参考論文

『計量国語学』収録論文

真田　治子

参考論文

『計量国語学』収録論文

真田　治子

目次

（１）日本語と西欧語の基本語彙の対照研究

（『計量国語学』第19巻第４号　1994年　収録）

計量国語学第十九巻第四号（Mathematical Linguistics, vol.19 no.4）1994年

調査報告

日本語と西欧語の基本語彙の対照研究

真田　治子　（学習院大学大学院）

ディスクリプタ：　基本語彙　日本語　ドイツ語　フランス語　スペイン語

0.　研究の目的と調査の概要

各国語において作成されている基本語彙表は，その意味的構成が語彙表によって異なり，これらの語彙表を重ね合せてみると，意味分野によって差異があることがわかる．このような構成の差異がどの意味分野に出るか，どの程度の大きさであるかは，言語本来の性格や語彙表の目的の違いに起因すると推測できる．特に対象言語と自国語の差異を理解することは，基本語彙表を使用する学習者にとって効果的であると考える．さらに，基本語彙表の設計段階では，その目的によって意味的構成のウェイトをどこに置くかを，予め考慮すべきであろう．

上記の理由から，各国語の基本語彙の意味的構成の差異を明らかにしたいと考え，国立国語研究所『日独仏西基本語彙対照表』［文献 7］を用いて，日本語基本語彙と外国語基本語彙との比較を行った．

この『日独仏西基本語彙対照表』は，国立国語研究所『日本語教育のための基本語彙調査』［文献 6］にあげられた約二千語（以下「基本二千語」）とそれを含む約六千語（以下「基本六千語」）の日本語基本語彙と，独語，仏語，西語の基本語彙を，国立国語研究所『分類語彙表』［文献 1］の意味分類番号に基づいて比較対照させたものである．「基本二千語」「基本六千語」の，各意味分類番号への配置は『日本語教育のための基本語彙調査』に見出し語とともにあげられた番号によっているが，外国語基本語彙の，意味分類番号への配置は『日独仏西基本語彙対照表』独自に行われている．

この対照表に日本語基本語彙として用いられた「基本二千語」「基本六千語」は，まず「第一次専門家判定」として，『分類語彙表』［文献 1］を判定材料に，「留学生等外国人の日本語学習者が，専門領域の研究または職業訓練に入る基礎としてはじめに学習すべき日本語の一般的・基本的な語彙について妥当な標準を得る」という方針のもとに，日本語教育・国語学・言語教育等の専門家22人が判定委員となって語群を抽出，更に第一次判定結果を検討し，判定の偏りや判定材料の不備による問題点を修正して（第二次選定），選び出された語彙であるという［文献 6］．

また，この対照表の外国語基本語彙の資料は以下によっている．

・独語『ドイツ基本語辞典』［文献 4］

SANADA Haruko (Gakushuin University) — Comparative Study of Basic Vocaburaries of Japanese and European Languages

・仏語『フランス基本語辞典』［文献 2］
・西語『スペイン基本語辞典』［文献 5］

『フランス基本語辞典』［文献 2］は，「ある程度の教養は備えていてもフランス語の知識が広くない外国人を特に対象とするもので，書き言葉にも話し言葉にも使われる語から成っている」辞書としてフランスで出版され，さらに「初級・中級程度の，使いやすくしかも内容的に現代フランス語にじゅうぶん対処しうる仏和辞典を学習者の手もとに届けたいという意図」［文献 2］から，日本で翻訳されたものである．原著者の語彙選定の目的が，外国人の大人の学習者がその国で言語を学習するために，その国で編集されたという点が「基本二千語」「基本六千語」の選定の状況と似ている．

『ドイツ基本語辞典』［文献 4］は，日本人が中心となって編集し，『フランス基本語辞典』と同じ出版社から出された日本人向けの辞書であるが，「まえがき」には，「本書はドイツ語の語彙のなかからもっとも重要かつ基本的と思われるもの5300余語を選んで，これを辞書の形にまとめたものである．（中略）本辞典の語彙選択にあたっても，最終的には編者たちの主観的な判断にもとづいて決めざるをえなかった」とあり，どのような学習者を対象としているか，選定の方針はどのようであったかについては明記されていない．

『スペイン基本語辞典』［文献 5］は日本人と外国人の共同編集で，「まえがき」には上記のフランス語，ドイツ語，そしてロシア語，英語の辞書の「（前略）姉妹編として，同じ理由と目的をもって企画され，同じ特長と体裁を備えている．（中略）初級・中級程度の学習者にとって必要かつ十分な基本語5000を選」んだとある．選定にあたっては「その基準を次の２書に求めた．」としてマドリッドで出版された『Vocabulario Usual, Comunity Fundamental 』とオランダで出版された『Frequency Dictionary of Spanish Words』の２冊をあげている．また，「参考書」として４冊の洋書があげられている．この「まえがき」からは対象となる日本人学習者が例えば現地生活を予定している社会人であるか試験のために学習する学生であるかなどは明らかでないが，スペインで出版された基本語彙表やスペイン語の使用度数を意識して作られたことは推測できる．

このように編集方針に若干違いはあるが，上記３冊は姉妹編として出版されているので，今回の比較に際して大きな問題となるような不公平はないだろうと考えた．

『日独仏西基本語彙対照表』［文献 7］の構成は次のようになっている．

各項目は，それぞれの訳語形に与えられた意味分類の番号の順によって配列されている．ほぼ中央の欄に「訳語形」を示し，その右側にドイツ語，フランス語，スペイン語の順にそれに対する語形と，それぞれの品詞が示されている．「訳語形」の左には，それに対応する，「日本語教育のための基本語彙調査」で設定された基本語彙としての語形が「日本語」として示され，その左には，それらの基本度「R」として示されている．全体の配列の第１キーとなった分類番号は「分類番号」という見出しで左端の欄に示されている．（中略）

（１）「分類番号」

「分類語彙表」の番号で示してある（中略）．（中略）一つの語形に対して，「分類語彙表」で二つ以上の「番号」が与えられている場合は，その項目に対して複数の意味番号が与えられ，本語彙のそれぞれの場所へ配分されることになる．

（２）「訳語形」

訳語形は，（中略）それぞれの辞書で訳語として示されているものである．表記はなるべくもとの辞書によったが，３言語をそろえる便利のため，意味の本質に関わらない範囲内で若干の変更を加えたものもある．（中略）

（３）「日本語」

「日本語」として示した欄は，二つの役割を持っている．

ひとつは，「訳語形」に「分類番号」を与える際の根拠を示すことにある．（中略）

もう一つの役割が，日本語の基本語彙，「基本二千」「基本六千」の語形を示し，さらには，「六千」には入らないが「分類語彙表」には入っている，ということを示すものである．（中略）

