2	1	1	4
05 時間が足りない	1	1	1	3
06 テキストの収集・取得が難しい	2	1	0	3
07 記憶装置の容量が足りない	0	2	0	2
08 出版業者との競争	0	0	1	1
09 標準画像フォーマットがどうなるか，はっきりしない	0	0	1	1
10 デスクフォーマットが異なっている	1	0	0	1
合計	15	12	11	

4. 著作権

4.1 問題と解決策 （T12, T13)

下記のコメントは，アンケートの項目T12（問題点）とT13（解決策）に書かれていたものである。機関名の前に付けられている番号は各機関の機関番号である。アンケート項目T12とT13は過去・現在・未来にわたっての問題点と解決策とを自由に書き込むようにしてあるため，機関によって扱う事柄はまちまちである。下記の機関はたまたま著作権の問題に触れているため

に，ここで取上げられたのである。また，機関によっては多くのコメントを書いたところもあったが，ここでは著作権に関する部分だけを示し，それ以外の事柄に関する部分は別の節で示している。次節以下でも同様である。

9　ストックホルム大学　スウェーデン　（過去・現在）

問題点：　We are building a balanced corpus with the same structure as Brown and LOB. Getting texts, and in particular getting copy-rights, from so many different sources has been extremely time-consuming.

（ブラウン・コーパスやLOBと同じ構造の安定したコーパスを構築しています。かなり多くの版権所持者から，テキストを得ること，とりわけ著作権を得ることにとんでもなく時間がかかっています。）

解決策：無回答

（注）“LOB”とは“Lancaster-Oslo-Bergen corpus”のことである。

12　ダートマス大学　アメリカ　（過去・現在・未来）

問題点：Quality + Copyright（質　＋　著作権）

解決策：None so far（今までのところなし）

13　ジョージタウン　アメリカ

a．（過去・現在）

問題点：Copyright permissions from publishers and authors

（出版者や作者からの著作権使用許可を得ること。）

解決策：Royalty payments on copies sold

（市販のテキストに対する印税の支払）

b．（未来）

問題点：Funding. Copyright. Competition from publishers.

（資金。著作権。出版者から仕掛けられる競争）

解決策：More purchases,less creation.

（テキストファイルの購入をより多くし，テキストファイルの作成をより少なくする）

14　ケベック州政府フランス語局　カナダ

問題点（過去・現在）：Acquisition et gestion des droits d'auteur（著作権の取得とその管理）

解決策（現在）：Établissement d'une procédure d'acquisition des droits d'auteur;négociation des droits d'auteur confiée àune personne dans l'organisme; centralisation de toute l'information concernant l'acquisition et la gestion des droits d'auteur.

（著作権の取得手続の確立。著作権の交渉を機関内で個人に依託する。著作権の取得と管理についてのすべての情報の集中化）

4.2　分析

この著作権の問題はどこの機関でも一番頭の痛い問題のようである。上記の回答から，著作権に関する問題としては（1）著作権保持者から使用許可をとるまでの交渉に時間がかかるということと，（2）著作権を取得したあとの管理が大変であることとが伺える。

回答には「出版者から仕掛けられる競争（13）」というのがあるが，イギリスでも同じようなことが起こっているらしく，まずソフトウェア開発業者がスペルチェッカーなどの開発にテキスト・ファイルが使えることに気づいたため，多くの業者がアーカイヴを開設している機関にその使用を打診してくる。その次に出版者が研究者に対するテキスト・ファイルの販売という重要で新しいマーケットの可能性に気づいたために，テキスト・ファイルを作成している研究者や研究機関を対抗者として見なすようになったといわれて

いる（文献［6］）。

この点は日本も例外ではなく，近年の新聞や辞書のＣＤ化やオンライン・データベース化はいよいよ盛んになってきているし，電子ブックのような日本生れの電子化テキスト読書用機器もその種類を増やしてきている。発売が予定されている電子化テキストのなかには単行本よりも先に供給されるものさえあるという。再版，管理，運搬，保管，どれをとっても電子媒体のほうが紙の印刷媒体よりも容易なうえに経費も安く押さえられる。今後，出版社にとっては電子化テキストはますます重要な商品となっていくはずで，「出版者から仕掛けられる競争」は日本でも激しいものとなっていくであろう。

回答にある著作権の管理については具体的なことは書かれていないが，これは利用者の使用目的の制限と無断コピーの防止のことを指していると推測される。実際，オックスフォード・テキスト・アーカイヴの目録（文献［9］）には次のように使用目的の制限と無断コピーの禁止とが明記されている。

> Except for texts in category P, all users are required to sign a declaration limiting their usage of the texts to private study and research and agreeing not to redistribute the texts without consultation.
>
> （あらゆる利用者は，区分Ｐ以外のテキスト・ファイルの用途を個人的な調査研究に限定し，また，無断で再配付しないという宣言文へのサインを要求される）

このサインでどれだけ効果が上がるかは分からないし，また，調べようもないが，少なくともあとでトラブルが起こったときに著作権法遵守のための努力を行っていたという証拠にはなる。なお，「区分Ｐ」というのはパブリック・ドメイン・テキスト（public domain texts）のことで，このテキストは著作権が放棄されているので，コピーや改造，変更などが自由に行えるものである。利用者の側からいえばこのようなテキストが理想的だが，1992年４月の時点でオックスフォード・テキスト・アーカイヴが保有している約1000タイトル中，区分Ｐのテキストはつぎに示した３タイトルだけである（文

献［9］)。もっともその3年前の1989年10月に発行された目録(文献［10］)では区分Pはまだ設けられていなかったので，1992年の時点では設けられて間もなかったという事情は考慮に入れるべきだが，ともかく，今後，この区分Pのテキストが増加することを期待したい。

P－1054－E　MRC Psycholinguistic database (expanded SOED entries). Depositor: Michael Wilson, Informatics Division, SERC Rutherford Appleton Laboratory. [On RLIN]

P－ 720－E　Oxford advanced learner's dictionary (expanded "Computer Usable" version). Ed. R. Mitton. Depositor:
Roger Mitton, D of Computer Science, Birkbeck College. [On RLIN]

P－1192－E　The CED Prolog Factbase. [On RLIN]

解決策としては13bのジョージタウン大学の「テキストファイルの購入をより多くし，テキストファイルの作成をより少なくする」という案が注目される。つまりどの機関も独自にテキストファイルを作成しようとすると，著作権者との交渉を個別にしていかなければならなくなるが，業者が商品化したテキストファイルであれば，その交渉の時間も，テキストファイルの作成の時間もゼロとなる。とりわけテキストファイルの間違いを訂正するのはかなりの労力と時間を必要とする。この点，商品化されたテキストファイルの校正は完璧になされているのが普通である。資金の面でも，入力や校正に使うアルバイターの謝金一つを考えてみても，市販のテキストファイルを購入した方が却って安くなるであろう。日本ではまだまだ市販のテキストファイルが少ないので，この方法を取るのには限界があるが，将来的にはかなり有望な解決策と判断される。

ただし，この方法は常に販売されているテキストファイルの範囲という限定がつきまとうため，その中に希望するテキストファイルがなければやはり

自分で入力していかざるをえない。そのような場合の解決策としては，やはり13ａのジョージタウン大学が取ってきた，版権者に対する「印税の支払」というのが実際的な方法であろう。

5. 資金

5.1 問題点と解決策 （T12, T13）

資金が足りない旨の回答を寄せた機関はつぎの4機関である。それぞれどのような方面で資金が足りないかを推測するために，他の質問項目に対する回答のなかから「年間予算，担当職員の人数，ファイルの収集法，ファイルの作成法」に関する情報も示した。

9 ストックホルム大学 スウェーデン（未来）

問題点：Money

解決策：無回答

※年間予算：'90, '91（Sek 800,000＝￥1728万）；'89（Sek 700,000）

※担当職員：9人以下

※ファイルの収集法：100%自作

※ファイル作成法：キーボード入力95%以上，OCR 5 %以下

10 ヴェストファーレン ヴィルヘルム大学 ドイツ（過去）

問題点：money－personnel（お金 － 職員の）

解決策：improvise and keep on requests for funding

（間に合わせで現状をしのぎ，資金を要求しつづける）

※年間予算：無回答（政府からの援助と個人からのテキストの寄贈はあるとのこと）

※担当職員：9人以下

※ファイルの収集法：寄贈，購入，自作

※ファイル作成法：無回答

13 ジョージタウン大学 アメリカ

問題点（過去・現在・未来）：Continued funding of editors' time.（編集アルバイターの謝金を供給し続けること。）

解決策（過去）：Support of university administration.（大学経営からの援助）

※年間予算：無回答

※担当職員：9人以下

※ファイルの収集法：購入66%，自作34%

※ファイル作成法：OCR100%

16　ラベル大学　カナダ（過去・現在・未来）

問題点：Money

解決策：Money

※年間予算：無回答

※担当職員：9人以下

※ファイルの収集法：寄贈70%，自作30%

※ファイル作成法：キーボード入力100%

5.2　分析

資金難の理由を示しているのは「10　ヴェストファーレン　ヴィルヘルム大学」と「13　ジョージタウン大学」でいずれも人件費である。上記の4機関のテキスト・アーカイヴ担当職員がみな9人以下と少ないことから，アルバイターや非常勤職員は不可欠と推測される。テキストファイルの作成法も「13　ジョージタウン大学」のようにOCR（Optical Character Reader：光学的文字読取，またはOptical Character Recognition：光学文字認識）装置だけで作成しているところでもアルバイターの謝金がたりないと回答しているわけだから，その大半をキーボード入力で作成しているその他の機関ではなおさらのことであろう。

年間予算では「9　ストックホルム大学」が91年度予算（Sek 800,000＝¥1728万）を示しているが，それではいくらなら充分かという点になると，それぞれアーカイヴの規模や業務内容などにより違ってくるため一概にいえ

ないが，さすがに年間予算が1億円を超える機関（表９，５ダゴスティーニ協会，８王立工学院，19延世大学）からは予算不足を伺わせるような回答は寄せられていない。

5.3　テキスト・アーカイヴを維持していくために資金面で援助を受けているか。（G06）

資金面での解決策としては，資金援助をもらうほかに方法がないが，どれだけの機関が援助を受けているかという状況をテキスト・アーカイヴ担当者数と対照できるようにまとめたのが表７である。援助を受けている機関と受けていない機関とがほぼ半々であることが分かる。ちなみに「5.1」で資金不足の回答をしていた4機関のうち，前２者（機関番号９，10）は資金援助を得ており，その他（13, 16）は援助を得ていない。

表７

テキスト・アーカイヴ担当者数			−9	10−50	51−100	101−	無回答
yes	10	:	6	2	1		1
no	9	:	7	2			

5.4　どこからの援助金か（G07）

「5.3」の質問G06でyesと回答した10機関がどこから援助を得ているかをまとめたのが表８である（複数回答可能）。「政府から」と回答している機関が10機関あるということは，援助を受けているすべての機関が政府から援助を受けていることを意味している。「民間の団体」や「その他」からの援助を受けている機関はかなり少ないことがわかる。「その他」というのは「外国の複数の図書館　“US Libraries(RLG)”」や「個人　“individuals”」からの援助である。なお，質問の意図からは外れるが「個人からのテキスト・ファイルの寄贈　“private donors of text 〈machine readable〉”」を「その他」として回答した機関もあった。

表8

テキスト・アーカイヴ担当者数		−9	10−50	51−100	101−	無回答
政府から	10 :	6	2	1		1
民間の団体から	2 :	1	1（注）			
その他	3 :	3				

（注）この「1」は1機関が2つの民間団体からの援助金を得ている例である。

5.5 テキスト・アーカイヴに対するこれまでの援助金と財源 （G08, 09）

1億円を超える援助をもらっている機関が3機関（5, 8, 19）もあるのが注目される（表9）。なお，「11 歴史文書学院」でもかなり大規模な古文書のテキスト・アーカイヴを計画しており，現在はその目録作成と整理に追われているため，援助金の全額を計算している時間がないという回答があった。「1 ケンブリッジ大学」の毎年£100＝¥22,908というのがどういう援助金で何に使っているのかは，アンケートから伺うことはできない。

以上，援助金は機関によって¥22,908から¥2億822万とかなりの幅があることが分る。できれば，その内訳まで知りたいところであるが，あまり質問項目が込入ってくるとアンケートの回収率が落ちることを予想し，あえて項目として設定しなかった。

表 9

（円換算レートは'91.6.29現在）

援助金取得機関	国	援助金額	財源
1 ケンブリッジ大学	イギリス	'88～'91（各£100＝¥22,908）	無回答
5 ダゴスティーニ協会	イタリア	'90, '91（US$ 1,000,000＝¥1億3875万）	無回答
		'89（900,000）, '88（800,000）	
		'87（700,000）, '86（600,000）	
		'85（500,000）	
7 イエテボリ大学	スウェーデン	'91（Sek 50,000＝¥108万）	政府
8 王立工学院	スウェーデン	'90,'91（Sek 9,640,000＝¥2億822万）	無回答(注1)
9 ストックホルム大学	スウェーデン	'90,'91（Sek 800,000＝¥1728万）	政府 （注2）
		'89（SEK 700,000）	政府
19 延世大学	韓国	'91（US$ 1500,000＝¥2億812万）	政府・財団
		'90（50,000）'89（70,000）	財団
2 エクセター大学	イギリス	None	None
11 歴史文書学院	フランス	'61～'91（回答困難）	政府

※回答不可能・・・メリーランド大学（イギリス），王立工学院（スウェーデン）

※知らない（don't know）・・・オックスフォード大学（イギリス）

上記以外の機関は無回答。

（注1）ここの予算はテキストアーカイヴ関係だけではなく，Speech Communication 関係全体の援助金（文献［11］）

（注2）We are building a corpus of modern Swedish and have received research grants for collecting and analyzing texts. When the corpus is ready, we will get more money.

