JICST の実用的全自動漢字—カナ変換システム，K-KACS について†

荒木啓介††　板山和彦††

1. はじめに

漢字，かな混り日本語文を分かち書きされたカナ文字文に自動変換する課題は，その逆が日本語文の入力法の一環として広く研究されているのに対し，あまり目立ったテーマではなく，過去に一，二の研究[1),2)]があるにすぎない．しかし漢字—カナ変換は意外と実用性がある．すなわち(1)漢字モードファイルをカナ文字モードファイルに自動変換できるため，より高速，低廉な入出力機器によるファイルの使用を可能とする．(2)表意文字を表音文字系に変換することによる，自動朗読，自動点訳などへの用途，(3)文章中の用語，キーワードの抽出とそのフリガナ付加機能による，用語のソート，自動化情報検索への利用，などが考えられる．

JICST は内外の科学技術情報を収集・加工して一般にサービスすることを本務とする科学技術庁傘下の特殊法人であるが，情報を扱っている性質上，当初より用語，言語と係わりあいを持ってきた．昨今言語問題が論じられてにぎやかであるが，私共における立場は，大学や国公立の研究所，あるいは情報処理のハードやソフトを売るメーカとは異なり，情報サービスの向上と内部における省力化が基本的な目的であって，役に立つ技術を積み上げることが第一に要求される．他方，実用システムにより大量データで検証した経験は，アカデミックなサイドへも新しい話題を提供できるかもしれない．

さて，JICST では，機械編集した科学技術情報を，72年4月から文献検索磁気テープとして一般に提供するに際し，困った問題が生じた．英カナモードのテープは，抄録文を含まない他，論文タイトルも欧文のものは原語のまま，ロシア語はキリル文字を英字に翻字したものしか含められず，日本語のものは，ノータイトルを避けるため人手によるカナ振りを余儀なくされた．これは 76年4月にスタートした JOIS オンラインサービスのファイル内容にも継続し，ユーザにも不便を強いてきた．全論文タイトル（年間36万件以上）に人手でフリガナ付けし，パンチ入力することは不可能に近く，内部でもさまざまな案が出されていたが，私共は別途アイデアを持ち，以下のごとく問題を整理して研究，開発を決意した．

(1) 目的の明確化
JICST 理工学文献漢字モードファイルのタイトルのカナ変換

(2) 範囲の限定
科学技術分野に限り，タイトル文に限る．固有名詞には深入りせず．漢字も，JICST 使用の 1842（凸凹を入れると 1844）のみ．

(3) 作業の徹底主義
入力者側に特に注文をつけず，自然に記入されたタイトルを変換できるよう，日本語文のさまざまな表記法に対処する．

(4) 機能の単純化
実用システムをめざすため，複雑な文章分析，合成処理，例外処理はやらない．代りに，辞書は多少大きくなっても良い．

(5) 拡張性，メンテナンスへの配慮
分野の拡張（現在未処理の医学，農学など），分かち書きのみの使用，抄録文への適用も可能なように配慮．メンテナンスの容易さも考慮．

2. システムの概要

2.1 基本方針

カナ変換する前に分かち書きするか，カナ変換してから分かち書きするか，の選択があるようだが，字種情報を含めた情報量の多い段階で分かち書きし，その結果を利用する他，文頭，文末にさらにスペースを挿入してすべての文字群を全く同格に置いてからカナ変

† Practical Kanji—Kana Conversion, JICST Full Automatic Kanji—Kana Conversion System, K-KACS by Keisuke ARAKI and Kazuhiko ITAYAMA (Information Div., the Japan Information Centre of Science and Technology).

†† 日本科学技術情報センター

換した方が万事好都合である．応用性も広まる．

（1） 分かち書き

日本語の場合，漢字，カタカナ，記号類は主として名詞であり，漢字を含む動詞，形容詞などはたいていひらがなの語尾を持つこと，それ以外の助詞などはひらがなであるから，文末より，ひらがなを検出してはじめて辞書照合させる．分かち書き辞書は，語尾ひらがなの助詞，助動詞，接続詞などの連で，最長照合する．漢字・カタカナのみの長大な文字連の切断は行わない．

ただしこの方式ではかなりの不備がある．つまり日本語においては，漢字書きせずひらがな書きをしたり，使用範囲外の漢字は JICST のルールによりひらがな書きされるため，分かち書きに際し，字種に頼り切れない．