（４）「R」（中略）

（５）「ドイツ語」，「フランス語」，「スペイン語」

それぞれ，ドイツ語，フランス語，スペイン語の語形を示す．計算機処理の都合上，すべて大文字を用いざるを得なかった．（中略）

（６）「品詞」（中略）

（７）その他

（中略）「日本語」の欄で「＊」としたものは語形がないことを示す．（中略）各外国語の欄で「＊＊」をもって示したものは，それぞれ該当する語形のないことを示す．

（国立国語研究所『日独仏西基本語彙対照表』［文献 7］より）

この対照表を用いて，以下の角度から検討した．

- 「基本二千語」・「基本六千語」にあって，三つの外国語のどれにも訳語の対応がない日本語

「基本二千語」「基本六千語」にあって，その語が独語，仏語，西語の三つの外国語の基本語彙の訳語のどれにも使われていないものを抽出した．これらの語は，西欧語圏に共通な，社会生活に最低限必要な語と違って，日本文化に大きく依存していたり，あるいは西欧語と異なる日本語という言語体系によっていたりするものと予想される．西欧語圏の日本語学習者が理解に注意を払うべき語とも解釈できる．

- 訳語が日本語基本語彙にない独語，仏語，西語の各基本語彙

外国語基本語彙にあって，訳語が日本語基本語彙にないものを抽出した．訳語を日本語と外国語の対照のキーワードにすることは，その基本語彙辞典の編訳者がどのような訳語をつけるかに依存しているという危険性もあるが，ここでは，訳語はある程度日本語の位相差を考慮してつけられているという仮定を前提に，日本語基本語彙との比較を試みた．例えば仏語のmèreには「母（親）」という訳語，maman には「おかあさん，ママ」という訳語がつけられている［文献 2］．また，訳語の形が１単語であるか，複合語であるか，短い文のような表現であるか，という点にも注目した．複合語や表現の場合，それらが日本語基本語彙のみから成り立っていれば，日本語基本語彙と１対１の単語単位で対応しているものに準じて理解が可能だと考えられる．「基本二千語」「基本六千語」には，その作成方針上，助詞・助動詞が含まれていないが，ここでは助詞・助動詞は理解の前提と仮

定した．ここで抽出された語は，西欧文化特有のものを当然含むであろうが，その他に，日本人の一般的認識では基本的でない概念の語で，西欧では比較的よく使用する語，基本的と考えられている語もあると思われる．外国語学習者が，国の違いによる「意識のずれ」から，見落とすことがないよう注意すべき語といえる．

今回の調査にあたって，国立国語研究所『日独仏西基本語彙対照表』のデータをフロッピー化し，修正を加えたものを，本書の担当者であった筑波大学の高田誠氏の好意により1990年に貸与していただくことができた（以下これをフロッピー版「日独仏西基本語彙対照表」とする）．書籍版の対照語彙表［文献 7］とフロッピー版の対照語彙表では，調査結果に書籍版の校正ミスと思われる差異がいくつか見られた．今回の分析はすべてフロッピー版を対象とした．また，このフロッピー版「日独仏西基本語彙対照表」の意味分類番号別，言語別の語数の集計表が，高田誠『基本語彙の対照研究』［文献 8］にあげられている．この表では，『日独仏西基本語彙対照表』の場合と異なり，「体の類」「用の類」「相の類」「その他」を区別する意味分類番号の１桁目が無視され，２桁目と３桁目を使って集計されている．これについて高田氏は次のように理由を述べている．

『日独仏西基本語彙対照表』では，意味の分類体系は『分類語彙表』の体系をそのままもちいた．すなわち，第１段階の大分類で「体」「用」「相」「その他」の四つに分類し，以下，大きな意味分野からだんだんに小さな分野へと４段階に分けられ，全体で５段階の下位分類がなされている．これらのカテゴリー分類は最大５桁の数字でコード化されていて，第１の桁は体・用・相・その他の別に用いられ，実質的な意味のコードは，二桁目からということになる．二桁目以下のコードは，それぞれの意味にしたがって体・用・相・その他のあいだで相対応するように考えられているが，桁が下がるにしたがってその対応は崩れてくる．

この分け方は，日本語については一応理にかなっているが，他の３言語に関しては具合いの悪いところが少なくない．たとえば，「漢語＋スル」のサ変動詞は漢語語幹で切られ，「体」として分類される．したがって，外国語のほうで動詞であっても語釈で漢語サ変があてられていると「体」すなわち，名詞という大分類の中にいれることになってしまう．同じように，形容詞あるいは副詞の語釈が，「漢語＋シタ」とか「漢語＋シテ」のような形であたえられていると，これらも「体」のところに置くことになる．したがって，体・用・相・その他という品詞分類を指向したカテゴリー分けは，外国語を考慮にいれた分類ではあまり適当ではないということになる．すなわち，実質的な意味分類である二桁目以下のコードによる分類が考察の対象となるというわけである．

そこで，本表では，体・用・相・その他の別は無視し，二桁目と三桁目の２桁ですべての意味分野をまとめて示した．四桁目以下は，意味が細分化しすぎ全体が見えなくなるおそれがあることや，体・用・相で対応のずれが大きくなってしまうことなどからすべてまとめてしまうことにした．したがって，表に示した意味分野の２桁のコードの第１桁は，１から５までの５分類で，それぞれ『分類語彙表』の分類カテゴリーに対応している．すなわち，１：抽象的関係，２：人間活動の主体，３：人間活動ー精神および行為，４：生産物および用具，５：自然物および自然現象に相当する．

１，３，５については，体・用・相のそれぞれ対応するものをふくみ，２，４は体のみである．第２桁は，『分類語彙表』ではとくに見出しはついていないが，本表では，便宜的に見出しを付けて示した．「その他１」は，『分類語彙表』の「その他」に含まれるもので，接続語や間投詞など実質的意味をもたない項目である．「その他２」としたものは，独仏西の方にあってどうにも分類のしようのないもの，表記の違いで空見出しで示されているものなどである．

（高田誠『基本語彙の対照研究』［文献 8］より）

1. 日本語基本語彙の特徴

対照語彙表の中から，対応する外国語基本語彙を持たない日本語基本語彙を抽出した．これは，対照語彙表の「R（ランク）」の欄が「２」（「基本二千語」）か「６」（「基本六千語」）の日本語見出し語で，独語，仏語，西語の各欄がいずれも「＊」になっている語のことである．

対応する外国語基本語彙を持たない日本語基本語彙は，「基本二千語」「基本六千語」合わせて6818語のうち，1990語であった．高田氏の表と比較するために，これを意味分類番号の第２，第３桁で集計し，さらに各項目においてどのくらいの率で残ったか，項目間の残存率を比較するため，比率のｚスコアを算出した．「基本二千語」「基本六千語」6818語を母集団とした時，抽出された結果は1990語で，残存率は1990／6818＝ 29.19％となる．これを全体比率という．比率のｚスコアは，各項目の残存率が全体比率の「相似形」であることを期待値として，その期待値から実際の比率がどれくらいずれているかを，０を中心に，＋－で表した指標である（注１）．計算方法と計算例を表1.1にあげた．また，意味分類番号の第２，第３桁で集計した語数と算出した比率のｚスコアを表1.にまと