（我々は現在，現代スウェーデン語のコーパスを作成しておりまして，テキストを収集・分析するための研究援助金を得ました。コーパスの準備が完了すれば，もっと資金をもらえるでしょう。）

6. 職員

6.1 問題点と解決策 (T12, 13)

職員の問題について回答を寄せたのはつぎの4機関である。

1 ケンブリッジ大学 イギリス (過去・現在・未来)

問題点： Lack of staff, time (スタッフと時間が足りない)

解決策： 無回答

10 ヴェストファーレン ヴィルヘルム大学 ドイツ (過去)

問題点：money − personnel (お金 − 職員の)

解決策：improvise and keep on requests for funding
(間に合わせで現状をしのぎ，資金を要求しつづける)

13 ジョージタウン大学 アメリカ

問題点 (過去・現在・未来)：Continued funding of editors' time. (編集アルバイターの謝金を供給し続けること。)

解決策 (過去)：Support of university administration. (大学経営からの援助)

19 延世大学 韓国 (未来)

問題点： Trained man power for analyzing the data, and for lexicographical projects
(データの分析や辞書編纂プロジェクトのために訓練された人)

解決策： Special seminars. Establishing lectures on computational lexicography, language information. Interns in computational linguistics and language engineering.
(特別のゼミナール。コンピュータ辞書学や言語情報に関する講義を確立すること。コンピュータ言語学や言語工学のインターン)

この問題は前節の「資金」の問題とも直接関係があるので，前節で資金不

足を回答した機関もほぼすべて職員不足の問題を抱えていると考えてよさそうである。上記の4機関の回答を比べて興味深いのは，「1,10,13」の機関が職員不足という職員数だけが問題となっているのに対し,「19　延世大学」だけは職員の質に関する問題だけで，職員数に関しての記述がない。表9でも見たとおり，延世大学では年間予算が3億円近くあるため，職員数の問題は存在しないのであろう。

職員不足の解決策は前節の解決策と同じになる。また，職員の質の解決策は職員や職員候補者に対する教育ということになる。国内に適当な職員がいなければ韓国以外から招くこともできるかと思うが，テキスト・ファイルの対象言語が韓国語だけのためそれも難しいのであろうか。

6.2　各機関の構成スタッフの人数　（G03）

つぎの表10は機関全体のスタッフの数に関する質問の結果である。

表10

01	10人未満	1	05	501～1,000人	3
02	10～ 50人	7	06	1,001～5,000人	2
03	51～200人	1	07	5,001人以上	4
04	201～500人	1			

「10～50人」が一番多いがこれにはアンケートの質問文の欠点が反映しているものと推定される。つまり，「機関」の範囲が明確でないため，ある回答者は「大学全体」の人数を答えたのに対し，別の回答者は「大学付属のコンピュータセンター」などの部所属の人数を答えたものと推定される。職員が「10～50人」の大学というのは考えにくい。また，「01　10人未満」の1機関はケンブリッジ大学だが，次の質問事項である「テキスト・アーカイヴ担当者の人数」も「10人未満」と答えている。

6.3　テキスト・アーカイヴ担当部局の人数　（G04）

表11は表10の機関の総人数とテキスト・アーカイヴ担当部局の人数との対照表である。

表11

機関の総人数→	1－9	10－50	51－200	201－500	501－1000	1001－5000	5001－	
担当部局の人数								
01 10人未満 13：	1	4	1		3	1	3	
02 10～ 50人 4：		2		1		1		
03 51～100人 1：							1	（注1）
04 100人以上 0								
無回答 1：		1						

（注1） 歴史文書学院（フランス）

機関の大小に関わらず担当部局の人数は「10人未満」が一番多く，意外に小人数で運営していることが分かる。このため「職員が足りない」という声が聞えてくるのであろう。また，機関の総職員数が5000以上，テキスト・アーカイヴ担当の職員数が51～100人という「歴史文書学院」はかなり大規模な組織であることが分かる。しかも，対象としているテキストは古文書に限られており，フランスの自国文化の研究に対する熱の入れようがわかる。

6.4 テキスト・アーカイヴ部局内で文献学，または書誌学を学んだことのあるスタッフの人数 （G05）

この質問項目は筆者がオックスフォード大学コンピューティングセンターを訪れたときの経験から敢えて設けた項目である。というのは，筆者は5年前，オックスフォード・テキスト・アーカイヴ（OTA）のテキストファイル数種を入手したが，ファイルにはそのテキストの底本に関する情報が明記されていなかったので，直接担当者に会って質問をしたことがある。その時，その担当者はいろいろ調べてくれたのだが，結局分からないと言ったのである。これがいかに重要な問題であるかということは，少しでも文献学を勉強したものであれば，すぐ分かることである。つまり，文献学の初歩の初歩の

段階で大きな躓きをしているのである。恐らく，コンピューティングセンターということで理工系出身の職員がほとんどなのであろうが，文献学を学んだ職員がいれば，こういうことは起こらなかったはずである。OTAといえば世界的に有名なアーカイヴであるが，そこでこのような管理・運営がなされていることは意外であった。それだけに他のアーカイヴの状態を知りたいと思いこの項目を設けたのである。

表12

テキスト・アーカイヴ担当者数→		−9	10−50	51−100	101−	無回答
該当者数↓						
01	0	3：	3			
02	1〜5	11：	9	2		
03	6人以上	3：	2	1		
	無回答	2：	1			1

表12から，文献学を学んだことのある職員がゼロという機関が3機関というのは予想したよりも少ない。そのような職員を置いている機関では1人から5人というところが一番多く，11機関に昇る。やはり，アーカイヴ担当職員が多い機関ほど文献学の知識をもつ職員が多い傾向が認められる。

7. テキストの収集と作成

7.1 問題点と解決策 (T12, 13)

テキストの収集と作成に関する問題点を挙げてきたのは，つぎの4機関である。

7 イエテボリ大学 スウェーデン (現在)

問題点： More difficult to obtain texts while organisations have found that they have a commercial value (spelling−checkers etc.)

（テキストが商業的価値，例えばスペリング－チェッカーなど，を持っていることを諸機関が知ったため，依然よりもテキストを得ることが難しくなった）

解決策：無回答

9　ストックホルム大学　スウェーデン　（過去）

問題点：　We are building a balanced corpus with the same structure as Brown and LOB. Getting texts, and in particular getting copy-rights, from so many different sources has been extremely time-consuming.

（ブラウン・コーパスやLOBと同じ構造の安定したコーパスを構築しています。かなり多くの版権所持者から，テキストを得ること，とりわけ著作権を得ることにとんでもなく時間がかかっています。）

解決策：　無回答

（注）“LOB”とは“Lancaster-Oslo-Bergen corpus”のことである。

13　ジョージタウン大学　アメリカ　（未来）

問題点：　Funding. Copyright. Competition from publishers.

（資金。著作権。出版社から仕掛けられる競争）

解決策：　More purchases,less creation.

（テキストファイルの購入をより多くし，テキストファイルの作成をより少なくする。）

19　延世大学　韓国（過去）

問題点：　Collecting representative texts

（代表的なテキストの収集）

解決策：　Social survey of reading habits of average adults

（平均的成人を対象とする読書習慣についての社会的展望）

テキストの収集と作成は「著作権」と関係の深い問題である。欧米の3機関（7，9，13）ではテキスト・ファイルが以前と比べて収集や作成がしにくくなったと答えている。その理由はテキスト・ファイルが商業的な価値があるということが広く知られるようになったためである。従来はただ同然で収集・作成できたものが，いちいち版権所持者から許可をとったり，対価を払ったりしなければならなくなったわけである。最後の「19　延世大学」の回答からは，韓国語の「代表的なテキスト」とは何かという問題が浮上していたことがわかる。解決策で言っている「平均的成人を対象とする読書習慣についての社会的展望」というのは，その代表的なテキストを決めるための調査の必要性を説いているのであろう。欧米諸国と違って著作権の問題が指摘されていないのが興味深い。韓国では著作権の問題がどうなっているかという点についての補充調査が必要であろう。

7.2　収集・管理しているテキスト・ファイルの種類　（T04）

表13はどのような種類のテキスト・ファイルを収集しているかという質問（T04）の回答をまとめたものである。この表からは「文学，辞書，新聞」を扱っている機関が多いことが分かる。なかには「パンフレット」や「試験問題」まで収集・管理している機関もあり，その多彩さには驚かされる。

表13　選択項目（複数回答可）と各項目の被選択回数

01　あるゆる種類	5	05　学術論文	5
02　文学	8	06　教科書	5
03　新聞	7	07　その他	8
04　辞書	8		

（注）「その他」は具体的にはつぎのようなものを指す。
学術書，雑誌，パンフレット，カタログ，試験，特定の研究プロジェクトに関係する資料など

表14は各機関ごとの回答状況で，表15はテキスト・ファイルの種類数と機関数との関係をまとめたものである。

ほとんどの機関が複数の種類のテキスト・ファイルを扱っており，あらゆる種類のテキスト・ファイルを扱っているところは5機関ある。その一方，1種類だけというところも3機関ある。これらの数字に関しては，どのような目的でテキスト・ファイルを収集・作成しているかという点と関わりが深いのだが，あいにくその点はアンケートの調査項目に入っていない。それに代る調査項目としては「G02　機関の主な活動」があるが，必ずしもアーカイヴに関係ある活動を記述しているわけではない。例えば，「6　ライデン大学」は辞書のテキスト・ファイルだけを集めているので，その目的はかなり限定されたものと推定されるが，“G02”の回答では“Education, Research”とあるばかりで参考にはならない。同様の回答は「2　エクセター大学」も寄せている。ただし，「3　メリーランド大学」が教科書のテキスト・ファイルだけを集めているのは，授業用の資料のデータベースを提供しているためであることが分かった（“The database is also used to distribute full text resources for class room use.”）。とにかく，特殊図書館ならぬ特殊テキスト・アーカイヴは極めて少ない。

表14　各機関ごとの回答状況

	機関	国	01	02	03	04	05	06	07
1	ケンブリッジ大学	イギリス		02	03			06	
2	エクセター大学	イギリス		02					
3	メリーランド大学	イギリス						06	
4	オックスフォード大学	イギリス	01						
5	ダゴスティーニ協会	イタリア	01						
6	ライデン大学	オランダ				04			
7	イエテボリ大学	スウェーデン		02	03	04	05	06	07 (注1)
8	王立工学院	スウェーデン		02	03	04			
9	ストックホルム大学	スウェーデン		02	03		05		07 (注2)
10	ヴェストファーレン ヴィルヘルム大学	ド イ ツ		02	03	04	05		
11	歴史文書学院	フランス							07 (注3)
12	ダートマス大学	アメリカ		02		04			
13	ジョージタウン大学	アメリカ	01						07 (注4)
14	ケベック州政府 仏語局	カ ナ ダ			(注5)	04	05	06	07 (注6)
15	ケベック州政府 教育省	カ ナ ダ	01						07 (注7)
16	ラベル大学	カ ナ ダ							07 (注8)
17	ケベック大学	カ ナ ダ			03	04			
18	ヘブライ語学院	イスラエル	01						
19	延世大学	韓 国		02	03	04	05	06	07 (注9)