（例） りん，ひ素，ほう素，たんぱく質，しゅよう，

これらの正しい処理のためには，後述する特別の辞書を用いる．

（2） 漢字一カナ変換

漢字の種類とその読み方に何らかの関連性があることは誰でも気づくが，その規則化となると容易ではない．単純な音訓規則だけでは正しい読み分けは望めない．我々は漢字の読み方とその種類との関係を，実際に即して分析し，当初表-1 の左欄，システム化の段階では右欄のように分類整理した．これらの例を挙げてみると，読みが1種の IA 類は亜，声，阿，案，医，胃，姻，員，院，逸，域，意，維，緯，磯，韻，宇，英，疫，液，駅，閲，演，衛，王，央，欧，恩，億，憶，……と，ア行だけでもこれほどあり，意味的にもバラエティに富んでいる．2A 類は，名詞性の漢字で，素直に音訓規則に従う，園，災，机，茎，的，滴，など．2B 類は形容詞，副詞性の漢字で，予(め)，再(び)，熱(い)，概(ね)，異(なる)，優(しい)，優(れた)など，一定の送りがなを伴った場合にはそれにしたがって読みを与えることができるものである．2C 類は，歌，歌う，歌唱，のように，名詞でも動詞でもあり得る漢字で，他に，願(ネガイ)，願う，飢(ウエ)，飢える，疑(ウタガイ)，疑う，などがある．3類は，度数の少ない読み方のみを辞書化し，他方の読みを他のすべての場合について与えるもので，鉛（鉛筆，亜鉛，黒鉛，有鉛，など 24 熟語をエンと読ませ，他はナマリを与える方が，酢酸鉛，硫酸鉛，等々と限りのないケースに正しく対処できる），型（熟語以外はガタとする），付（付着，接着等々の熟語以外は，ツキ，とする），東（熟語以外は，ヒガシ，とする）等々．4A類，4B類は，2A類，2B類に対応し，しかも例外読みの熟語を持つもので，前者は，手（手順，手動，手），汗，火，歯，など，後者は形容詞，動詞性の漢字の，取（取る，下取，取締，取水），吸，巻，など，4C類は，行，通，など．“行った”，“通った”，の区別に，へ，に，を，等の助詞を利用するものである（ただし間に副詞が入って判別できない場合には，科学技術上多いと思われる，オコナッタ，トオッタ，の方を優先させる）．

表-1 漢字の分類

分類	定義	字数	実システムの分類	処理タイプ
1類	読み方一つのみ	682	IA	A
2類	音読規則あり	275		
2A	送りがな無し（名詞性）	38	NA	A
2B	送りがなあり	212	VA	A
2C	読みに送りがな用いる場合と用いない場合	25	VB	B
3類	少数の例外以外一つの読み方のみ	114	NA	A
4類	例外読み複合語あり	771		
4A	送りがな無し（名詞性）	319	NB	B
4B	送りがなあり（動詞，形容詞性）	448	VB	B
4C	助詞まで考慮して読み分ける必要	4	CA	A
（8種類）			（6種）	（2種）

（注） 字数は辞書ごとに変動した．

実際のシステムでは，送りがなも熟語も含めて辞書化し，処理は，辞書以外は読み一つを持つA処理と音訓規則を適用するB処理の二つとし，名詞性のNと非名詞性のVという形に単純化した．辞書化にあたっては，2通り以上の読みがある場合，種類の少ない方を辞書化するようにしたことが，このシステムの特徴である．

2.2 分かち書き処理

分かち書き辞書と処理は以下のとおりである．

タグ，処理タイプ，処理結果，

に対する，A，△に△対する△

に及ぼす，A，△に△及ぼす△

の時の，A，△の△時△の△

マッチしたら処理結果を単純に転送する．

2文字，1文字のタグは，ミスマッチを起す可能性が高いので，そのタグの前または後に特定の文字がある時は分かち書きタグと見なさないよう指示する．

後例外字，前例外字

より，A，△より△，糸線戻数，ただこ

これにより，より糸，より線，より戻，たより，などは切断されない．

1文字のタグについては，“の”と“と”のみをA処理（無条件でタグと見なす）とし，それ以外は，前後がひらがな以外の場合のみタグと見なすB処理とし，それぞれに，前後の例外を指定する．