表1.1. 日本語基本語彙の比率のｚスコアの計算方法と計算例

全体比率 ：P　　P = W / N

W ：外国語に対応のない日本語基本語彙の語数

N ：「基本二千語」「基本六千語」の総語数

P = 1990 / 6818 = 0.2919

期待値 ：F　　F = n * P

n ：「基本二千語」「基本六千語」の各意味分類番号別の語数

意味分類番号（10:こそあど)の期待値の例

F = 98 * 0.2919 = 28.60

標準偏差 ：σ　　$\sigma = \sqrt{n * P * (1-P)}$

意味分類番号（10:こそあど)の標準偏差の例

$= \sqrt{98 * 0.2919 * 0.7081} = 4.50$

ｚスコア ：Z　　Z = f - 0.5 - F / σ

f ：各意味分類番号の、外国語に対応のない日本語基本語彙の語数

意味分類番号（10:こそあど)のｚスコアの例

Z = 28 - 0.5 - 28.60 / 4.50 = -0.245

表1.2.　外国語に対応のない日本語基本語彙の，意味分類番号別ｚスコア
（●：ｚスコアが +2.0以上，または -2.0以下のもの）

意味分類番号		語数
抽象的関係	(.10) こそあど	28
	(.11) 異同・関係	72
	(.12) 出現・存在	42
	(.13) 様相・性質	55
	(.14) 力	8
	(.15) 動き・変化・変形	138
	(.16) 時間	156
	(.17) 空間・場所	37
	(.18) 形	20
	(.19) 量・単位・程度	172
人間活動の主体	(.20) 自他	30
	(.21) 家族	19
	(.22) 相手・仲間・主客	3
	(.23) 人種・民族・階級	14
	(.24) 成員・職・地位	37
	(.25) 国・都市・地名	38
	(.26) 事務所・店	18
	(.27) 機関・機構	12
	(.28) 団体	3
精神および行為	(.30) 感情・思考・論理	182
	(.31) 言語	154
	(.32) 創作・芸術	32
	(.33) 生活・遊び・動作	121
	(.34) 身上・行為・活動	59
	(.35) 交わり・争い	37
	(.36) 支配・人事・待遇	36
	(.37) 所有・経済・取引	54
	(.38) 仕事・産業	46
生産物および用具	(.40) 物品・貨幣	10
	(.41) 資材	8
	(.42) 衣料	19
	(.43) 食料・薬品	30
	(.44) 住居	29
	(.45) 道具	33
	(.46) 器具・機械	25
	(.47) 土地・道足・施設	7
自然物及び現象	(.50) 光・色・音・におい	34
	(.51) 自然・物体・物質	47
	(.52) 天地	27
	(.55) 生物・植物	33
	(.56) 動物・魚・虫	19
	(.57) 体	23
	(.58) 生死・健康・病気	23
その他		0
計		1990

めた（注２）．意味分野名は高田氏に従った．各項目の母集団の数は高田氏の表によっている．一般に，データが正規分布をなしていれば，ｚスコアの -2.0 ～ +2.0 にデータの 95.45％は入るといわれるので，「極端に偏っている」という基準として今回は +2.0 以上，-2.0 以下の項目に注目した．

比率のｚスコアが +2.0 以上の項目は次の７つである．

意味分類番号	意味分野名		日本語基本語彙総語数	外国語に対応のないもの	ｚスコア
(.11)	抽象的関係	異同・関係	183	72	+2.941
(.16)	抽象的関係	時間	376	156	+5.190
(.19)	抽象的関係	量・単位・程度	432	172	+4.806
(.25)	人間活動の主体	国・都市・地名	87	38	+2.855
(.31)	精神および行為	言語	337	154	+6.607
(.32)	精神および行為	創作・芸術	72	32	+2.718
(.33)	精神および行為	生活・遊び・動作	290	121	+4.631

意味分類番号(.11）の（異同・関係）では「体の類」の「クラス，等級」を表す言葉が目立つ．自己と対物，対人の関係を絶対的でなく相対的にとらえ，常に全体の中での自己の位置を確かめたいという日本の思考方法の現れとも考えられる．また，「すると」「たとえば」など基本的な接続詞がこの意味分類番号に集まっていることもｚスコアを高くしている．

意味分類番号(.16）の（時間）には「毎朝」「今月」「昨年」「来週」など外国語では複数語で表現されると思われる語，元号，「元日」「盆」「暮れ」など一年のうちの特定の時を表す語，「上旬」「下旬」など月のうちのある時期を示す語，「夜明け」「夕暮れ」など一日のうちのある時間を示す語が数多く含まれている．日本語では，ある区切られた時間のうちの特定の部分を示すために，ただ単に「初め」「中ほど」「終わり」と呼ぶのではなく，個別の呼び方をきめ細かくおこなっていることがわかる．「相の類」にある語も，「前の時」を表すために「さっき」「さきほど」「先だって」，「後の時」を表すために「あくる」「翌」，という形で細かな言いかえがなされている．高田氏の先行論文［文献８］でもこの項目は「日本語がかなり高く」，「『今月』『今晩』（中略）といった表現が１語で示される点がやや特徴的であろうか．」としている．

意味分類番号(.19）の（量・単位・程度）では「ここのか」「ふたり」など数詞に関するものや「軒」「匹」「杯」など日本語に特徴的な「単位」の語が多い．高田氏の先行論文［文献８］では「（独語，仏語，西語と比べて）日本語の高い値がややめだつ．助数詞がここに含まれていることが大きく作用していると考えられる．」となっているが，後述する西語のように，西欧語は３言語ともここに数詞，助数詞を配しており，むしろ「単位」の方が日本語の高い値に貢献しているように思われる．

意味分類番号(.25）の（国・都市・地名）では日本の国内の地方を表す語（近畿，九州，山陰，山陽，四国，東海道，東北，北陸など）を網羅していることが目立つ．独，仏，西語の基本語彙の場合は，首都や大都市を含む地域名はあっても，国全土に及ぶすべての地

域名が含まれてはいなかった（注３）．

意味分類番号(.31)の（言語）では，「かな」「漢字」「口語」「感動詞」「形容動詞」「漢語」「外来語」「促音」「濁音」「送りがな」「ふりがな」「部首」などの文法用語が多い．また，「おっしゃる」「伺う」「申し上げる」など言語に関する敬語表現も見られる．