(注1)　se brochure！（パンフレットです！）
(注2)　governmental, administrative magazines
（政府発行雑誌，行政関係雑誌）
(注3)　catalogs of manuscripts（写本のカタログ）
(注4)　Philosophy, Art Commentary（哲学や芸術の解説書）
(注5)　Il s'agit plus particulièrement de fichiers de donnèes terminologiques ou documentaires.
（特に専門用語資料または参考資料のファイルに関してより多くの問題がある。）
(注6)　Documents spécialisés dans un des domaines où il y a des projets terminologiques en cours.
（講義に関する専門用語のプロジェクトの分野に限定された資料）
(注7)　Guidelines, Tests（教育指導書，試験）
(注8)　First－hand data collected by our field workers.
（実地調査員によって収集された源データ）
(注9)　Scholarly writings in social, cultural, hitorical studies, sports and recreation.
（社会的，文化的，歴史的研究，やスポーツ，レクリエーションに関する学術的著作物）

表15　テキスト・ファイルの種類数と機関数との関係

種類数	all	6	5	4	3	2	1
機関数	5	2	0	3	2	2	5

7.3　テキスト・ファイルの言語　（T05）

表16は収集・管理しているテキスト・ファイルの言語の種類に関する質問（T05）をまとめたものである。この表から英語と仏語のテキスト・ファイルを扱っている機関が多いことがわかるが，これはアンケートを返送してきた機関のなかで英語圏や仏語圏の国の機関が多かったことと対応している（英語圏　イギリス　4，アメリカ　2，仏語圏　フランス　1，カナダ・ケベック州　4）。

表17は言語の種類の数と機関数とをまとめたものだが，この表からは72%の機関が1種類の言語のファイルしか収集していないことが分かり，一般的には言語種はかなり限定されているという傾向が見られる。

表18は収集しているテキスト・ファイルの言語種と各国の公用語との関係をまとめたものだが，公用語だけの機関が67%，公用語と非公用語の機関が28%と，公用語のファイルだけを収集・作成している機関がほとんどであることがわかる。また，なかには「1　ケンブリッジ大学」（表19）のように非公用語（仏語）だけを扱っている機関もある。以上の結果は，公用語を対象とする研究が一番やりやすいということを考えれば，当然のことである。

表16　選択項目（複数回答可）と被選択回数

01　英語	8	05　スペイン語	2
02　仏語	5	06　ギリシャ語	2
03　スウェーデン語	3	07　ヘブライ語	2
04　独語	2		

その他　（イタリア語・オランダ語・ロシア語・ラテン語・
アラビア語・韓国語・ＬＡＯ）

表17　言語の種類の数と機関数との関係

言語の種類	7	4	2	1	?
機関数	1	2	1	13	1

表18　収集しているテキスト・ファイルの言語種と公用語との関係

関係	公用語だけ	公用語と非公用語	非公用語だけ
機関数	12	5	1

表19　各機関ごとの回答状況

			01 英	02 仏	03 スェーデン	04 独	05 西	06 ギリシャ	07 ヘブライ	その他
1	ケンブリッジ大学	イギリス		02						
3	メリーランド大学	イギリス	01							
4	オックスフォード大学	イギリス	01 (注1)							
5	ダゴスティーニ協会	イタリア								全欧州語
6	ライデン大学	オランダ								Dutch
7	イエテボリ大学	スウェーデン			03					
8	王立工学院	スウェーデン			03					
9	ストックホルム大学	スウェーデン			03					
10	ヴェストファーレン ヴィルヘルム大学	ド イ ツ	01			04	05	06		
11	歴史文書学院	フランス	01	02				06	07	Latin,Russian Arabic etc
12	ダートマス大学	アメリカ	01							
13	ジョージタウン大学	アメリカ	01			04				
14	ケベック州政府 フランス語局	カナダ		02						
15	ケベック州政府 教育省	カナダ		02						
16	ラベル大学	カナダ	01	02			05			LAO (Solomon Islands)
17	ケベック大学	カナダ	01							
18	ヘブライ語学院	イスラエル							07	Hebrew only
19	延世大学	韓　国								Korean only

(注1) but not exclusively (しかし，排他的ではない。)

(注2) エクセター大学　イギリス　Texts used in the past are no longer held, thus not applicable.
(かつて使っていたテキストはもはや保管していない。このように我々は当てはまらない。)

7.4　管理しているテキスト・ファイルの量

管理しているテキスト・ファイルの量についての質問項目はないのだが，回答を寄せた機関のなかにはその量を手紙や注記のかたちで明示したり，あるいは同封された機関紹介パンフレットなどに明記されている場合があったので，それをまとめてみたのが表20である。量の示しかたとしては，「のべ語数，タイトル数，記憶媒体の容量」の三つにわたっているため，すべてを同じ基準で比較することができない。できれば，この三つの量の示し方すべてについて明示してほしかったが，これも補充調査の1項目とすべきであろう。現在，国立国語研究所が保有しているテキスト・ファイルはのべ語数で約600万語であるが，それに比べると「7　イエテボリ大学」の3000万語，「19　延世大学」の2000万語などかなり膨大な量のファイルを管理していることがわかる。しかし，この量と年間予算とは必ずしも比例していない。たとえば，イエテボリ大学と延世大学の保有量は大差ないが，'91年度予算は前者がSek 50,000＝¥108万で，後者はUS＄ 1500,000＝¥2億812万と雲泥の差がある。これは蓄積年数の違いが反映したものと思われる。つまりイエテボリ大学は少ない予算の中で少量のテキスト・ファイルを長期間にわたって蓄積しつづけたのに対し，延世大学は豊富な予算で短期間のうちに大量に蓄積したのであろう。この蓄積年数という点も補充調査項目に加える必要がある。

タイトル数からいえば「3. メリーランド大学」の1,250が「4. オックスフォード大学」の875を上回っているが，オックスフォード大学がフルテキストのファイルを扱っているのに対し，メリーランド大学はパンフレットも含んでいるため，メリーランド大学の方が扱っている量が多いとは単純にいうことはできない。「のべ語数，タイトル数，記憶媒体の容量」の3つの情報がほしいというゆえんである。

「11. 歴史文書学院」のタイトル数が42,000とあるが，同機関ではまだテキスト・ファイル化はしておらず，目録を整理中ということであるが，このタイトル数はオックスフォード大学のものを遥かに凌駕しており，今後の動

向が注目される。

表20

		のべ語数	タイトル数	容量（Mb）
3.メリーランド大学	イギリス		1,250（注1）	
4.オックスフォード大学	イギリス		875	1,058(注2)
7.イエテボリ大学	スウェーデン	3000万（注3）		
8.王立工学院	スウェーデン	4.2万		
11.歴史文書学院	フランス		42,000（注4）	
19.延世大学	韓　国	2000万		

（注1）パンフレットも含む。（文献［12］）

（注2）文献［6］

（注3）文献［13］

（注4）　16世紀のテキストの目録情報データベースのタイトル数。テキストアーカイヴはまだ作成していない。図書館の蔵書数　80.000冊，マイクロフィルム図書館の所蔵数　55.000巻，写真図書館のミニアチュール所蔵数　30.000枚。これらは将来機械可読データにしていくとのこと。(文献[14]）

The I. R. H. T. has not yet enterd in database the machine－readable texts, because we have too much work （establishment of the repertories and catalogs of manuscripts and of printed books of Europe, of Near Orient and of North Africa） and because of many languages treated （ancient and modern）. But one day we will have this opportunity, I hope.

（我々I. R. H. T. はまだ機械可読化テキストのデータベース化に取組んでいません。というのは，我々はあまりにも多くの仕事を抱えているからです。その仕事というのはヨーロッパや近東や北アフリカといった所の写本や活字本の所蔵庫とカタログを整備することです。また，古代語から現代語にわたる多くの言語を扱っているという点もデータベース化を遅らせている原因の一つです。ただし，何時の日かデータベース化の機会が持てるでしょうし，また個人的にもそうなることを望んでいます）　※アンケートの書込み

7.5 テキスト・ファイルの収集法に関する状況 （T01）

表21は各機関がどのようにしてテキスト・ファイルを収集・蓄積しているかを調査した結果である。この結果をさらにまとめると次のようになる。寄贈 21，購入 9，作成 16。こうして見ると「寄贈」という方法が一番多く取られており，「作成」がそれに続く。現在「購入」が一番少ないが，この方法は著作権・資金・時間・職員などの問題をかなり解決してくれるであろうことは本稿の冒頭で述べた。将来はこの「購入」という方法が盛んになっていくと予想される。

表21 選択項目（複数回答可）と各項目の被選択回数

01	個人研究者からの寄贈	8
02	主要な研究プロジェクトからの寄贈	4
03	他の機関からの寄贈	9
04	個人研究者からの購入	1
05	主要な研究プロジェクトからの購入	1
06	他の機関からの購入	7
07	自機関での作成	16
08	その他	0

表21はただ単にどのような方法を取っているかを示しているだけなので，実際のテキスト・ファイルの量との関係は不明である。そこで各機関が何種類の方法をとっており，それぞれの方法でどの位のファイルを蓄積しているかという状況をまとめたのが表22である。%ではなく，ただ質問項目の番号だけあげているのは，%の回答がなかったことを示している。%を回答した機関は13（72.22%）で，選択項目のチェックだけをした機関は5（27.78%）であった。

表22　各機関ごとの回答状況１

機関	国	寄贈 (01	02	03)	購入 (04	05	06)	作成 (07)
1. ケンブリッジ大学	イギリス	10%		10%				80%
2. エクセター大学	イギリス	01						07
3. メリーランド大学	イギリス			20%				80%
4. オックスフォード大学	イギリス	01	02	03				07
5. ダゴスティーニ協会	イタリア						100%	
6. ライデン大学	オランダ							100%
7. イエテボリ大学	スウェーデン			75%			15%	10%
8. 王立工学院	スウェーデン	75%						25%
9. ストックホルム大学	スウェーデン	(注1)						100%
10. ヴェストファーレン ヴィルヘルム大学	ド イ ツ		02	03			06	07
11. 歴史文書学院	フランス	5%						95%
12. ダートマス大学	アメリカ	01	02				06	07
13. ジョージタウン大学	アメリカ						66%	34%
14. ケベック州政府 仏語局	カ ナ ダ	01	02	03	04	05	06	07
16. ラベル大学	カ ナ ダ	70%						30%
17. ケベック大学	カ ナ ダ			50%				50%
18. ヘブライ語学院	イスラエル							100%
19. 延世大学（注2）	韓 国			10%				90%

※無回答・・・15. ケベック州政府教育省（カナダ）

（注1）　We have gathered machine－readable texts from many various sources and prepared and standardized them for our own purposes.
（我々は豊富な収集元から機械可読化テキストを集め，準備し，我々自身の目的のために集めたテキストを標準化しました。）

（注2）　So far, we have collected 20 million running words of Korean texts published since 1970.
（これまで1970年以降に出版された韓国語のテキストから2000万語を収集しました。）

表22をもとに収集方法の種類数と機関数との関係を示したのが表23である。2種類の方法をとる機関が44.44%と最も多く，つぎに多い1種類（22.22%）の倍になっている。かなり限定された方法でテキスト・ファイルを蓄積していることが分かる。

表23　収集方法の種類数と機関数との関係

種類数	8	7	6	5	4	3	2	1
機関数	0	1	0	1	2	2	8	4
		5.56%		5.56%	11.11%	11.11%	44.44%	22.22%

さきの表22から%を回答した機関だけを抽出し，さらに収集方法を「寄贈，購入，作成」だけにまとめたのがつぎの表24である。蓄積したファイルの80%以上を「作成」によったという機関は13機関中7機関（53.8%）となっており，さきの表21の結果とは異なって「作成」という方法が主流を占めていることがわかる。(なお，この「作成」の中に数えられた機関のなかには「9. ストックホルム大学」のようにテキスト・ファイルを他の機関から収集したあとで，ストックホルム大学の研究目的に使えるように手を加えたというところも入っている。）その他の方法で50%以上のファイルを収集したという機関は「寄贈」が13機関中4機関（30.8%），「購入」が13機関中3機関（23.1%）である。冒頭の各機関の問題点の解決方法で「市販のテキスト・ファイルの購入」をあげた「13. ジョージタウン大学」ではさすがに「購入」が66%，「作成」が34%と，「購入」への移行が進んでいると解釈される。