や，B，△や△，金
な，B，△な△，染，恵
と，A，△と△，殺場石粒畜るれっき……
の，A，△の△，うむめっんりるれち，も

さらに，先に触れた，かな書き名詞の誤った切断を救済するため，名詞を含んだ特殊な辞書を用意した．

のりん，A，△の△，りん
そのりん，A，△その△りん

これらの措置により，

のり　　（食べるのり）
△の△りん酸
△その△りん酸

と正しく分かち書きできる．

“と”の場合，と殺　や　と石　を連続させるため，石油と石炭　などが続いてしまうのを最終的に避けるため，やむを得ず原則を崩して，カナ変換の段階で

と石炭 ⟶ と△セキタン
と石油 ⟶ と△セキユ

のように変換させることにし，解決した．

分かち書き辞書の実行形式は，タグの末尾のひらがなによるテーブル，そのアドレスの下での，同一末尾ひらがなを持つタグの長い順ソートに格納した辞書ファイルとなっており，全部をコア上に展開し，高速処理を期している．タグ数は最新の辞書 D_6 で，1,602語である．分かち書き処理の最終段階では，この処理によって新たに作られたスペースが重複しているかどうかチェックし，一つに減らす処理を行う．

2.3 漢字—カナ変換

先ず入力文字連の前後にスペースを入れ，すべての文字群を同格に置く．処理は次の2通りのみに簡略化した．

A処理：例外辞書中でマッチしない場合には一つの読みのみを与える．

B処理：例外辞書中でマッチしない場合には前後の文字種を調べ，漢字なら音，ひらがなやスペースなら訓を与える．

実行辞書は，送りがな，熟語，前後の助詞を含む連をすべて同等と見なし，最初の漢字を親漢字見出しとして，漢字に関してはすべて後の方向へのみ照合するようにし，前方に漢字以外の文字連がある場合には，その字数だけ逆のぼって照合する．漢字以外は読み飛ばして処理ずみのエリアに転送してあるので，この場合にも，熟語全体中，漢字部分のみの読みを与える．

例えば，　歯　　の変換では，

歯，シ，ハ，B
歯そう，シソウ
歯茎，ハグキ
歯並び，ハナラビ
むし歯，バ
ムシ歯，バ
歯車，ハグルマ
……

これにより，

……の△歯△を△……

歯肉炎，　　歯根　歯車　　などを正しく読み分けられる．　　虫歯，奥歯，などは，それぞれ　虫，　奥，　の親漢字のもとに格納される，

ただし人手で辞書を作り，メンテナンスする場合にはこれでは不便なので，虫歯，奥歯　は，歯のところでまとめて入れても良く，実行形式辞書を作るときに重複消去して，しかるべき親漢字のもとに集めるようにしている．

辞書の工夫は，以下の通り．

（1） 音訓逆転，　　鉛　はナマリ　を主たる読み方とし，鉛筆，亜鉛　などを辞書化した方が有利．東西南北，　川（ガワとする），　付（ツキとする）なども同じ．

（2） スペースの利用

△形△　をカタチ，△型△　をカタ　とし，辞書熟語（形式，形成など）の他はガタ　と読ませる．

（3） サ変動詞等のミスマッチの防止

移す，ウツス　とした時，転移す，変移す　がテンウツス，ヘンウツス，となることを避けるため，転移，変移，など，通常は辞書なしで読める文字連も辞書化する必要があった．

（4） 助詞の利用

に 通っ，カヨッ を 通っ，トオッ
に 行っ，イッ を 行，オコナッ

など．

（5） 単位を表わす漢字

1発，3発，6発，……1分，2分，3分……0分 まで辞書化，さらに 2分子，（2プンシ），3分岐（3プンキ），も辞書化して区別させた．

以上のような，さまざまな工夫を要し，また，科学技術用語の思いの外の多彩さ，固有名詞も地名，会社名中心にかなり必要とした，などのため，79年5月の D_6 で，24,605語という大きさになった．

3. 変換結果，精度

78年3月にプログラムシステムが完成し，ほぼ同時に，全漢字を収録した辞書もでき上ったので，2カ月に1度の周期でテストランと辞書メンテナンスを行い，精度の向上を図った．

3.1 プログラム，オペレーション関係

プログラム 15本
コボル 3,830 ステップ
アセンブラ 1,550 ステップ
使用計算機， HITAC M-170
占有 CPU 約 120 KB （分かち書きタグを含むため）
ディスク 約80シリンダー，20 MB