意味分類番号(.32)の（創作・芸術）では「和歌」「能」「落語」「歌舞伎」など日本文化に関連のあるものが集まっていた．

意味分類番号(.33)の（生活・遊び・動作）では，「就職」「受験」「休講」「出勤」「冬休み」「衣食住」「朝寝坊」など外国語では複数語に相当すると考えられるもの，「お参り」「花見」「剣道」「柔道」「すもう」など日本文化と関連のあるもの，「いただきます」「ごちそうさま」「お帰りなさい」「ただいま」「おかげさまで」「ご苦労さま」など日本独特のニュアンスを含んだ挨拶があった．高田氏もこの分野の日本語の値は高いとしている［文献８］．

総じていえば，日本文化に関する語，日本語という言語構造に関連する語，外国語では複数語に相当する語などのほかに，位置づけや時間に関する語など日本固有の概念や思考体系に基づく語が抽出されたと考えられる．

比率のｚスコアが -2.0 以下の項目は次の６つである．

意味分類番号	意味分野名	日本語基本語彙総語数	外国語に対応のないもの	ｚスコア
(.15) 抽象的関係	動き・変化・変形	652	138	-4.549
(.17) 抽象的関係	空間・場所	174	37	-2.382
(.30) 精神および行為	感情・思考・論理	755	182	-3.111
(.36) 精神および行為	支配・人事・待遇	174	36	-2.549
(.41) 生産物および用具	資材	70	8	-3.400
(.99) その他		158	0	-8.157

これらの分野は，最初に日本語基本語彙として配置された語のうち，多くの割合で外国語基本語彙と合致したということである．これらの項目の結果からさらに日本語基本語彙と外国語基本語彙の類似点を考察していくこともできるが，今回は特に両者の相違点について論じているので，このような両基本語彙の合致については別の機会に考える．

2. 外国語基本語彙の特徴

次に，外国語基本語彙にあって，訳語が日本語基本語彙にないものを抽出した．これにより，独語，仏語，西語の各の基本語彙の特徴が，よりはっきりすると考えた．

外国語基本語彙は，見出し語と訳語が１対１対応のものと，１対複数対応のものとがある．訳語が複数ついているものは，各訳語の意味分類番号に従って，１つの見出し語が複数箇所に配置されている．そこで，対照語彙表の見出し語を訳語の側から検討していく方法を採った場合には，同じ外国語見出し語でありながら，ある意味分類の訳語では日本語基本語彙と対応し，他の意味分類の訳語では対応しない，ということが起こる．今回の調

査ではこのような語は採用しないことにし，どのような訳語の観点からその見出し語を見た場合にも，日本語基本語彙と対応しない外国語基本語彙を選び出すことにした．独語，仏語，西語の各について，どのような訳語も対照語彙表の「ランク」の欄が「７」か「９」で，「基本二千語」「基本六千語」と一致しない語を抽出した．

この結果，抽出された語数は以下の通りである．（１見出し語が複数の訳語を持つことがあるが，「語数」はこの場合，訳語数を指す．また，「対照語彙表全体の語数」は，高田誠『基本語彙の対照研究』［文献 8］による．）

	抽出語数	対照語彙表全体の語数
独語	2361語	9775語
仏語	2862語	12322語
西語	2544語	17390語

この対照語彙表における「基本二千語，基本六千語，または分類語彙表の見出し語と訳語との一致の認定方法」については，表の「解説」の中で次のように触れられている．

> 全体の７割程度の項目については，形の上でもほぼ一致し，分類番号を与えることは容易であったが，約３割の項目については，単位の長さの点で「分類語彙表」のそれと合致しないわけである．これについては，（中略）基本的には，担当者の主観によって分類の所属を決定した．（中略）（体の類には）漢語サ変動詞の語幹が含まれている．従って，訳語が「漢語＋スル」として与えられているものは，すべて，「体」の類に入れざるを得ないことになる．（中略）「漢語＋させる」は「体」に入れ，さらに，文法的なアスペクト，動作の様態などが，「する」の活用，語尾変化で示されている漢語動詞も，同じく「体」に入ってくることになった．また，形容詞の訳語として「～した」，「～している」という形で，同じく漢語動詞の活用語尾を伴った訳語が与えられているものも，「体」に組み入れられることになった．
>
> （国立国語研究所『日独仏西基本語彙対照表』［文献 7］より）

このほか，仮名と漢字による異表記の見出し語も意味の変わらない範囲で「一致」していると認定されているように観察できた．

訳語が「基本二千語」「基本六千語」と一致しないものは，さらに，「給仕」「真珠」のように訳語が１語で成立しているもの，「宇宙＋飛行」「雪＋が＋降る」のように訳語が複合語や表現で成り立っているもの，の２つにわけることができる．このうち，複合語や表現の訳語の場合は，その造語成分が「基本二千語」「基本六千語」と一致していれば，類推によって，１語の訳語に準じた理解をすることがある程度可能だと思われる．そこで，複合語や表現の訳語を短単位（国立国語研究所『電子計算機による新聞の語彙調査』［文献 3］の「短単位の区切り方」を参照のこと）の造語成分にわけ，すべての造語成分が「基本二千語」「基本六千語」と一致する見出し語は，この抽出結果から取り除いた．助詞・助動詞は「基本二千語」「基本六千語」に含まれていないが，基本語彙の知識の前提となるものと考え，「基本語彙と一致している」と仮定した．この造語成分と基本語彙との突き合わせには，上記の，対照語彙表における訳語と「基本二千語」「基本六千語」との突き合わせと同じルールを適用した．例えば漢語サ変動詞はその語幹に注目し，漢字と

仮名の異表記は意味を考慮した上で認定をした．また，動名詞はそのままの形で見出し語がない場合，動詞との突き合わせを行った．但し，「腹をたてる」「口の堅い」など複合語の意味が各造語成分の意味の総和と異なる場合や，慣用表現の特殊な例については，一致しているとは見做さなかった．

抽出結果の語数を意味分類番号の第２，第３桁で集計した．合計は以下の通りである．

	再抽出語数		再抽出語数		再抽出語数
独語	1892語	仏語	2294語	西語	2075語

意味分類番号の第２，第３桁でまとめた各項目において独語，仏語，西語のうち，どれが特徴的に偏っているかを調べるため，項目ごとに独語，仏語，西語の各総数に対する比率を出し，比率のｚスコアを算出した．ある項目で，独語，仏語，西語のうち，特に比率のｚスコアが高いものがあれば，その言語の，日本語と対応しない基本語彙の総数に対して，特にその項目に偏って語が配置されているということになる．比率のｚスコアの計算方法と計算例を表2.1.にあげた．集計語数と算出結果は表2.2.にあげた（注４）．前章の日本語の場合と同じく，比率のｚスコアが +2.0 以上，-2.0 以下の項目に注目した．但し比率のｚスコアは，もともとデータが 0を中心に＋－対称分布をするよう考えられてい