表24　各機関ごとの回答状況 2

機関	国	寄贈	購入	作成
1. ケンブリッジ大学	イギリス	20%		80%
3. メリーランド大学	イギリス	20%		80%
5. ダゴスティーニ協会	イタリア		100%	
6. ライデン大学	オランダ			100%
7. イエテボリ大学	スウェーデン	75%	15%	10%
8. 王立工学院	スウェーデン	75%		25%
9. ストックホルム大学	スウェーデン			100%
11. 歴史文書学院	フランス	5%		95%
13. ジョージタウン大学	アメリカ		66%	34%
16. ラベル大学	カナダ	70%		30%
17. ケベック大学	カナダ	50%	50%	
18. ヘブライ語学院	イスラエル			100%
19. 延世大学	韓国	10%		90%

7.6　テキスト・ファイルの受入れ選別担当者　（T06）

収集するテキスト・ファイルの受入れを決定する人は誰かという質問への回答結果をまとめたのがつぎの表25である。

表25　テキスト・ファイルの受入れ選別担当者

01　係となっている職員	11	
02　運営委員会	3	
03　その他	5	
03－1　研究者が個人的に		2
03－2　プロジェクトのリーダー		1
03－3　得られるファイルは何でも受入れている		2

「係となっている職員」というのが11機関と一番多く，つぎの「運営委員会」の3機関とは大きな開きがある。なかには「得られるファイルは何でも受入れている」というのも2機関あり，わりと受け入れはゆるいところが多

いようである。

7.7 テキスト・ファイルの作成方法 （T02）

テキスト・ファイルを作成している機関ではどのような方法で作成しているかという質問（T02）の結果が表26である（複数回答可能）。キーボード入力を採用している機関が14機関，OCR装置，あるいはソフトウェアによる自動読取りを採用している機関が9機関であるが，そのほとんどはキーボード入力との併用となっている。

表26 選択項目（複数回答可）と各項目の被選択回数

01	キーボードを使った入力	14
02	ＯＣＲによる自動読取り	9
03	その他	3

表27は各機関がどの方法でどれくらいのテキスト・ファイルを入力したかという状況を示したものである。この表から作成したファイルの85%以上をキーボード入力によっている機関がほとんどであることが分るが，これはOCRが実用化される前まではキーボード入力以外に方法がなかったということが反映しているためであろう。しかし，中には「13. ジョージタウン大学」のように100%OCRで入力している機関もあり注目される。機関によってはOCRをまったく利用していないところがあるが，これには経済的理由と技術的理由の2つに分れるであろう。技術的理由というのはOCRの文字読取り精度が実用段階に達していないということである。キーボード入力だけの機関のなかにはヘブライ語だけを扱っている「18. ヘブライ語学院」や韓国語だけを扱っている「19. 延世大学」が入っている。もしもヘブライ文字やハングルで入力しているとすれば，OCRの読取精度が原因となっている可能性がある。また，両国語をアルファベットに翻字してテキスト・ファイルを作成しているとしても，もともとヘブライ語や韓国語がアルファベット表記で刊行されることが少ないであろうことを考えれば，この場合もキーボード入力とならざるを得ないはずである。

表27　各機関ごとの回答状況

機　関	国	キーボード 01	OCR 02	その他 03
1. ケンブリッジ大学	イギリス	90%	10%	
2. エクセター大学	イギリス	01（100%？）		
3. メリーランド大学	イギリス	60%	40%	
4. オックスフォード大学	イギリス	01	02	
6. ライデン大学	オランダ	01（100%？）		
7. イエテボリ大学	スウェーデン	85%	15%	
8. 王立工学院	スウェーデン	100%		
9. ストックホルム大学	スウェーデン		5%未満	
10. ヴェストファーレン ヴィルムヘルム大学	ドイツ			03(100%？)
11. 歴史文書学院	フランス	90%	10%	
12. ダートマス大学	アメリカ	01	02	
13. ジョージタウン大学	アメリカ		100%	
14. ケベック州政府 フランス語局	カナダ	97%	1%	2%
15. 教育省	カナダ	100%		
16. ラベル大学	カナダ	100%		
18. ヘブライ語学院	イスラエル	100%		
19. 延世大学	韓国	90%		10%

（注1）無回答・・・17. ケベック大学（カナダ）

（注2）“5. ダゴスティーニ協会”はOCRの項（02）だけにチェックの印を付けていたが表27には含めていない。これは表22，24において100%購入していると回答しているからである。これだけでは矛盾した回答のようであるが，あるいは購入先の作成方法がOCRであるという意味か。

その他の方法をチェックしている機関が3機関あるが，具体的にどのような方法によっているかという点についての回答は得られなかった。ただ「10. ヴェストファーレン　ヴィルヘルム大学」だけは“magnetic tape”という注記が加えられているので，磁気テープによる寄贈や購入の意味と解釈される。また，「19. 延世大学」では「キーボード入力」が90%，「その他」が10%となっているが，この割合は表22で「作成」が90%，「寄贈」が10%という割合とも符合していることから，その他の10%は寄贈を意味している可能性が高い。いずれにしても，作成方法を聞いている質問の意図からは外れていることになる。

全体からいうと英語のテキストを作成している機関でのOCR利用が目立つがこれはアルファベット用OCRがもっとも早く実用段階に達したことが背景にあるためであろう。

7.8　ＯＣＲの機種　（T03）

表28はOCRを活用している機関だけを対象にしたもので，活用しているOCRはどの機種を使っているかを質問した結果である。

“Kurzweil4000” は一番早く実用化されたOCRの一つで，イギリスと北アメリカの多くの研究機関で導入されたと言われるだけあって（文献［15］），確かに一番多く利用されている。1986年の時点では一番優れた機種だったようだが（文献［16］），1992年までに発売された最新機種と比べると処理速度と価格の点で見劣りがするという（文献［15］）。“Mac scanner”は写真や絵といったようなイメージデータはかなり綺麗に読取るらしいが，OCR用のイメージスキャナーとしての精度（むしろソフトウェアが問題か）はそれ程でもないらしい（国立国語研究所外国人研究員　ジョン・フィリップス氏談）。これら以外の機種についての情報がないのでこれ以上の比較はできないが，全体的には多少の優劣はあるにしても，パーソナルコンピュータに見られるような，ある特定の機種が圧倒的なシェアを占めるという状況には至っていないようである。

表28　機種購入状況

機種	機関数と機関名
1. Kurzweil ―――――――――	5
a. Kurzweil 4000・・・	3 ［オックスフォード大学（イギリス），イエテボリ大学（スウェーデン），ジョージタウン大学（アメリカ）］
b. Kurzweil 5100・・・	1 ［ジョージタウン大学（アメリカ）］
c. Kurzweil Plus・・・	1 ［ダートマス大学（アメリカ）］
Acutext	
2. Apple ――――――――――	3
a. Apple・・・・・・・	2 ［ストックホルム大学（スウェーデン），歴史文書学院（フランス）］
b. Mac Scanner・・・・	1 ［ケンブリッジ大学（イギリス）］
3. Omnypage-Typist ――――	2 ［ダゴスティーニ協会（イタリア），ダートマス大学（アメリカ）］
4. Calera Truescan ――――――	1 ［メリーランド大学（イギリス）］
5. Hewlett Packard ――――――	2 ［ケベック州政府教育省（カナダ），ケベック州政府仏語局（カナダ）］
ScanJet　11c	

（注）“2”の“a.Apple”とあるのは，機種ではなく会社名だけの回答であったこと示している。

8.　テキストファイルの質

8.1　問題点と解決策　（T12, 13）

テキストファイルの質に関する問題を回答したのは次の4機関であった。

3　メリーランド大学　イギリス　（過去）

問題点：　Final print version of text is not consistent with machine readable form being offered. The text has been amended by a typesetter at the final stage.

（印刷されたテキストの最終版が，提供された機械可読形式と一致しない。テキストは最終段階で植字工によって修正されている。）

解決策：　Use OCR（注1） and scan final text.

（OCRを使って最終的なテキストを読取ること）

（注1）原文“ICR”

4　オックスフォード大学　イギリス　（過去）

問題点：Recognition（認識）

解決策：無回答

5　ダゴスティーニ協会　イタリア

問題点：無回答

解決策（未来）：“Fast” Optical Reading System

12　ダートマス大学　アメリカ

問題点（過去・現在・未来）：Quality + Copyright（質　＋　著作権）

解決策（過去）：None so far（今までのところなし）

一般的にデータの入力間違いはどこでも問題となっているようだが，その解決策としてはOCRの導入を答えているところが目立つ。これも予算に直結する問題であるが，今後テキスト・ファイルを作成していく機関ではOCRを導入していくところが増加すると予想される。また，過去にはそのOCR自体の認識率が問題であったと「4　オックスフォード大学」が回答しており，現在その点が依然として問題となっている日本と比べるとコンピュータ処理のしやすい文字を持っている国とそうでない国との差をあらためて感じざるをえない。

8.2　新しく受入れるテキスト・ファイルのどのような点をチェックするか（T07）

「6.4」において，筆者はオックスフォード大学コンピューティングセンターにテキスト・ファイルの底本の書誌情報を問い合せに行ったと述べたが，

その原因は，オックスフォード・テキスト・アーカイヴのテキスト・ファイルの中に入力間違いが多いものがあったからである。そのテキスト・ファイルは留学先となった大学院の授業（コンピュータ言語学）で分析対象とされたものであったが，あまりにその間違いが目立つため，分析を行う前にその間違いを発見するプログラムシステムを開発せざるを得なかったということがあったのである（文献［17］，［18］）。このような経緯から，さきの「6.4」での質問やここのテキスト・ファイルの質に関する質問はぜひ聞いておきたかったのである。

チェック項目としては「テキスト・ファイルの入力間違い」がもっとも多く，「テキスト・ファイルの底本の書誌的情報」「テキスト・ファイルの作成方法」と続く（表29）。「入力間違い」のチェックが多いというのは予想どおりだが，「書誌的情報」のチェックをしている機関が「入力間違い」をチェックしている機関の半数というのは意外である。理屈からいえば，非文のような誰でも分る間違いは別として，一般的には「入力間違い」かどうかの判断はそのテキストファイルの底本が手元にないことには不可能であり，そのためにも底本の書誌的情報のチェックは欠かせないはずである。機関のなかには「何もチェックしていない」というところが5機関もあり，その管理体制には疑問を抱かざるをえない。

表29

01	何もチェックしていない。	5
02	テキスト・ファイルの入力間違い	8
03	テキスト・ファイルの底本の書誌的情報	4
04	テキスト・ファイルの作成方法	3
05	その他	2
06	無回答	3

表30

			01	02	03	04	05	06
1	ケンブリッジ大学	イギリス		02	03			
3	メリーランド大学	イギリス		02				
4	オックスフォード大学	イギリス			03	04		
5	ダゴスティーニ協会	イタリア	01					
6	ライデン大学	オランダ	01					
7	イエテボリ大学	スウェーデン	01					
9	ストックホルム大学	スウェーデン		02				
10	ヴェストファーレン ヴィルヘルム大学	ド イ ツ			03			
11	歴史文書学院	フランス					05 (注1)	
12	ダートマス大学	アメリカ	01					
13	ジョージタウン大学	アメリカ		02			05 (注2)	
14	ケベック州政府 フランス語局	カ ナ ダ		02				
15	ケベック州政府 教育省	カ ナ ダ		02				
16	ラベル大学	カ ナ ダ		02		04		
17	ケベック大学	カ ナ ダ	01					
19	延世大学	韓 国		02	03	04		

(無回答)

2 エクセター大学 イギリス,8 王立工学院 スウェーデン,18 ヘブライ語学院 イスラエル

(注1) Our institute treats by computer the bibliographic database. If we do research on a text, we study it profoundly from the point of a scientific criterium.
(当研究所ではコンピュータにより図書目録のデータベースを取扱っています。もしも我々がテキストの研究をするとすれば，全く科学的観点から研究することになります。)

(注2) Documentation

8.3 テキスト・ファイルの底本を収集しているか (T08)

テキスト・ファイルの間違いをチェックして，それを訂正するためにはそのテキスト・ファイルの底本が必要だということは前節でも述べたが，その底本の収集状況をまとめたのが表31である。収集している機関には「9 ストックホルム大学」,「16 ラベル大学」,「19 延世大学」のように管理しているテキスト・ファイルのすべての底本を収集しているところもあり，その管理体制の手堅さを伺うことができる。また，収集していないと回答しているところも3機関あるが，入力間違いの扱いをどのようにしているのか疑問が残る。

表31

(収集している)

1 ケンブリッジ大学 イギリス (20%)
9 ストックホルム大学 スウェーデン (striving for 100%)
16 ラベル大学 カナダ (100%)
19 延世大学 韓国 (100%)
10 ヴェストファーレン ヴィルヘルム大学 ドイツ
14 ケベック州政府フランス語局 カナダ