3.2 テスト対象

JICST 理工学文献ファイル，全分野，校正完の分，6回のメンテナンスでランさせた．

総タイトル数，86,644 タイトル（年間の24%）
総漢字数 933,738 漢字
（平均 10.7字／タイトル）
平均全文字数 34.7字／タイトル

3.3 変換結果

タイトルの変換例は，**図-1** のとおりである．D_4 の段階で抄録文に適用してみたが，大略良好であった（**図-2**）．

辞書メンテナンスによる精度の向上と，処理速度の関係を**図-3** に示す．

なお，分かち書き精度は平均 98.6% で，100 タイトル中 1.4 の誤りが見られる．

一方，人手によりフリガナの精度は 99.64% 平均であり，D_3 の段階で機械変換が追い越した．

```
G010 金線超音波熱圧着ボンディングにおけるAl－Si2元蒸着薄膜のボンディング性
G011 金線超音波熱圧着ボンディング に おける Al－Si2元蒸着薄膜 の ボンディング性
G012 キンセンチョウオンパネツアッチャクボンディング に おける Al－Si2ゲンジョウチャクハクマク の ボンディングセイ
H010 キンセン チヨウオンパネツ アツチヤクボンデイング ニ オケル AL－SI 2ゲン ジヨウチヤクハクマク ノ ボンデイングセイ
G010 島状金属薄膜におけるトンネル電流と熱電子放射電流
G011 島状金属薄膜 に おける トンネル電流 と 熱電子放射電流
G012 シマジョウキンゾクハクマク に おける トンネルデンリュウ と ネツデンシホウシャデンリュウ
H010 シマジヨウ キンゾクハクマク ニ オケル トンネルデンリユウ ト ネツデンシホウシヤデンリユウ
G010 ケイ酸塩ガラス内のNd3+の吸収及び発光特性の組成依存性
G011 ケイ酸塩ガラス内 の Nd3+ の 吸収 及び 発光特性 の 組成依存性
G012 ケイサンエンガラスナイ の Nd3+ の キュウシュウ オヨビ ハッコウトクセイ の ソセイイゾンセイ
H010 ケイサンエンガラス ナイ ノ ND3+ ノ キユウシユウ オヨビ ハツコウトクセイ ノ ソセイイゾンセイ
G010 Ⅳ族，Ⅴ族の電子構造と結晶構造
G011 Ⅳ族，Ⅴ族 の 電子構造 と 結晶構造
G012 Ⅳゾク，Ⅴゾク の デンシコウゾウ と ケツショウコウゾウ
H010 4ゾク，5ゾク ノ デンシコウゾウ ト ケツシヨウコウゾウ
G010 極低温におけるマグネトフォノン共鳴とフォノンアシステッドサイクロトロン共鳴
G011 極低温 に おける マグネトフォノン共鳴 と フォノンアシステッドサイクロトロン共鳴
G012 キョクテイオン に おける マグネトフォノンキョウメイ と フォノンアシステッドサイクロトロンキョウメイ
H010 キヨクテイオン ニ オケル マグネトフオノン キヨウメイ ト フオノンアシステツド サイクロトロンキヨウメイ
G010 自動組立のドクメンテーションによる低コストでのより大きな効果
G011 自動組立 の ドクメンテーション に よる 低コストでの より 大きな 効果
G012 ジドウクミタテ の ドクメンテーション に よる テイコスト での より オオキな コウカ
G010 固定式締付ねじ Ⅰ 用途
G011 固定式締付ねじ Ⅰ 用途
G012 コテイシキシメツケねじ Ⅰ ヨウト
G010 はんだ付用フラックス Ⅱ
G011 はんだ付用フラックス Ⅱ
G012 はんだヅケヨウフラックス Ⅱ
G010 自動検査と試験における新しい機械化組立
G011 自動検査 と 試験 に おける 新しい 機械化組立
G012 ジドウケンサ と シケン に おける アタラシイ キカイカクミタテ
G010 英国における空港内鳥衝突対策について
G011 英国 に おける 空港内鳥衝突対策 に ついて
G012 エイコク に おける クウコウナイトリショウトツタイサク に ついて
```