表2.1.　日本語に対応のない独語，仏語，西語基本語彙の比率のｚスコアの計算方法と計算例

全体比率 ：P　　$P = W / N$

W ：ある意味分類番号における、日本語に対応のない独語、仏語、西語基本語彙の合計語数

N ：日本語に対応のない独語、仏語、西語基本語彙の合計語数

意味分類番号（.20:自他）の全体比率の例

$P = (13(独語) + 8(仏語) + 14(西語)) / (1892(独語) + 2294(仏語) + 2075(西語)) = 0.005590$

期待値　：F　　$F = n * P$

n ：日本語に対応のない独語（または仏語、西語）基本語彙の総語数

意味分類番号（.20:自他）の独語の期待値の例

$F = 1892 * 0.005590 = 10.58$

標準偏差 ：σ　　$\sigma = \sqrt{n * P * (1-P)}$

意味分類番号（.20:自他）の独語の標準偏差の例

$= \sqrt{1892 * 0.005590 * 0.994410} = 3.24$

ｚスコア ：Z　　$Z = f - 0.5 - F / \sigma$

f ：各意味分類番号の、日本語に対応のない独語（または仏語、西語）基本語彙の語数

意味分類番号（.20:自他）の独語のｚスコアの例

$Z = 13 - 0.5 - 10.58 / 3.24 = +0.593$

表2.2.　日本語に対応のない独語・仏語・西語の意味分類番号別のｚスコア

（G：独語，F：仏語，S：西語，

＊：ｚスコアが+2.0以上，または-2.0以下のもの）

意味分類番号		語数 独	仏	西	ｚスコア
	(.10) こそあど	17	9	8	FS　　G
抽	(.11) 異同・関係	40	41	26	S　　F　G
象	(.12) 出現・存在	33	29	21	S　F　　G
的	(.13) 様相・性質	28	35	35	GF S
関	(.14) 力	9	7	7	FS　G
係	(.15) 動き・変化・変形	173	166	177	F　　S　G
	(.16) 時間	58	60	75	F　G　　S
	(.17) 空間・場所	46	40	23	*S　　F　　G*
	(.18) 形	32	39	23	S　GF
	(.19) 量・単位・程度	63	102	123	*G　　F　　S*
	(.20) 自他	13	8	14	F　　SG
人	(.21) 家族	18	21	16	S　FG
間	(.22) 相手・仲間・主客	3	6	6	G　FS
活	(.23) 人種・民族・階級	66	83	128	*GF　　S*
動	(.24) 成員・職・地位	49	133	81	*G　S　　F*
の	(.25) 国・都市・地名	28	39	58	G　F　　S*
主	(.26) 事務所・店	32	26	28	F　S　G
体	(.27) 機関・機構	4	21	9	*G　S　　F*
	(.28) 団体	2	4	5	G　F　S
	(.30) 感情・思考・論理	256	287	263	FS　G
精	(.31) 言語	127	171	139	SG　F
神	(.32) 創作・芸術	22	30	34	G F　S
お	(.33) 生活・遊び・動作	79	96	89	GFS
よ	(.34) 身上・行為・活動	40	49	57	GF　S
び	(.35) 交わり・争い	53	54	37	S　F　G
行	(.36) 支配・人事・待遇	61	83	60	GS　F
為	(.37) 所有・経済・取引	56	37	51	*F　　S　G
	(.38) 仕事・産業	29	47	32	GS　F
生	(.40) 物品・貨幣	12	14	10	S　FG
産	(.41) 資材	29	35	22	S　GF
物	(.42) 衣料	20	48	26	G S　　F*
お	(.43) 食料・薬品	33	36	24	S　F G
よ	(.44) 住居	32	54	40	G S　F
び	(.45) 道具	50	68	51	S G　F
用	(.46) 器具・機械	27	45	26	S G　F
具	(.47) 土地・道足・施設	14	20	16	GS F
自	(.50) 光・色・音・におい	32	34	40	F　G　S
然	(.51) 自然・物体・物質	33	40	36	SGF
物	(.52) 天地	16	32	26	G　　S F
及自	(.55) 生物・植物	27	23	29	F　　SG
び然	(.56) 動物・魚・虫	24	16	24	F　　S G
現	(.57) 体	29	30	28	FS G
象	(.58) 生死・健康・病気	44	35	37	F　S　　G
その他		33	41	15	*S　　G F
計		1892	2294	2075	

るため，比較しているデータの中に比率のｚスコアが極端に高いものがあると，その影響で相対的にｚスコアが低くなるデータも現れる可能性がある．特に今回の調査では１項目（意味分類番号別の１意味分野）に独語，仏語，西語と３つしかデータがないので，比率のｚスコアが -2.0 以下の項目は他の言語にｚスコアが高いものがあるかどうかについて注意を払った．また，このｚスコアの値が高くても，もともとその言語ではその項目の語が多かったことも考えられるので，独語，仏語，西語の基本語彙全体についても各項目ごとに，意味分類番号の第２，第３桁で比率のｚスコアを出してみた．この２種類の比率のｚスコアは分母が異なるので直接比較することはできないが，前者の値が高い項目について，母集団となる語彙全体でもともと偏りのある項目かどうか検査する目安となると考えた．

2.1. 日本語基本語彙に対応のない独語基本語彙

日本語基本語彙に対応のない独語基本語彙について述べる．

比率のｚスコアが +2.0 以上の項目は次の項目である．

意味分類番号	意味分野名	語数	ｚスコア	母集団のｚスコア
(.17) 抽象的関係	空間・場所	46	+2.208	+2.755

意味分類番号(.17) の（空間・場所）では，「下方へ（HERAB）（HINAB）（HERUNTER）（HINUNTER)」「下方に向かって（ABWÄRTS)」「～の下方に（UNTERHALB)」「上方へ（HERAUF）（HINAUF）（EMPOR)」「～の上方に（OBERHALB）」など方向を示す語とその複合形が多くみられた．この項目の基本語彙全体の独語のｚスコアは+2.755で，仏語，西語に比較してもともと独語はこの項目へ多くの語を配していることがわかる．また，語成分の膠着による造語という独語の特性が表れているとも考えられる．

比率のｚスコアが -2.0 以下の項目は次の４つである．

意味分類番号	意味分野名	語数	ｚスコア	母集団のｚスコア
(.19) 抽象的関係	量・単位・程度	63	-2.692	-1.531
(.23) 人間活動の主体	人種・民族・階級	66	-2.035	-2.567
(.24) 人間活動の主体	成員・職・地位	49	-3.550	-3.558
(.27) 人間活動の主体	機関・機構	4	-2.119	-3.327

意味分類番号(.19)の(量・単位・程度）では，仏語のｚスコアは-0.401，西語は+2.835と，西語の値が高い．また，意味分類番号(.23）の（人種・民族・階級）では，仏語のｚスコアは-1.928，西語は+3.811と，西語の値が高い．これらの項目は西語の影響が強いと考えられるので，後述の西語の項を参照されたい．