(収集していない)

3 メリーランド大学 イギリス
4 オックスフォード大学 イギリス
15 ケベック州政府教育省 カナダ

(無回答)

2 エクセター大学 イギリス
5 ダゴスティーニ協会 イタリア
6 ライデン大学 オランダ
7 イエテボリ大学 スウェーデン
8 王立工学院 スウェーデン
11 歴史文書学院 フランス
12 ダートマス大学 アメリカ
17 ケベック大学 カナダ
18 ヘブライ語学院 イスラエル

9. ＴＥＩについて

TEI（Text Encoding Initiative）は，ACH（Association of the Computer and the Humanities），ALLC（Association for the Literary and Linguistic Computing），ACL（Association for Computational Linguistics）の諸学会が中心となって，テキストデータベースの開発と共有のための国際標準規格を提案することを目的にしている研究プロジェクトである。TEI「ガイドライン」の第一草稿（Document: TEI P1）は1990年，第二草稿（Document: TEI P 2）の一部は1992年に出ており，現在，第三草稿（Document: TEI P 3）が作成されている（文献［19］～［24］）。

9.1 ＴＥＩを知っているか。 （T09）

TEIそのものを知っているかどうかという質問への回答をまとめたのが表32である。TEIを知らないという機関が6機関，それもすべて欧米の機関だというのは意外である。

表32

（知っている） 11

1 ケンブリッジ大学 イギリス
2 エクセター大学 イギリス
4 オックスフォード大学 イギリス
7 イエテボリ大学 スウェーデン
9 ストックホルム大学 スウェーデン
10 ヴェストファーレン ヴィルヘルム大学 ドイツ
12 ダートマス大学 アメリカ
13 ジョージタウン大学 アメリカ
16 ラベル大学 カナダ
18 ヘブライ語学院 イスラエル
19 延世大学 韓国

（知らない） 6

3 メリーランド大学 イギリス
6 ライデン大学 オランダ
8 王立工学院 スウェーデン
14 ケベック州政府フランス語局 カナダ
15 ケベック州政府教育省 カナダ
17 ケベック大学 カナダ

（無回答） 2

5 ダゴスティーニ協会 イタリア
11 歴史文書学院 フランス

9.2　もしもＴＥＩを知っているのなら，将来，テキスト・エンコーディングの共通規格として受入れるか。　（T10）

受入れを決定している機関と検討中の機関とが同数であるという点から，まだTEIが全面的な支持を得ていないということが分かる。もっともTEIはまだ最終的な草稿が出来ていないため，その完成を待ってから判断するという機関も多いと思われる。

表33

（受入れる）

4　オックスフォード大学　イギリス
10　ヴェストファーレン ヴィルヘルム大学　ドイツ
7　イエテボリ大学　スウェーデン
13　ジョージタウン大学　アメリカ
9　ストックホルム大学　スウェーデン

（受入れない）

18　ヘブライ語学院　イスラエル

（検討中）

1　ケンブリッジ大学　イギリス
16　ラベル大学　カナダ
2　エクセター大学　イギリス
19　延世大学　韓国
12　ダートマス大学　アメリカ

9.3　もしも受入れないのであれば，その理由は何か。

前節で受入れないと回答している機関は「18　ヘブライ語学院」だけであったが，その理由はすでに独自のエンコーディングの方法で長年にわたってヘブライ語のテキスト・ファイルを作成してきているということである。その他の機関では，エンコーディングにかかる時間やその内容自体に関する問題が指摘されている。「14　ケベック州政府フランス語局」が問題としている“SGML（Standard Generalized Markup Language）”というのは，電子化テキストの文書構造をマークを使って記述するためのメタ言語で，1986年10月に国際標準化機構（ISO）の規格になっている（ISO 8879, 文献［25］，［26］）。このマークアップ言語は本来，オフィス文書，出版・印刷などの目的で作成されたのだが，一般性が高かったため，電子化テキストを記

述する方法として広く利用されるようになった。そこでTEIでも，SGMLを文書構造記述の文法の枠組みとして採用したのである。

13　ジョージタウン大学　アメリカ　（担当者：10人以下）

We will adopt at least a modified version,but encoding & verifying will be more time−consuming than editing text.

（せめて限定版は採用したいと思っていますが，とにかくエンコーディングやその確認の作業が本文ファイルの作成よりも時間がかかるので....。）

14　ケベック州政府フランス語局　カナダ

Nous considérons davantage l'étude de la norme SGML (Standard Generalized Markup Language)

（SGMLの基準のさらなる検討を考慮しています。）

18　ヘブライ語学院　イスラエル

We have developed our own Encoding methods since 1959, designed especially for Hebrew and for our Historical Dictionary Project.

（1959年から我々独自のエンコーディングの方法を開発してきています。その方法はヘブライ語と我々の歴史辞書プロジェクト用に特別に設計したものです。）

10.　補充調査項目

今回のアンケート調査を行ってその結果を分析してみると，調査票自体に行届かない点がいろいろとあることが分かってきた。今後，補充調査をするとすれば，以下のような点を追加する必要があろう。

a．テキストアーカイヴを設けた理由

b．テキストファイルの収集・作成を始めた年

c．収集したテキストファイルの量はどれくらいか。（のべ語数，タイトル数，容量）

d．公開しているか。また，その範囲は。

e．テキストファイル中に入力間違いが発見された時，すでにそのファイルを購入しているユーザーに対するメンテナンスはどうやっているか。

f．新しいテキストファイルを受入れる際にコンピュータウイルスに対するチェックを行っているか。

g．著作権をめぐるトラブルがあったか。その具体例は？

文献

［1］伊藤雅光「電子化テキストと画像データ」（『国語学』170, 1992. 09, p. 111－122）

［2］伊藤雅光「音声付き用例検索システムについて　－　『平曲』録音資料批判を事例研究として　－」（要旨,『国語学』174, 1993. 09, p. 69－70）

［3］伊藤雅光「日本語情報資料データベース構築のための準備的研究　平成3年度」（要旨,『平成3年度　国立国語研究所年報43』1993. 03, p. 56－7

［4］伊藤雅光「画像付き単語検索システムの研究」（要旨,『平成3年度国立国語研究所年報43』1993. 03, p. 114－5）

［5］伊藤雅光「日本語情報資料データベース構築のための準備的研究　平成4年度」（要旨,『平成4年度　国立国語研究所年報44』1993. 12, 8p予定）

［6］Proud, Judith K. : The Oxford Text Archive ; A Refport to the British Library Research and Development Department. (February 1989, British Library R＆D Report no. 5985, 44p)

［7］Oxford University Computing Service (ed.) : Oxford Text Archive Shortlist (December 1988, 40p)

［8］The Language Industries Survey (ed.) : The 1990 Language Engineering Directory (1989, Commission of the Eurorean Communities)

［9］Oxford University Computing Service (ed.) : Text Archive ; A Shortlist of Machine－Readable Texts Held at Oxford. (April 1992, iv, 86 p)

[10] Oxford University Computing Service (ed.) : Text Archive ; A

Shortlist of Machine－Readable Texts Held at Oxford. (October 1989, iii, 13p)

[11] Royal Institute of Technology (ed.) : Annual Report 1991 ; Department of Speech Communication & Music Acoustics. (1991, 22p)

[12] NERIS Trust (ed.) : NERIS... ; The national service that supports the teaching of the whole curriculum. (Maryland College, 10p)
※NERIS (National Educational Resources Information Service)

[13] Department of Computational Linguistics (ed.) : The Language Bank (University of Gothenburg, 16p) ※データは1991年1月現在のもの

[14] Centre National de La Recherche Scientifique (ed.) : L'institut de Recherche et d'Histoire des Textes (1987, 10p)

[15] Burnard, Lou : Tools and Tecniques for Computer－assisted Text Processing. (1992, Butler, C. S. (ed.) : Computers and Written Texts, p. 1－28)

[16] Hockey, S. M. : OCR ; The Kurzweil Data Entry Machine. (1986, Literary and Linguistic Computing, 1－2, p. 63－7)

[17] 伊藤雅光「テキストファイル訂正システム － TEFCOS －」(要旨,『計量国語学』17－3, 1989. 12, p. 136－7)

[18] Ito, Masamitu : PAFCOS ; Page File Correcting System. (1990, University of Nottingham, 111p)

[19] Sperberg－McQueen, C. M. and Burnard, L. (ed.) : Guidelines for the Encoding and Interchange of Machine－readable Texts. (October 1990, TEI Document Number : P1, version 1. 1, 290p) Chicago and Oxford : ACH－ACL－ALLC Text Encoding Initiative.

[20] Hockey, Susan : The ACH－ACL－ALLC Text Encoding Initiative : An Overview. (June 1991, 17p) ※第13回テキスト・データベース研究会〈JACH〉(1991. 6. 28, 会場：東京大学大型計算機センター) 配付資料

[21] ホッキー, スーザン (富岡克哉 訳)「TEI概観」(1991. 06, 15p) ※第13回テキスト・データベース研究会〈JACH〉(同上) 配付資料, 文献

[20] の翻訳
[22] Sperberg－McQueen, C. M. and Burnard, Lou : Living with the Guidelines ; An Introduction to TEI Tagging. (March 1991, TEI Document Number : TEI EDW18, 37p)
[23] Sperberg－McQueen, C. M. and Ploctkin, Wendy : Text Encoding Initiative : Current Documents. (October 1991, TEI Document Number : TEI A0, 34p)
[24] Sperberg－McQueen, C. M. and Burnard, Lou : Outline of TEI P2. (November, 1991, TEI Document Number : TEI ED W23, 6p)
[25] Bryan, Martin : SGML ; an author's guide to the standard generalized markup language. (1988, Addison－Wesley Publishing Company, xvii, 364p)
[26] ブライアン，マーチン（山崎俊一　監訳，福島誠　訳）『SGML入門』(1991, アスキー出版局, 379p) ※文献 [25] の翻訳

今回のアンケート調査では以下の方々のご協力を得ました。ご多忙中にもかかわらずご回答をお寄せくださったことに対し，深く感謝いたします。

1 R. Rodd Cambridge University
LLCC, Sidgwick Ave., Cambridge, CB3 9DA, U. K.
2 W. S. Dodd University of Exeter
Exeter, Ex 4 4QH, U. K.
3 David Taylor Maryland College
Leighton Street, Woburn, Milton Keynes MK17 9JD, U. K.
4 Lou Burnard Oxford University
13 Banbury RD, Oxford OX2 6RB, U. K.
5 D'agostini Giovanni D'agostini Organizzazione
Via Giusti, 17 33100 Udine, Italia
6 K. J. M. J. de Smedt Leiden University
P. O. Box 9555, 2300 RB Leiden, Netherlands
7 Martin Gellerstam University of Gothenburg
S－412 98 Göteborg, Sweden
8 Sheri Hunnicutt Royal Institute of Technology
Box 70014, S－100 44 Stockholm, Sweden

9 Gunnel Källgren Stockholm University
S－106 91 Stockholm, Sweden

10 Wolf Paprotté Westfälische Wilhelms－Universität Münster
Hüfferstrsse 27, D－4400 Münster, Germany

11 Ishigami－Iagolnitzer (Mitchiko) 石上美智子 Institut de Recherche et d'Histoire des Textes
40 avenue d'Iéna 75116 Paris, France

12 Otmar K. E. Foelsche Dartmouth College
6192 Bartlett Hall, Hanover, NH 03755－3530, U. S. A

13 Michael Neuman Georgetown University
238 Reiss Science Building, Washington, DC 20057, U. S. A

14 Mme Mireille Lacasse Gouvernement du Québec, Office de la langue française
700, boulevard Saint－Cyrille Est Québec (Québec) Canada G1R 5G7

15 Le'o Laroche Gouvernement du Québec, Ministere de l'Education
1035, de la Chevrotière Québec (Québec) Canada G1R 5A5

16 Pierre Maranda Université Laval
Québec QC GIK 7P4, Canada

17 Douglas O'shaughnessy University of Quebec
16 place du Commerce, Nuns Island, Quebec H3E 1H6, Canada

18 Reuven Merkin The Academy of Hebrew Language
Giv' at Ram Campus, Jerusalem, Israel

19 Lee, Sangsup Yonsei University
134 Shinchon－dong, Sudaemoon－ku, Seoul 120, 749, Korea

資料１　　アンケート回答状況

目　　次

［凡例］

1. 各質問項目は【Question】と【Answers】とから構成される。
2. 【Question】にはアンケート用紙にある質問文とその選択肢とをそのまま挙げた。
3. 【Answers】には各機関から送られてきた回答をまとめた。最初にある番号1～19は各機関に付けた機関番号である。その次に機関名、国名、選択肢番号の順で示してある。
4. 選択肢番号が入るべきところに“N”が書かれてある場合は、その項目に関する回答がなかったことを示している。

General information

【Question】

G01. Does your organisation belong to the private sector or to the public sector?