図-1 タイトルの変換例（Hセグメントは人手によるもの）

人手によるミスは，機械のミスとはパターンが異なり，段落，誤読，フスヌやク，コとユの誤りなどが多い．

計算機時間に関しては，分かち書きはすべてコア内

G010 昭和53年1月14日に，伊豆大島，伊豆半島東南部を中心に，かなり大きな地震が発生した。被害は伊豆半島東南部で大きかったので，建設省建築研究所で調査団を現地に派遣した。被害の概要として，がけ崩れ，ＲＣ造の被害，鋼構造の被害，木構造及び二次部材の被害の概要について報告した

G011 昭和53年1月14日に，伊豆大島，伊豆半島東南部を中心に，かなり大きな地震が発生した。被害は伊豆半島東南部で大きかったので，建設省建築研究所で調査団を現地に派遣した。被害の概要として，がけ崩れ，ＲＣ造の被害，鋼構造の被害，木構造及び二次部材の被害の概要について報告した

G012 ショウワ53ネン1ガツ14ニチ ニ，イズオオシマ，イズハントウナンブ ヲ チュウシン ニ，カナリ オオキナ ジシン ガ ハッセイシタ。ヒガイ ハ イズハントウナンブ デ オオキカッタ ノデ，ケンセツショウケンチクケンキュウショ デ チョウサダン ヲ ゲンチ ニ ハケンシタ。ヒガイ ノ ガイヨウ トシテ，ガケクズレ，ＲＣゾクリ ノ ヒガイ，コウコウゾウ ノ ヒガイ，キコウゾウ オヨビ ニジブザイ ノ ヒガイ ノ ガイヨウ ニ ツイテ ホウコクシタ

G010 濃尾平野の地盤沈下の現状について述べるとともに，東海三県地盤沈下調査会および濃尾平野で現在とられている対策のうち，代表的な堤防かさ上げ強化，地下水揚水規制，代替用水の供給及び水源の開発についてのべる。さらに代替用水供給及び水資源開発の問題点についても報告

G011 濃尾平野の地盤沈下の現状について述べるとともに，東海三県地盤沈下調査会および濃尾平野で現在とられている対策のうち，代表的な堤防かさ上げ強化，地下水揚水規制，代替用水の供給及び水源の開発についてのべる。さらに代替用水供給及び水資源開発の問題点についても報告

G012 ノウビヘイヤ ノ ジバンチンカ ノ ゲンジョウ ニ ツイテ ノベル ト ともに，トウカイサンケンジバンチンカチョウサカイ および ノウビヘイヤ デ ゲンザイ とられて いる タイサク の うち，ダイヒョウテキ な テイボウかさアゲキョウカ，チカスイヨウスイキセイ，ダイタイヨウスイ の キョウキュウ オヨビ スイゲン の カイハツ に ついて の べる。さらに ダイタイヨウスイキョウキュウ オヨビ ミズシゲンカイハツ の モンダイテン に ついて も ホウコク

図-2 抄録文の変換例

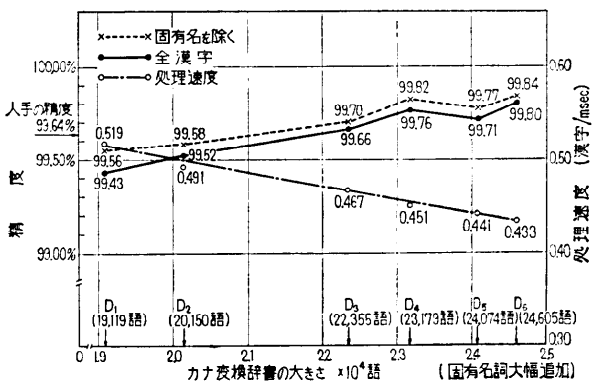

図-3 K-KACS 精度の向上

処理のため極めて高速で，3,000 タイトルの処理で CPU 時間10～20秒程度，漢字—カナ変換も例外熟語を持たない漢字はコア内に持ち，例外熟語も，テーブルサーチにより，概当する親漢字以下の全熟語を一挙にディスクからコアに転送できるため，高速であり，全処理時間も，3,000 タイトル，3万漢字で，CPU時間60～70秒，実ラン時間35～40分程度である．