意味分類番号(.24)の(成員・職・地位）では，仏語のｚスコアは+3.761，西語は-0.729と，仏語の値が高い．また，意味分類番号(.27）の（機関・機構）では，仏語のｚスコアは+2.285，西語は-0.827と，仏語の値が高い．これらの項目は仏語の影響が強いと考えら

れるので，後述の仏語の項を参照されたい．

2.2. 日本語基本語彙に対応のない仏語基本語彙

日本語基本語彙に対応のない仏語基本語彙について述べる．

比率のｚスコアが +2.0 以上の項目は次の３つである．

意味分類番号	意味分野名		語数	ｚスコア	母集団のｚスコア
(.24)	人間活動の主体	成員・職・地位	133	+3.761	+6.981
(.27)	人間活動の主体	機関・機構	21	+2.285	+1.206
(.42)	生産物および用具	衣料	48	+2.242	+2.603

意味分類番号(.24) の（成員・職・地位）では，「外科医（CHIRURGIEN）」「宣教師(MISSIONNAIRE)」「タイピスト（DACTYLO)」「外交官（DIPLOMATE)」「羊飼い（BERGER)」「配管工（PLOMBIER)」「将校（OFFICIER)」といった職業名のほか，何かをする人を表す「殺害者（MEURTRIER)」「創始者（FONDATEUR)」「亡命者（RÈFUGIÈ)」「立会人（TÈMOIN)」「納税者（CONTRIBUABLE)」「列席者（ASSISTANCE)」などの語があった．この項目の基本語彙全体の仏語のｚスコアは+6.981で，もともと基本語彙全体でみた時も，仏語はこの意味分野に多くの語を配していることがわかる．高田誠『基本語彙の対照研究』［文献 8］でも「この分野では，フランス語が職業，地位にある人を表す語形を数多くあげていることは注目してよい．」としている．但し高田氏は「これらの多くは派生形であり，語構成の上で日本語と異なってい」て，「日本語の語彙単位としては複合的であり独立した語彙項目としては立項されないためこのような差となってあらわれた」としている．しかし今回はこのような複合語を短単位にわけて調べ，それでもなお多くの職業や動作の種類を示す語成分は日本語の基本語彙に含まれていないことが明らかになった．つまりこれらは，日本語では複合語であるから立項されにくいだけでなく，日本語の基本語彙にもともと入っていない語の派生形であるといえる．

意味分類番号(.27) の（機関・機構）では，「公使館（LÈGATION）」「天文台（OBSERVATOIRE)」といった建物や「上院（SÈNAT)」「デモ隊（MANIFESTATION)」 といった組織の名前，「騎兵隊（CAVALERIE)」「憲兵隊（GENDARMERIE)」「参謀部（ÈTAT-MAJOR）」「歩兵隊（INFANTERIE）」「砲兵隊（ARTILLERIE）」など軍関係の組織名が特に多くみられた．この項目の基本語彙全体の仏語のｚスコアは+1.206で，一応正ではあるが特に多くはない．つまりこの母集団自身は特徴的ではないが，日本語基本語彙と対応しないものを数多く含んでいるということである．

意味分類番号(.42) の（衣料）では，「毛皮（FOURRURE)」「チョッキ（GILET)」 「半ズボン(CULOTTE)」「スカーフ(FOULARD)」「ベレー帽(BÈRET)」「サンダル靴(SANDALE)」「首飾り（COLLIER)」「腕輪（BRACELET）」など洋服やアクセサリーに関する語がみられた．この項目の基本語彙全体の仏語のｚスコアは+2.603で，独語，西語にくらべると多いといえる．つまりこの意味分野にも仏語は基本語としてもともと多くの語を配しており，ファッションという現代フランスの文化的背景を表していると考えられる．

比率のｚスコアが -2.0 以下の項目は次の項目である.

意味分類番号	意味分野名	語数	ｚスコア	母集団のｚスコア
(.37)	精神および行為　所有・経済・取引	37	-2.265	-1.820

意味分類番号(.37) の（所有・経済・取引）では，独語のｚスコアは+1.838，西語は+0.407であった．また，この項目の基本語彙全体のｚスコアは独語が-0.563，仏語が-1.820，西語が+1.887で，もともと仏語は基本語彙全体でみてもこの項目への語の配し方が少ない.

2.3. 日本語基本語彙に対応のない西語基本語彙

日本語基本語彙に対応のない西語基本語彙について述べる.

比率のｚスコアが +2.0 以上の項目は次の３つである.

意味分類番号	意味分野名	語数	ｚスコア	母集団のｚスコア
(.19)	抽象的関係　量・単位・程度	123	+2.835	+0.818
(.23)	人間活動の主体　人種・民族・階級	128	+3.811	+0.696
(.25)	人間活動の主体　国・都市・地名	58	+2.523	+0.360

意味分類番号(.19) の（量・単位・程度）では，数詞（TREINTA等)と，数詞に「の」がついた形（CUARENTA 等)，数に関連した「- 分の一（OCTAVO 等)」，「- 等分の（NOVENO等)」，「- 倍（の)（TRIPLE等)」「- 重（の)（DOBLE等)」の形，が多くみられた．この項目の基本語彙全体の西語のｚスコアは+0.818であった.

意味分類番号(.23) の（人種・民族・階級）では，「アステーカ族（AZTECA）」「イベリア人（IBERICO)」「グワテマーラ人（GUATEMALTECO)」「コロンビア人（COLOMBIANO)」「セビーリャ人（SEVILLANO)」などスペイン語圏を中心とした民族名や「皇帝（EMPERADOR)」「公爵（DUQUE)」「平民（PLEBEYO)」「法王（PAPA)」「大司教（PONTÍFICE）」などの社会的階級名が多くみられた．この項目の基本語彙全体の西語のｚスコアは+0.696であった.

意味分類番号(.25) の（国・都市・地名）でも，「パナマの（PANAMÈNO）」「メキシコの（MEJICANO)」「アンダルシーアの（ANDALUZ)」「コルドバの（CORDOBÈS)」「ベネズエーラの（VENEZOLANO）」などスペイン語圏を中心とした地域名が多かった．この項目の基本語彙全体の西語のｚスコアは+0.360であった.

以上の３項目について，意味分類番号第２，第３桁での基本語彙全体の西語のｚスコアをみると，基本語彙全体からみた，もともとの各項目への語の配置の仕方は特に多くはない．いずれの場合も，日本語の基本語彙と対応しない語を，独語，仏語に比べて多く含んでいるということになる.