01 □ private sector
02 □ public sector
03 □ other (please specify).....

【Answer s】

1 Cambridge University イギリス 02
2 University of Exeter イギリス 02
3 Maryland College イギリス 03 ※ Educational Trust
4 Oxford University イギリス 02
5 D'agostini Organizzazione イタリア 01
6 Leiden University オランダ 02
7 University of Gothenburg スウェーデン 01
8 Royal Institute of Technology スウェーデン 02
9 Stockholm University スウェーデン 02
10 Westfälische Wilhelms－Universität Münster ドイツ 02
11 Institut de Recherche et d'Histoire des Textes フランス 02
12 Dartmouth College アメリカ 01
13 Georgetown University アメリカ 01
14 Gouvernement du Québec, Office de la langue française カナダ 02
15 Gouvernement du Québec, Ministère de l'Education カナダ 02
16 Université Laval カナダ 02
17 University of Quebec カナダ 02

18 The Academy of Hebrew Language イスラエル 02

19 Yonsei University 韓国 01

【Question】

G02. Indicate the main activities of your organisation.

【Answers】

1 Cambridge University Humanities Computing Research

2 University of Exeter Education & Research

3 Maryland College

Provision of a national database distributed on CD－Rom carrying information about learning resources. The database is also used to distribute full text resources for class room use.

4 Oxford University Archiving & Distribution of Electronic Texts

5 D'agostini Organizzazione

Industrial property, Automatic Translation Software, Automatic Translation Service

6 Leiden University Education, Research

7 University of Gothenburg Linguistic Study of Swedish

8 Royal Institute of Technology Research, Teaching

9 Stockholm University

Normal university activities : teaching, research

10 Westfälische Wilhelms－Universität Münster

Research, Teaching, Development

11 Institut de Recherche et d'Histoire des Textes

1) Photocopy of manuscripts of European countries, (East and West) , of Orient countries, and North African countries and conservation of microfilms.

2) Establishment of card－indexes of these manuscripts.

3) Establishment of catalogs of these titles by means of treatment of database by computers and publication.

4) Scientific research on these texts. Organisation of conferences and of congress on the study of texts.

5) Consultant of scolors for their reseach and for providing micro－films.

12 Dartmouth College

Liberal arts college with grad schools in business, engineering, medicine, biology, computer science etc.

13 Georgetown University

Creation, dissemination, & Analysis of Electronic text ; Cataloguing of Projects in Electronic Text

14 Gouvernement du Québec, Office de la langue française

Francisation des milieux de travail ; recherche, production et diffusion linguistiques et terminologiques ; normalisation terminologique; promotion du statut et de la qualité du fran aisçau Québec.

15 Gouvernement du Québec, Ministère de l'Education Educaton

16 Université Laval Research + Teaching

17 University of Quebec

Research + Education (Master's + Doctoral) in Speech + Image Processing and Computer Networks

18 The Academy of Hebrew Language

Producing Historical Dictionary of Hebrew

19 Yonsei University

1) Structuring large－scale corpora of Modern Korean.
2) Developing language engineering tools.
3) Compiling dictionaries of Korean.
4) Describing Korean on corpus linguistic princeples.

【Question】

G03. How many people are employed by your organisation in your country (do not include foreign offices) , including all departments and／or divisions?

01 ☐ less than 10
02 ☐ 10 to 50
03 ☐ 51 to 200
04 ☐ 201 to 500
05 ☐ 501 to 1, 000
06 ☐ 1, 001 to 5, 000
07 ☐ more than 5, 000

【Answers】

1 Cambridge University イギリス 01
2 University of Exeter イギリス 05
3 Maryland College イギリス 02
4 Oxford University イギリス 02
5 D'agostini Organizzazione イタリア 02
6 Leiden University オランダ 07
7 University of Gothenburg スウェーデン 02
8 Royal Institute of Technology スウェーデン 02
9 Stockholm University スウェーデン 03
for the Institute of Linguistics, Stockholm University
10 Westfälische Wilhelms－Universität Münster ドイツ 07
11 Institut de Recherche et d'Histoire des Textes フランス 07
The C. N. R. S. is the National research organisation which includes many hundreds of laboratories and Institutes in aural and Human sciences.
12 Dartmouth College アメリカ 05
13 Georgetown University アメリカ 07 ※for entire Univ.
14 Gouvernement du Québec, カナダ 04
Office de la langue française
15 Gouvernement du Québec, Ministère de l'Education カナダ 06
16 Université Laval カナダ 06
17 University of Quebec カナダ 05
18 The Academy of Hebrew Language イスラエル 02
19 Yonsei University 韓国 02

【Question】

G04. How many people are involved in the text repository or archive maintaining activities in the relevant department or division?

01 ☐ less than 10
02 ☐ 10 to 50
03 ☐ 51 to 100
04 ☐ more than 100

【Answers】

1 Cambridge University イギリス 01
2 University of Exeter イギリス 01
4 Oxford University イギリス 01
5 D'agostini Organizzazione イタリア 01
6 Leiden University オランダ 01
7 University of Gothenburg スウェーデン 01
8 Royal Institute of Technology スウェーデン 01
9 Stockholm University スウェーデン 01
10 Westfälische Wilhelms－Universität Münster ドイツ 01
11 Institut de Recherche et d'Histoire des Textes フランス 03
12 Dartmouth College アメリカ 01
13 Georgetown University アメリカ 01
14 Gouvernement du Québec, Office de la langue française カナダ 02
15 Gouvernement du Québec, Ministère de l'Education カナダ 02
16 Université Laval カナダ 01
17 University of Quebec カナダ 01
18 The Academy of Hebrew Language イスラエル 02
19 Yonsei University 韓国 02
［無回答の機関］ 3

【Question】

G05. Among the people who are employed by your organisation in the text repository or archive maintaining activities, how many have studied philology or bibliography?

01 □ 0
02 □ 1 to 5
03 □ more than 5

【Answers】

1 Cambridge University イギリス 01
2 University of Exeter イギリス 02
4 Oxford University イギリス 02
5 D'agostini Organizzazione イタリア 02
6 Leiden University オランダ 02

7 University of Gothenburg スウェーデン 02
8 Royal Institute of Technology スウェーデン 01
9 Stockholm University スウェーデン 02
10 Westfälische Wilhelms－Universität Münster ドイツ 02
11 Institut de Recherche et d'Histoire des Textes フランス 03
12 Dartmouth College アメリカ 02
13 Georgetown University アメリカ 02
14 Gouvernement du Québec, Office de la langue française カナダ 02
15 Gouvernement du Québec, Ministère de l'Education カナダ 02
16 Université Laval カナダ 01
18 The Academy of Hebrew Language イスラエル 03
19 Yonsei University 韓国 03
［無回答の機関］ 3, 17

【Question】

G06. Has the relevant department or division received any financial aid in respect to the text repository or archive maintaining activities?

01 □ yes
02 □ no

【Answers】

1 Cambridge University イギリス 02
2 University of Exeter イギリス 02
3 Maryland College イギリス 01
4 Oxford University イギリス 01
5 D'agostini Organizzazione イタリア 02
6 Leiden University オランダ 02
7 University of Gothenburg スウェーデン 01
8 Royal Institute of Technology スウェーデン 01
※ as part of general funding
9 Stockholm University スウェーデン 01 ※ research grants
10 Westfälische Wilhelms－Universität Münster ドイツ 01
11 Institut de Recherche et d'Histoire des Textes フランス 01
12 Dartmouth College アメリカ 01
13 Georgetown University アメリカ 02

14 Gouvernement du Québec, Office de la langue française カナダ 02
15 Gouvernement du Québec, Ministère de l'Education カナダ 02
16 Université Laval カナダ 02
17 University of Quebec カナダ 02
18 The Academy of Hebrew Language イスラエル 01
19 Yonsei University 韓国 01

【Question】

G07. If yes, please indicate who has aided the department or division. Provide information for as many as are applicable.

01 Government of your country
02 private foundation (s) (please specify).....
03 other (please specify).....

【Answers】

3 Maryland College イギリス 01
4 Oxford University イギリス 01, 03 ※03 (US Libraries – RLG)
7 University of Gothenburg スウェーデン 01
8 Royal Institute of Technology スウェーデン 01
9 Stockholm University スウェーデン 01
10 Westfälische Wilhelms－Universität Münster ドイツ 01, 03
※03 (private donors of text 〈machine readable〉
11 Institut de Recherche et d'Histoire des Textes フランス 01
12 Dartmouth College アメリカ 01, 02, 03 ※03 (individuals)
18 The Academy of Hebrew Language イスラエル 01
19 Yonsei University 韓国 01, 02
※02 (Yonsei University, Seung－kok Foundation)
［無回答の機関］ 1, 2, 5, 6, 13～17

【Question】

G08. Estimate the relevant department's or division's 1991 budget from your organisation in respect to the text repository or archive maintaining activities. If the department or division received any financial aid, please indicate the sum in the first round brackets, and

also indicate the aid source (Government・・・1, foundation・・・2, other・・・3) in the second round brackets.

(　　) (　　)

.....

【Answers】

1 Cambridge University イギリス 100 pounds N

2 University of Exeter イギリス None None

3 Maryland College イギリス

Not possible to separate from rest of funded activities N

4 Oxford University イギリス don't know don't know

5 D'agostini Organizzazione イタリア U.S. $ 1,000,000 N

7 University of Gothenburg スウェーデン Sek 50,000 1

8 Royal Institute of Technology スウェーデン nothing specific

9 Stockholm University スウェーデン Sek 800,000 1

11 Institut de Recherche et d'Histoire des Textes フランス N 1

14 Gouvernement du Québec, Office de la langue française カナダ

Sans objet

19 Yonsei University 韓国 US $ 1500,000 1,2

［無回答の機関］ 6, 10, 12, 13, 15～18,

【Question】

G09. Estimate the relevant department's or division's past budgets from your organisation, for as many years as information is available, in respect of the text repository or archive maintaining activities. If the department or division has received any financial aid, please indicate the sum in the first round brackets, and also indicate the aid number (Government・・・1, foundation・・・2, other・・・3) in the second round brackets. (continue on a separate page as necessary)

1990 (　　) (　　)

.....

1989 (　　) (　　)

.....

【Answers】

1 Cambridge University イギリス '88 (100) ,'89 (100),'90 (100)

2 University of Exeter イギリス None

3 Maryland College イギリス see above

4 Oxford University イギリス Don't know

5 D'agostini Organizzazione イタリア '90 (US$ 1,000,000) ,'89 (900,000), ' 88 (800,000),'87 (700,000),'86 (600,000),'85 (500,000)

9 Stockholm University スウェーデン A '90 (SEK 800,000) (1) , '89 (SEK 700,000) (1) We are building a corpus of modern Swedish and have received research grants for collecting and analyzing texts. When the corpus is ready, we will get more money.

11 Institut de Recherche et d'Histoire des Textes フランス
It is difficult to evaluate the sum concerning the text repository or archive maintaining activities. '61－'90 (1)

14 Gouvernement du Québec, Office de la langue française カナダ '90 (Sans objet)

19 Yonsei University 韓国 '90 (US$ 50,000) (2),'89 (US$ 70,000) (2)

［無回答の機関］ 6～8, 10, 12, 13, 15～18

Text Archive information

【Question】

T01. Please indicate ways which your organisation has gathered machine－readable texts, and also indicate their rough percentages in your repository or archive of machine－readable texts. Provide information for as many ways as are applicable.