辞書による処理については，全漢字の個別の統計はとっていないが，処理タイプ別の集計では以下のとおりである．

例外熟語を持たない漢字の処理	6.8%
辞書サーチした場合	93.2%
・そのうち辞書マッチ処理	14.1%
・アンマッチ処理	79.1%
・・そのうちA処理（一義的）	65.1%
・・そのうちB処理（音訓判定）	14.0%

すなわち，辞書を全く持たなくても，86% 程度は何とか正しく変換できる．これは，文献 1) の精度ともほぼ一致するが，実用のためには十分でない．

4. JICSTにおける実用と，費用効果分析

一部固有名詞に問題が残るが，精度，処理時間共に問題が少ないので，79年4月分からの JICST 理工学ファイルに適用し，人手によるフリガナが不要になった他，英，独，仏以外の全外国語文献に適用し，日本語カナ文字タイトルを可能とした．JOIS オンライン検索システムの回答書の改善例を**図-4** に示す．

費用の節減は，日本文フリガナの入力費が不要になった分が，直接に年間190万円程度であるが，71年にフリガナづけを外部委託した時に料金を 4.3% up しているので，今日の料金をもとに試算すると，さらに200万円，合計390万円程度の人力を節約したことになる．

一方，JICST で処理している年間36万件のすべてにフリガナをふり，入力するとなると，抄録の料金 up が1,400万円，入力パンチ代が1,300万円を要するが，81年からの JOIS II では，K-KACS により全タイトルカタカナ化を行うことに決定されたので，年間2,700万円分の節約をすることになり，また5年分のバックファイルにも適用するので，その分，1億3,500万円の仕事を

```
     S: ｻｰﾋﾞｽ ｵ ｶｲｼ ｼﾏｽ    1979.05.31   13:13:39
        ﾃﾞｰﾀﾍﾞｰｽ   JICST     ( [illegible].01 - 1979.06 )     436,869 ｹﾝ
        ｶｲｼ ﾊﾞﾝｺﾞｳ      13
[ 1] U: ｷｶｲﾎﾝﾔｸ
     S:        10 ｹﾝ
[ 2] U: ｼﾞﾄﾞｳｼｮｳﾛｸﾎｳ
     S:         6 ｹﾝ
[ 3] U: ｼﾞﾄﾞｳｻｸｲﾝﾎｳ
     S:        47 ｹﾝ
[ 4] U: 1*2*3
     S:        [illegible] ｹﾝ
[ 5] U: L/3, 4: N+DE+FR
     S:        10 ｹﾝ
[ 6] U: [illegible]
#001
CC= 025.3/.4, (FA020708)
CN= B790[illegible]
TI= [illegible]
AU= POPOV EH V
JN= R255AA    , IZV AKAD NAUK SSSR TEKH KIBERN
UN= NO.5, PAGE.84-90, 78
CI= (1) (A1) (RU       ) (SUN) (  , 4,  8)
KW= [illegible]; ﾛｼｱｺﾞ

#002
CC= 61:5/6, (FA020409)
CN= B79030674
TI= (ﾃｷｽﾄ ﾉ ﾌﾞﾝｾｷ ｵﾖﾋﾞ ﾄｳｺﾞｳ [illegible] ｼﾞﾄﾞｳｼｮｳﾛｸｼｽﾃﾑ ﾉ ｻｸｾｲ)
AU= KARASEV S A
JN= R272AA    , NAUCHNO TEKH INF 2
UN= NO.5, PAGE.22-28, 78
CI= (1) (A1) (RU       ) (SUN) (  , 6,   )
KW= ｼｮｳﾛｸ; ｼﾞﾄﾞｳｼｮｳﾛｸﾎｳ; ﾃｷｽﾄｼｮﾘ

#003
CC= 002.5, 025.3/.4:800/801, (FA0[illegible]01), (FA0209[illegible]4)
CN= B79030727
TI= (ﾌﾞﾝﾙｲ ｻｸｲﾝｱﾙｺﾞﾘｽﾞﾑ ﾉ ｼﾞﾄﾞｳｻｸｾｲ ﾉ ｶﾀﾞｲ(ｹﾞﾝｺﾞ ﾉ ﾌﾞﾝﾎﾟｳ ﾉ ｶｲｾｷ))
AU= SEREBRYANYJ A I
JN= R272AA    , NAUCHNO TEKH INF 2
UN= NO.5, PAGE.[illegible]-16, 78
CI= (1) (A1) (RU       ) (SUN)
KW= ｻｸｲﾝｺﾞﾊﾞﾘﾝｸﾞ; ｻｸｲﾝｻｸｾｲ; ｼﾞﾄﾞｳｻｸｲﾝﾎｳ; ｼﾞｮｳﾎｳｹﾝｻｸ; ｼｿｰﾗｽ

#004
CC= 025.3/.4:800/801, (FA020801)
CN= B78121265
TI= "PONIMANIE" FRAZ OGRANICHENNOGO RUSSKOGO YAZYKA.
AU= SAKLANOV V M; POPOV EH V
JN= R255AA    , IZV AKAD NAUK SSSR TEKH KIBERN
UN= NO.4, PAGE.66-74, 78
CI= (1) (A1) (RU       ) (SUN) (  , 3,  4)
KW= ｶｲﾜｶﾞﾀｼｮﾘ; ｷｶｲﾎﾝﾔｸ; ﾘｶｲ; ﾛｼｱｺﾞ

#005
CC= 002.6+02+061.66, (FA020502)
CN= B78060920
TI= VSESOYUZNYJ TSENTR PEREVODOV.
AU= GERASIMOV V N
JN= R187CA    , NAUCHNO TEKH INF 1
UN= NO.11/12, PAGE.43-45, 77
CI= (1) (D3) (RU       ) (SUN)
KW= ｷｶｲﾎﾝﾔｸ; ｼﾞｮｳﾎｳｾﾝﾀ; ｿﾋﾞｴﾄ; ﾎﾝﾔｸｻｰﾋﾞｽ

#006
CC= 025.3/.4:800/801, (FA020801)
CN= B78050061
TI= ANALIZ NEIZMENYAEMYKH OPREDELITELEJ K SUSHCHESTVITEL'NYM.
AU= KNORINA L V
JN= R272AA    , NAUCHNO TEKH INF 2
UN= NO.2, PAGE.20-27, 77
CI= (1) (A1) (RU       ) (SUN) (  , 2, 15)
KW= ｷｶｲﾎﾝﾔｸ; ｹﾞﾝｺﾞｶﾞｸ; ｺｳｺﾞｳｺﾞ; ﾛｼｱｺﾞ

     S: ｼｮｳﾘｮｳ ｵﾜﾘﾏｼﾀ                JICST-COPYRIGHT
```