01 ☐ donated by individual scholars .....%
02 ☐ donated by major research projects .....%
03 ☐ donated by other organisations .....%
04 ☐ bought from individual scholars .....%
05 ☐ bought from major research projects .....%
06 ☐ bought from other organisations .....%
07 ☐ prepared at your organisation .....%
08 ☐ other (please specify and use block capitals).....%

【Answers】

1 Cambridge University イギリス 01 (10%) ,03 (10%) ,07 (80%)
2 University of Exeter イギリス 01 () ,07 ()
3 Maryland College イギリス 03 (20%) ,07 (80%)
4 Oxford University イギリス 01, 02, 03, 07
5 D'agostini Organizzazione イタリア 06 (100%)
6 Leiden University オランダ 07 (100%)
7 University of Gothenburg スウェーデン 03(75%), 06 (15%), 07 (10%)
8 Royal Institute of Technology スウェーデン 01 (75%) ,07 (25%)
9 Stockholm University スウェーデン 07
100%, We have gathered machine-readable texts from many various sources and prepared and standerdized them for our own purposes.
10 Westfälische Wilhelms-Universität Münster ドイツ
02 (), 03() , 06 () ,07 (), 08 ※08 (donated by publishing industry)
11 Institut de Recherche et d'Histoire des Textes フランス 01 (5%) ,07 (95%)
12 Dartmouth College アメリカ 01 () ,02 () ,06 () ,07 ()
13 Georgetown University アメリカ 06-46%, 07-34%
14 Gouvernement du Québec, カナダ 01, 02, 03, 04, 05, 06, 07
Office de la langue française
16 Université Laval カナダ 01 (70%) ,07 (30%)
17 University of Quebec カナダ 03 (50%) ,06 (50%)
18 The Academy of Hebrew Language イスラエル 07 (100%)
19 Yonsei University 韓国 03 (10%) ,07 (90%)
※ (03) So far, we have collected 20 million running words of Korean texts published since 1970.

［無回答の機関］ 15

【Question】

T02. If you have prepared any machine-readable texts, please indicate the methods for inputting the text data, and also indicate their rough percentages. Provide information for as many methods as are applicable.

01 □ traditional keyboarding methods.....%
02 □ OCR devices.....%

03 □ other (please specify and use block capitals).....%

【Answers】

1 Cambridge University イギリス 01 (90%) ,02 (10%)
2 University of Exeter イギリス 01
3 Maryland College イギリス 01 (60%) ,02 (40%)
4 Oxford University イギリス 01, 02
5 D'agostini Organizzazione イタリア 02
6 Leiden University オランダ 01
7 University of Gothenburg スウェーデン 01 (85%) ,02 (15%)
8 Royal Institute of Technology スウェーデン 01 (100%)
9 Stockholm University スウェーデン 02 (< 5%)
10 Westfälische Wilhelms－Universität Münster ドイツ 03
※03 (magnetic tape)
11 Institut de Recherche et d'Histoire des Textes フランス
01 (90%) ,02 (10%)
12 Dartmouth College アメリカ 01, 02
13 Georgetown University アメリカ 02－100%
14 Gouvernement du Québec, カナダ 01－97%, 02－1%, 03－2%
Office de la langue française ※03 (Programe de transfert informa－tique à partir de disquettes fourn-ies)
15 Gouvernement du Québecç Ministère de l'Education カナダ 01(100%)
16 Université Laval カナダ 01 (100%)
18 The Academy of Hebrew Language イスラエル 01 (100%)
19 Yonsei University 韓国 01 (90%) ,03 (10%)
※accepted machin－readable texts only
［無回答の機関］ 17

【Question】

T03. If you have used any OCR devices, please indicate the OCR model. (e. g. the Kurzwei 4000, Palantir 3000, etc.)

【Answers】

1 Cambridge University イギリス Mac scanne

3　Maryland College　イギリス　Calera Truescan
4　Oxford University　イギリス　Kurzweil 4000
5　D'agostini Organizzazione　イタリア　Omnypage－Typist
7　University of Gothenburg　スウェーデン　Kurzweil 4000
9　Stockholm University　スウェーデン　Apple
11　Institut de Recherche et d'Histoire des Textes　フランス　Apple
12　Dartmouth College アメリカ　all Kurzweil Plus Acutext, Omnipage etc.
13　Georgetown University　アメリカ　Kurzweil 4000, 5100
14　Gouvernement du Québec,　カナダ　　Hewlett Packard ScanJet 11c
Office de la langue française
［キーボード入力のみの機関］　8, 10, 15～19
［無回答の機関］　2, 6, 17

【Question】

T 04.　What kinds of machine－readable texts does your organisation gather?
You may tick as many kinds as are applicable.
01 ☐ every kind of machine－readable texts
02 ☐ literature
03 ☐ newspaper
04 ☐ dictionary
05 ☐ academic paper
06 ☐ school textbook
07 ☐ other　(please specify and use block capitals)

【Answers】

1　Cambridge University　イギリス　02, 03, 06
2　University of Exeter　イギリス　02
3　Maryland College　イギリス　06
4　Oxford University　イギリス　01
5　D'agostini Organizzazione　イタリア　01
6　Leiden University　オランダ　04
7　University of Gothenburg　スウェーデン　02, 03, 04, 05, 06, 07　※07
(se brochere！)
8　Royal Institute of Technology　スウェーデン　02, 03, 04

9 Stockholm University スウェーデン 02, 03, 05, 07
※07 (governmental, administrative, magazines)
10 Westfälische Wilhelms－Universität Münster ドイツ 02, 03, 04, 05
11 Institut de Recherche et d'Histoire des Textes フランス 07
※ catalogs of manuscripts
12 Dartmouth College アメリカ 02, 04
13 Georgetown University アメリカ 01, 07 ※07 (Philosophy, Art Commentary)
14 Gouvernement du Québec, Office de la langue française カナダ
04, 05, 06, 07
※04 (Il s'agit plus particulièrement de fichiers de données terminologiques ou documentaires.)
07 (Documents spècialisès dans un des domaines où il y a des projets terminologiques en cours.)
15 Gouvernement du Québec, Ministère de l'Education カナダ
01, 07 ※ 07 (Guidelines, Tests)
16 Université Laval カナダ 07
※ 07 (First－hand data collected by our field workers)
17 University of Quebec カナダ 03, 04
18 The Academy of Hebrew Language イスラエル 01
19 Yonsei University 韓国 02, 03, 04, 05, 06, 07
※07 (Scholarly writings in social, cultural, hitorical studies, sports and recreation.)

【Question】
T05. Please indicate the main language－readable texts.

【Answers】
1 Cambridge University イギリス French
2 University of Exeter イギリス Texts used in the past are no longer held, thus not applicable.
3 Maryland College イギリス English
4 Oxford University イギリス English, but not exclusively
5 D'agostini Organizzazione イタリア All European Languages
6 Leiden University オランダ Dutch

7 University of Gothenburg スウェーデン Swedish
8 Royal Institute of Technology スウェーデン Swedish
9 Stockholm University スウェーデン Swedish
10 Westfälische Wilhelms－Universität Münster ドイツ German, English, Spanish, Modern Greek
11 Institut de Recherche et d'Histoire des Textes フランス latin (but also French, English, Russian, Arabic, Hebraic, Greek etc)
12 Dartmouth College アメリカ English
13 Georgetown University アメリカ English, German, Italian
14 Gouvernement du Québec, カナダ Français Office de la langue française
15 Gouvernement du Québec, Ministère de l'Education カナダ Franch
16 Université Laval カナダ French, English, Spanish, LAO (Solomon Islands)
17 University of Quebec カナダ English
18 The Academy of Hebrew Language イスラエル Hebrew
19 Yonsei University 韓国 Korean only

【Question】

T06. Who selects new machine－readable texts for gathering?

01 □ staff
02 □ a steering committee
03 □ other (please specify and use block capitals)

【Answers】

1 Cambridge University イギリス 03 ※03 (Scholars)
2 University of Exeter イギリス 01
3 Maryland College イギリス 01
4 Oxford University イギリス 03 ※03 (We accept whatever get)
5 D'agostini Organizzazione イタリア 01
6 Leiden University オランダ 01
7 University of Gothenburg スウェーデン 01
8 Royal Institute of Technology スウェーデン 01
9 Stockholm University スウェーデン 01
10 Westfälische Wilhelms－Universität Münster ドイツ 03

※03 (opportunistic intake : we accept 〈more or less〉 what we get)

11 Institut de Recherche et d'Histoire des Textes フランス 01

12 Dartmouth College アメリカ 03

※03 (individuals in various departments)

13 Georgetown University アメリカ 01

14 Gouvernement du Québec, Office de la langue française カナダ 01

15 Gouvernement du Québec, Ministère de l'Education カナダ 02

16 Université Laval カナダ 03 ※03 (Project head 〈P. Maranda〉

17 University of Quebec カナダ 01

18 The Academy of Hebrew Language イスラエル 02

19 Yonsei University 韓国 02

【Question】

T 07. If your organisation checks newly gathered machine-readable texts, please indicate the main items to be checked. Provide information for as many items as are applicable.

01 ☐ no check

02 ☐ errors in the machine-readable texts

03 ☐ bibliographical data of the copy texts used in preparing the machine-readable texts

04 ☐ methods of inputting the machine-readable texts (e.g. keyboard, OCR, etc.)

05 ☐ other (please specify and use block capitals)

【Answers】

1 Cambridge University イギリス 02, 03

3 Maryland College イギリス 02

4 Oxford University イギリス 03, 04

5 D'agostini Organizzazione イタリア 01

6 Leiden University オランダ 01

7 University of Gothenburg スウェーデン 01

9 Stockholm University スウェーデン 02

10 Westfälische Wilhelms-Universität Münster ドイツ 03

11 Institut de Recherche et d'Histoire des Textes フランス 05

※05 (Our institute treats by computer the bibliographic database.

If we do research on a text, use study it profoundly from the point of a scientific criterium.)

12 Dartmouth College アメリカ 01
13 Georgetown University アメリカ 02, 05 ※05 (Documentation)
14 Gouvernement du Québec, Office de la langue française カナダ 02
15 Gouvernement du Québec, Ministère de l'Education カナダ 02
16 Université Laval カナダ 02, 04
17 University of Quebec カナダ 01
19 Yonsei University 韓国 02, 03, 04
［無回答の機関］ 2, 8, 18

【Question】

T08. For checking, does your organisation have been gathering the same copy texts used in preparing the machine－readable texts? If it has been doing so, what is the rough percentage of the macine－readable texts whose copy texts were gathered in the total machine－readable texts.

01 □ yes.....%
02 □ no

【Answers】

1 Cambridge University イギリス 01 (20%)
3 Maryland College イギリス 02
4 Oxford University イギリス 02
9 Stockholm University スウェーデン 01 ※striving for 100%
10 Westfälische Wilhelms－Universität Münster ドイツ 01 ()
13 Georgetown University アメリカ ※not for purchased electronic texts
14 Gouvernement du Québec, Office de la langue française カナダ 01
15 Gouvernement du Québec, Ministère de l'Education カナダ 02
16 Université Laval カナダ 01 (100%)
19 Yonsei University 韓国 01 (100%)
［無回答の機関］ 2, 5～8, 11, 12, 17, 18

【Question】

T09. Does your organisation know Text Encoding Initiative (TEI) ?

01 □ yes

02 □ no

【Answers】

1 Cambridge University イギリス 01

2 University of Exeter イギリス 01

3 Maryland College イギリス 02

4 Oxford University イギリス 01

6 Leiden University オランダ 02

7 University of Gothenburg スウェーデン 01

8 Royal Institute of Technology スウェーデン 02

9 Stockholm University スウェーデン 01

10 Westfälische Wilhelms－Universität Münster ドイツ 01

12 Dartmouth College アメリカ 01

13 Georgetown University アメリカ 01 ※ Affiliated Project

14 Gouvernement du Québec, Office de la langue française カナダ 02

15 Gouvernement du Québec, Ministère de l'Education カナダ 02

16 Université Laval カナダ 01

17 University of Quebec カナダ 02

18 The Academy of Hebrew Language イスラエル 01

19 Yonsei University 韓国 01

［無回答の機関］ 5, 11,

【Question】

T10. If your organisation knows TEI, will your organisation adopt the TEI Gidelines as a common text－encoding scheme in the future ?

01 □ yes

02 □ no

03 □ under consideration

04 □ other (please specify and use block capitals)

【Answers】

1 Cambridge University イギリス 03

2 University of Exeter イギリス 03
4 Oxford University イギリス 01
7 University of Gothenburg スウェーデン 01
9 Stockholm University スウェーデン 01
10 Westfälische Wilhelms－Universität Münster ドイツ 01
12 Dartmouth College アメリカ 03
13 Georgetown University アメリカ 01
16 Université Laval カナダ 03
18 The Academy of Hebrew Language イスラエル 02
19 Yonsei University 韓国 03
［無回答の機関］ 3, 5, 6, 8, 11, 14, 15, 17

【Question】
T11. If your organisation will not adopt the TEI Gidelines, please indicate its reason. (please specify and use block capitals)

【Answers】
13 Georgetown University アメリカ
We will adopt at least a modified version, but encoding & verifying will be more time－consuming than editing text.
14 Gouvernement du Québec, Office de la langue française カナダ
Nous considédrons davantage l'étude de la norme SGML (Standard Generalized Markup Language)
18 The Academy of Hebrew Language イスラエル
We have developed our own Encoding methods since 1959, designed especially for Hebrew and for our Historical Dictionary Project.
［無回答の機関］ 1～12, 15～17, 19

【Question】
T12 a. Please indicate past, present and anticipated future problems in maintaining your repository or archiver of machinereadable texts. (Please specify and use block capitals)
[past]

【Answers】

1 Cambridge University イギリス Lack of staff, time

3 Maryland College イギリス

1) Final print version of text is not consistent with machine readable form being offered the text has been amended by typesetten at the final stage.

2) Word－processor formats are different.

3) Disc formats are different.

4 Oxford University イギリス Recognition

9 Stockholm University スウェーデン

We are building balanced corpus with the same structure as Brown and LOB. Getting texts, and in particular getting copyrights, from so many different sources has been extremely time－consuming.

10 Westfälische Wilhelms－Universität Münster ドイツ money－personnel

11 Institut de Recherche et d'Histoire des Textes

The I. R. H. T. has not yet enterd in database the machine－readable texts, because we have too much work (establishment of the repertories and catalogs of manuscripts and of printed books of Europe, of Near Orient and of North Africa) and because of many languages treated (ancient and modern). But one day we will have this opportunity, I hope.

12 Dartmouth College アメリカ Quality + Copyright

13 Georgetown University

Continued funding of editors' time. Copyright permissions from publishers and authors

14 Gouvernement du Québec, Office de la langue française

Acquisition des droits d'auteur

16 Université Laval カナダ Money

19 Yonsei University 韓国 Collecting representative texts

［無回答の機関］ 2, 5～8, 15, 17, 18

【Question】

T12 b. Please indicate past, present and anticipated future problems in maintaining your repository or archiver of machine－readable

texts. (Please specify and use block capitals)
[present]

【Answers】

1 Cambridge University イギリス Lack of staff, time

2 University of Exeter イギリス
A small number of short texts have been used for specific projects in the past, but none are corrently retained, thus there is no problem, and there is no archive at present.

3 Maryland College イギリス 1 and 2 as past

7 University of Gothenburg スウェーデン
More difficult to obtain texts while organisations have found that they have commercial value (spelling－checkers etc.)

9 Stockholm University スウェーデン
Still copyrights, problem which I think is often overlooked

12 Dartmouth College アメリカ Same (Quality + Copyright)

13 Georgetown University アメリカ Same

14 Gouvernement du Québec, Office de la langue française
Acquisition et gestion des droits d'auteur

16 Université Laval カナダ Money

17 University of Quebec カナダ Lack of sufficient disk space－memory

19 Yonsei University 韓国 Storage and maintenance mechanism

［無回答の機関］ 4～6, 8, 10, 11, 15, 18

【Question】

T12 c. Please indicate past, present and anticipated future problems in maintaining your repository or archiver of machine－readable texts.
(Please specify and use block capitals)
[future]

【Answers】

1 Cambridge University イギリス Lack of staff, time

9 Stockholm University スウェーデン Money

12 Dartmouth College アメリカ Same (Quality + Copyright)

13 Georgetown University アメリカ
Same, Competition from publishers, Uncertainty over imaging standards
16 Université Laval カナダ Money
19 Yonsei University 韓国
Trained man power for analyzing the data, and for lexicographical projects
［無回答の機関］ 2～8, 10, 11, 14, 15, 17, 18

【Question】
T13 a. Please indicate solutions to the problems.
(Please specify and use block capitals)
[past]

【Answers】
3 Maryland College イギリス
1) Use ICR and scan final text. 2) Use Ascii base. 3) MS－DOS has become norm.
10 Westfälische Wilhelms－Universität Münster ドイツ
improvise and keep on requests for funding
12 Dartmouth College アメリカ None so far
13 Georgetown University アメリカ
Support of university administration rayalty payments on copies sold
16 Université Laval カナダ Money
19 Yonsei University 韓国 Social survey of reading habits of average adults
［無回答の機関］ 1, 2, 4～9, 11, 14, 15, 17, 18

【Question】
T13 b. Please indicate solutions to the problems.
(Please specify and use block capitals)
[present]

【Answers】
13 Georgetown University アメリカ as above

14 Gouvernement du Québec, Office de la langue française カナダ

Établissement d'une procédure d'acquisition des droits d'auteur; négociation des droits d'auteur confiée à une personne dans l'organisme ; centralisation de toute l'information concernant l'acquisition et lagestion des droits d'auteur.

16 Université Laval カナダ Money

19 Yonsei University 韓国

Purchase or lease of appropriate hardware [very expensive]

[無回答の機関] 1～12, 15, 17, 18

【Question】

T13 c. Please indicate solutions to the problems.
(Please specify and use block capitals)
[future]

【Answers】

5 D'agostini Organizzazione イタリア

"Fast" Optical Reading System

10 Westfälische Wilhelms－Universität Münster ドイツ

(develop applocations)

13 Georgetown University アメリカ

More purchases, less creation, select highest quality formats

(E. C., resolution of images)

16 Université Laval カナダ Money

19 Yonsei University 韓国

Special seminars, Establishing lectures on computational lexicography, language information, Interns in computational linguistics and language engineering.

[無回答の機関] 1～4, 6～9, 11, 12, 14, 15, 17, 18

- The 1991 Machine-readable Text Archives Survey -

Background

The National Language Research Institute has instituted a plan to establish a repository or archive of machine-readable texts. For this purpose, the Department started in April 1991 a preparatory study to survey repositories or archives of machine-readable texts throughout the world and to analyse problems of maintaining them. The results of this study will be published in 1993.

Notes on completing the questionnaire

The questionnaire below consists of two parts. Questions G01 to G09 request general information on the nature of your organization. Questions T01 to T13 inquire about the repository or archive of machine-readable texts your organization maintains.

If you find any of the questions to be too difficult to answer, please proceed to the next question.

Please complete the questionnaire in block capitals using blue or black ink.

Thank you for your kind cooperation!

Reference information (*please use block capitals*)

respondent's name .....

job title .....

academic background .....

organization .....

department .....

address .....

.....

telephone .....

telefax .....

e-mail address (*if any*).....

date .....

G01. Does your organization belong to the private sector or to the public sector?

01 ☐ *private sector*
02 ☐ public sector
03 ☐ other (*please specify*) .....

G02. Indicate the main activities of your organization.

.....

.....

.....

.....

G03. How many people are employed by your organization in your country (do not include foreign offices), including all departments and/or divisions?

01 ☐ less than 10
02 ☐ 10 to 50
03 ☐ 51 to 200
04 ☐ 201 to 500
05 ☐ 501 to 1,000
06 ☐ 1,001 to 5,000
07 ☐ more than 5,000

G04. How many people are involved in the text repository or archive maintaining activities in the relevant department or division?

01 ☐ less than 10
02 ☐ 10 to 50
03 ☐ 51 to 100
04 ☐ more than 100

G05. Among the people who are employed by your organization in the text repository or archive maintaining activities, how many have studied philology or bibliography?

01 ☐ 0
02 ☐ 1 to 5
03 ☐ more than 5

G06. Has the relevant department or division received any financial aid in respect to the text repository or archive maintaining activities?

01 ☐ yes
02 ☐ no

G07. If yes, please indicate who has aided the department or division. Provide information for as many as are applicable.

01 ☐ Government of your country

02 ☐ private foundation(s) (*please specify*).....

.....

03 ☐ other (*please specify*) .....

G08. Estimate the relevant department's or division's 1991 budget from your organization in respect to the text repository or archive maintaining activities. If the department or division received any financial aid, please indicate the sum in the first parentheses, and also indicate the aid source (Government···1, foundation···2, other···3) in the second parentheses.

( ) ( )
.....

G09. Estimate the relevant department's or division's past budgets from your organization, for as many years as information is available, in respect of the text repository or archive maintaining activities. If the department or division has received any financial aid, please indicate the sum in the first parentheses, and also indicate the aid number (Government···1, foundation···2, other···3) in the second parentheses. (*continue on a separate page as necessary*)

1990 ( ) ( )
.....
1989 ( ) ( )
.....
1988 ( ) ( )
.....
1987 ( ) ( )
.....
1986 ( ) ( )
.....
1985 ( ) ( )
.....

1984	( )	( )
1983	( )	( )
1982	( )	( )
1981	( )	( )
1980	( )	( )
1979	( )	( )
1978	( )	( )
1977	( )	( )
1976	( )	( )
1975	( )	( )
1974	( )	( )
1973	( )	( )
1972	( )	( )
1971	( )	( )
1970	( )	( )
1969	( )	( )
1968	( )	( )
1967	( )	( )
1966	( )	( )
1965	( )	( )
1964	( )	( )
1963	( )	( )
1962	( )	( )
1961	( )	( )

[4]

Text Archive information(*duplicate if necessary*)

T01. Please indicate ways which your organization has gathered machine-readable texts, and also indicate their rough percentages in your repository or archive of machine-readable texts. Provide information for as many ways as are applicable.

01 ☐ donated by individual scholars .....%

02 ☐ donated by major research projects .....%

03 ☐ donated by other organizations .....%

04 ☐ bought from individual scholars .....%

05 ☐ bought from major research projects .....%

06 ☐ bought from other organizations .....%

07 ☐ prepared at your organization .....%

08 ☐ other (*please specify and use block capitals*) .....%

.....

.....

T02. If you have prepared any machine-readable texts, please indicate the methods for inputting the text data, and also indicate their rough percentages. Provide information for as many methods as are applicable.

01 ☐ traditional keyboarding methods .....%

02 ☐ OCR devices .....%

03 ☐ other (*please specify and use block capitals*) .....%

.....

.....

T03. If you have used any OCR devices, please indicate the OCR model. (*e.g. the Kurzwei 4000, Palantir 3000, etc.*)

.....

[5]

T04. What kinds of machine-readable texts does your organization gather? You may tick as many kinds as are applicable.

01 ☐ every kind of machine-readable texts
02 ☐ literature
03 ☐ newspaper
04 ☐ dictionary
05 ☐ academic paper
06 ☐ school textbook
07 ☐ other (*please specify and use block capitals*)

.....

.....

T05. Please indicate the main language of your machine-readable texts.

.....

T06. Who selects new machine-readable texts for gathering?

01 ☐ staff
02 ☐ a steering committee
03 ☐ other (*please specify and use block capitals*)

.....

T07. If your organization checks newly gathered machine-readable texts, please indicate the main items to be checked. Provide information for as many items as are applicable.

01 ☐ no check
02 ☐ errors in the machine-readable texts
03 ☐ bibliographical data of the copy texts used in preparing the machine-readable texts
04 ☐ methods of inputting the machine-readable texts (*e.g. keyboard, OCR, etc.*)
05 ☐ other (*please specify and use block capitals*)

.....

T08. For checking, does your organization also maintain a collection of source copy texts used in preparing the machine-readable texts? If so, what is the rough percentage of the machine-readable texts whose copy texts

were gathered in the total machine-readable texts.

01 ☐ yes .....%
02 ☐ no

T09. Is your organization familiar with *Text Encoding Initiative* (*TEI*)?

01 ☐ yes
02 ☐ no

T10. If your organization is familiar with *TEI*, will your organization adopt the *TEI Guidelines* as a common text-encoding scheme in the future?

01 ☐ yes
02 ☐ no
03 ☐ under consideration
04 ☐ other (*please specify and use block capitals*)

.....

.....

T11. If your organization will not adopt the *TEI Guidelines*, please indicate the reason(s). (*please specify and use block capitals*)

.....

.....

.....

T12. Please indicate past, present and anticipated future problems in maintaining your repository or archive of machine-readable texts. (*please specify and use block capitals*)

[past]

.....

.....

.....

.....

.....

.....

[present]

.....

.....

.....

.....

.....

.....

[future]

.....

.....

.....

.....

.....

.....

T13. Please indicate solutions to the problems.
(*please specify and use block capitals*)

[past]

.....

.....

.....

[8]

.....

.....

.....

[present]

.....

.....

.....

.....

.....

.....

[future]

.....

.....

.....

.....

.....

.....

Please return completed questionnaires to:-

Masamitsu Ito
The National Language Research Institute
3-9-14 Nisigaoka, Kita-ku
Tokyo 115, Japan
E-mail : m-ito@tansei.cc.u-tokyo.ac.jp