79年4月からの改善の分（#001–#003）

従来の出力（#004–）

図-4 オンライン回答書の改善例

することになる．

これに要する費用は，システム作成費は100万円足らず，人件費は，時間外の任意作業が主のため僅少，計算機料金も年間 80 時間で，機器構成を大きく占有しないため130万円程度である．また，全タイトルの分かち書き，フリガナづけの処理を行えるために，タイトルからのキーワード自動抽出も可能となり，目下その準備中である．

この他 K-KACS は，JICST における他の場面，例えばシソーラス作成における入力時の人手によるフリガナミスのチェックにも活躍し，目検では発見しにくい誤りを見つけ出して担当者を援助している．

なお，本研究は，JICST の中でも情報部という現場で主として時間外に行われ，それが実務に採用された．これは JICST はじまって以来のことであるが，言語研究は，ある程度までカードと鉛筆のみで可能であることも示されたと思う．

参考文献

1) 田中章夫：漢字かなまじり文を全文カナ書き・ローマ字書きに変換するシステムについて，国立国語研究所報告 (34) pp. 17-22 (1969)．
2) 堀強：日本語情報の漢字カナシステム，第12回情報科学技術研究集会発表論文集， pp. 171-187 (1975)．
3) 板山和彦，荒木啓介：JICST 理工学文献ファイルの文献和文標題の漢字一カナ変換用の辞書作成の試み，第 13 回情報科学技術研究集会発表論文集，pp. 219-227 (1976)．
4) 板山和彦，荒木啓介，佐原 卓，坂上安彦，日夏健一：JICST 理工学文献ファイル専用漢字一カナ変換システム (K-KACS) について，情報管理 Vol. 21, No. 5, pp. 365-378 (1978)．
5) 板山和彦，荒木啓介，坂上安彦，佐原 卓，日夏健一：JICST 理工学文献ファイル専用，漢字一カナ変換システム K-KACS による実験，第15回情報科学技術研究集会発表論文集， pp. 169-173 (1978)

(昭和54年6月4日受付)