比率のｚスコアが -2.0 以下の項目は次の２つである．

意味分類番号	意味分野名	語数	ｚスコア	母集団のｚスコア
(.17)	抽象的関係　空間・場所	23	-2.287	-1.515
(.99)	その他	15	-2.781	-1.162

意味分類番号(.17) の（空間・場所）では，独語のｚスコアは前述の通り+2.208（独語の項を参照），仏語は-0.070と，独語の値が高い．この項目は独語の影響が強いと考えられる．

意味分類番号(.99) の（その他）では，独語のｚスコアは+1.089，仏語は+1.392であった．この項目には訳語の付けられていない，冠詞や前置詞の縮約形，融合形，それから基本形を参照するよう指示した異形などが配置されている．西語のｚスコアの値が低いのは，西語の言語そのものとしての文法的特性と，例えばどのような異形を見出し語に立てるかなど，この西語基本語彙のもととなった『スペイン語基本語辞典』の編集方針に起因するものと思われる．

2.4.　独語，仏語，西語に共通して，日本語基本語彙に対応のない語

日本語基本語彙に対応のない独語，仏語，西語の訳語を突き合わせ，３言語に共通するものを抽出した．訳語数にして 162語あった．

月名，数詞，序数詞，「天使」「牧師」「ミサ」など宗教関係の語，「オペラ」「チップ」「シャワー」など西欧の文化を背景としている語が散見された．

2.5.　日本語基本語彙に対応のない，高頻度訳語

以上の，日本語基本語彙と独語，仏語，西語の訳語との突き合わせの過程で，どのような日本語が，訳語に頻出するかを観察した．日本語基本語彙にはないが，西欧語基本語彙の訳語に出てくる語のうち，使用度数が５回以上で複数言語にまたがってあらわれるのは以下の９語である．

		（内訳）			新聞の語彙調査	
	合計度数	独語	仏語	西語	全体順位	出現率
－所	25	10	11	4	302	.287
－師	20	6	9	5	1019	.089
不正	11	7	4		10844	.005
房	11		6	5	－	－
教徒	11		6	5	－	－
服従	10		6	4	－	－
相続	10		4	6	6615	.011
－類	9		4	5	1999	.045
節度	8	4	4		－	－

この９語に，国立国語研究所『電子計算機による新聞の語彙調査』［文献 3］の新聞での頻度順位を重ねてみる（全見出し数 13206）と，接尾辞である「- 所」「- 師」「- 類」は新聞の語彙調査の中でも比較的頻度が高い．つまり，これらの語は今は「基本二千語」「基本六千語」に含まれていないが，独語，仏語，西語の基本語彙の訳語として，複数の言語にまたがってよく出現し，新聞などでも割合よく使われる語であるといえる．このような語は，次回の「基本二千語」「基本六千語」の改訂の際には，考慮すべき語と考えられる．

3. おわりに

日本語，独語，仏語，西語の４か国語で，延べ 46219語の基本語彙を，国立国語研究所『日独仏西基本語彙対照表』［文献 7］の分類番号別配置の助けを得て，以上のように鳥瞰することができた．異なった文化，生活様式を持つ各国の言語の中から，それぞれの社会において特に基本的であると選ばれた語彙を対照した．これらの語彙は外国語学習者にとって最初に出会う語群であり，学習者は自国語での「基本的」か否かという先入観を捨てて，異文化を柔軟に受け入れなければならないであろう．さらに，学習者は自国語では基本的でない語を学ぶことを通して，その語を基本的だと考える他言語社会の文化的，歴史的背景，事情をもかいま見ることができるのである．

また，高田氏の先行論文［文献８］では，言語別に各意味分類番号の語数とその比率だけが示され，この比率が項目間で「やや低い」「かなり高い」という表現で比較されていたが，今回比率のｚスコアを用いることにより，比較の方法がより客観的になったと考えられる．

今回の調査を手がかりとして，他の言語の基本語彙との比較対照や，各国で用いられている小学校教科書などの語彙の比較調査をすすめ，言語にみられる文化的差異を細かく考察していきたいと考えている．

謝　辞

フロッピー版「日独仏西基本語彙対照表」を快く貸与下さった筑波大学の高田誠氏に謝辞を申し上げる．また，同データをNEC LANFILE 用からMS-DOS用へ変換する際にご協力いただいた日本電気株式会社第二官庁　相馬氏，朝山氏にも併せて謝辞を申し上げる．

注

注１）比率のｚスコアの精度を保つためにはある程度の数のデータが必要であるが，今回のデータを比率の標本分布とみなし，必要な標本数を求めた場合，仮にその精度を95％とすると，比率の標本分布が近似的に正規分布に従えば，1.96Ｓ（Ｓは標準誤差）が有意水準0.05より小さければよい．従って，

$$1.96S \leqq 0.05$$

今回のデータは無限母集団（言語全体）からの標本抽出と考えられるので，全体比率は今回の値と同じと仮定すると，

$$1.96\sqrt{\frac{p(1-p)}{n}} \leqq 0.05$$

（ｐは全体比率．ここでは0.2919．）
（ｎは標本数）

となり、これを解くと標本は 317.6以上とればよいことがわかる.

注２）この分布に対し，カイ二乗検定を用いて一様性の検定をしたところ，有意水準0.05で有意であるという結果が得られた.

注３）筆者の約半年のヨーロッパ留学中の経験では，その国がどのように行政区分され，どのように名付けられているかはその都度地図をみればよいことであって，最初に覚えるべきことではなかった．逆に国内を移動する際には主立った都市名を覚えておくことが安全確実に移動するための必要手段となった．これは経験的意見で，個人差もあろうが，あえて注記しておく.

注４）この分布に対し，カイ二乗検定を用いて一様性の検定をしたところ，有意水準0.05で有意であるという結果が得られた.

文献

1. 国立国語研究所(1964)『分類語彙表』秀英出版
2. ジョルジュ・マトレ（野村二郎・滑川明彦訳編）(1967)『フランス基本語辞典』白水社
3. 国立国語研究所(1970)『電子計算機による新聞の語彙調査』秀英出版
4. 岩崎英二郎，早川東三，子安美知子，平尾浩三，鉄野善資(1971)『ドイツ基本語辞典』白水社
5. 高橋正武，瓜谷良平，宮城昇，エンリケ・コントレラス(1972)『スペイン基本語辞典』白水社
6. 国立国語研究所(1984)『日本語教育のための基本語彙調査』秀英出版
7. 国立国語研究所(1986)『日独仏西基本語彙対照表』秀英出版
8. 高田誠(1991)『基本語彙の対照研究』文芸言語研究19　言語篇

（1993年5月7日受付，1994年1月14日再受付）

Descriptors and Abstracts

Report

COMPARATIVE STUDY OF BASIC VOCABULARIES OF JAPANESE AND EUROPEAN LANGUAGES

SANADA Haruko (Gakushuin University)

Descriptors: basic vocaburary; Japanese; German; French; Spanish;

Abstract:

The aim of this paper is to clarify the difference in semantic construction between Japanese and European basic vocabularies. It can be assumed that this difference results from the characteristics of each language and the purpose of selecting each vocabulary. Understanding this difference is important for learners of these languages and developers of basic vocabularies. The present paper analyzes a comparative table in "A Contrastive Study of the Fundamental Vocabulary of Japanese, German, French and Spanish" by The National Language Research Institute which includes 6818 Japanese words, 9775 German words, 12322 French words, and 17390 Spanish words.The results are as follows:

<Characteristic Semantic Areas>

Japanese: Relations; Time; Unit; Countries and Cities; Languages; Art; Living;

German: Place;

French: Positions; Functions; Clothes;

Spanish: Unit; Race; Countries and Cities;

（２）諸言語の基本語彙の有効性の比較

（『計量国語学』第18巻第７号　1992年　収録）

計量国語学第十八巻第七号〔Mathematical Linguistics, vol.18 no.7〕1992年

諸言語の基本語彙の有効性の比較

真田　治子　（学習院大学大学院）

ディスクリプタ：　基本語彙　日本語　英語
フランス語　スペイン語　有効性

0.　はじめに

一般に「基本語彙」として示されている語彙には色々なものがあるが，その「基本語彙」の有効性を調査研究することを試みた．今回は日本語，英語，仏語，西語をとりあげ，これらの基本語彙が実際の文章の延べ語数のうち，どれだけの割合をカバーするかという観点から有効性を比較・検証した．

日本語の基本語彙としては，今回の調査では国立国語研究所『日本語教育のための基本語彙調査』〔文献27〕に示された二段階の語群――約二千語（以下「基本二千語」）とそれらを含む約六千語（以下「基本六千語」）を用いた．これは「基本二千語」「基本六千語」が大人のための語彙であり，ある程度の生活語彙，知識語彙が含まれていて，現代日本社会の語彙を偏りの少ない形で反映していると考えたからである．

「基本二千語」「基本六千語」はまず「第一次専門家判定」として，国立国語研究所『分類語彙表』〔文献４〕を判定材料に，「留学生等外国人の日本語学習者が，専門領域の研究または職業訓練に入る基礎としてはじめに学習すべき日本語の一般的・基本的な語彙について妥当な標準を得る」という方針のもとに，日本語教育・国語学・言語教育等の専門家２２人が判定委員となって語群を抽出，更に第一次判定結果を検討し，判定の偏りや判定材料の不備による問題点を修正して（第二次選定），「基本二千語」「基本六千語」を得たという．

仏語の基本語彙としては，国立国語研究所『日独仏西基本語彙対照表』〔文献31〕に用いられた，ジョルジュ・マトレ『フランス基本語5000辞典』〔文献５〕を用いた．この辞

SANADA Haruko (Gakushuin University) — Comparison of Effectiveness of Various Basic Vocabularies

典は，「ある程度の教養は備えていてもフランス語の知識が広くない外国人を特に対象とするもの」［文献５］としてフランスで出版され，「初級・中級程度の，使いやすくしかも内容的に現代フランス語にじゅうぶん対処しうる仏和辞典を学習者の手もとに届けたいという意図」［文献５］から，日本で翻訳されたものである．この仏語語彙の原著の目的が，外国人の大人の学習のためにその国で編集されたという点で，「基本二千語」「基本六千語」の選定目的と一致することから，この語彙を使用した．また，この語彙が国立国語研究所『日独仏西基本語彙対照表』［文献31］に使用されており，今後の筆者の様々な調査との共通の目安となりうることも使用理由の一つである．

西語の基本語彙としては，同じく国立国語研究所『日独仏西基本語彙対照表』［文献31］に用いられた高橋正武他『スペイン基本語5000辞典』［文献11］を使用した．この語彙は前述のジョルジュ・マトレ『フランス基本語5000辞典』［文献５］と「同じ理由と目的をもって」［文献11］日本で編集され，同じく５０００語が選定されているので，仏語と同じ基準の語彙となりうると考えてこれを選んだ．

英語の基本語彙には，ジョルジュ・マトレ『フランス基本語5000辞典』［文献５］及び『スペイン基本語5000辞典』［文献11］と同じシリーズとして日本で出版されたＪ・Ｒ・ショー『ラダー英和基本語5000辞典』［文献12］を用いた．この辞典は仏語の場合と同様，「英語を外国語として使う人たちのため」［文献12］アメリカで編集され，日本で翻訳されたものである．この辞典は見出し語のうち，５０００語を１０００語ずつに区切って重要度「１」から「５」がつけてある．これを日本語の「基本二千語」「基本六千語」となるべく近い形で対応させるため，重要度「２」までの２０００語，「５」までの５０００語の二段階で調査した．ただし，この辞書では一つの見出し語に，名詞，動詞，副詞，形容詞といった派生語がみな含まれているので，実際の収容語数は５０００語よりかなり多いと考えられる．今回の調査では，派生語で下位の見出しに含まれているものも，重要度の付いた上位の見出し語と同等に扱った．

資料としては，国際連合で１９５９年に採択された「児童の権利に関する宣言」を用い，「表題」「前文」「第１条」～「第１０条」「児童の権利に関する宣言の普及」の部分を範囲とした．資料としてこれを選んだのは，同じ内容の各種外国語版が作成されているためである．日本語版の他に，国際連合の公用語である英語版，仏語版，西語版を入手した．

国立国語研究所『日独仏西基本語彙対照表』［文献31］では，この他独語が比較されていて，基本語彙として岩崎英二郎他『ドイツ基本語5000辞典』［文献10］が同じシリーズ

で出版されているため，独語の調査も行いたかったが，ドイツ語圏の国連参加加盟国でも，ドイツ語圏の国々により出資された国連内のドイツ語への翻訳部門（現在は解体されている）でも，資料のうちの「児童の権利に関する宣言の普及」の部分の翻訳を行っておらず，資料全文の入手が不可能だったので，独語版の調査は断念した．

1. 日本語版「児童の権利に関する宣言」における「基本二千語」「基本六千語」の有効性

日本語版「児童の権利に関する宣言」の一部を図１．a. に示す．

> 第１条
> 児童は、この宣言に掲げるすべての権利を享有する。すべての児童は、いかなる例外もなく、自己又はその家族について、人種、皮膚の色、性、言語、宗教、政治上その他の意見、国民的もしくは社会的出身、財産、門地もしくは他の地位のため差別を受けることなく、平等に前記の権利を享有することができる。

図１．a.　日本語版「児童の権利に関する宣言」の一部

ここでの語の区切りは，基本的には短単位[注]としたが，「基本二千語」「基本六千語」に含まれるものは一語とみなした．例えば「そのほか」という語は「基本六千語」に入っているので，ここでは一語としてある．また，後述する英語版，仏語版，西語版の「児童の権利に関する宣言」との比較のために，「名詞＋する」を一語，「名詞＋だ／な」を一語とし，名詞の部分が「基本二千語」「基本六千語」と対応するかどうかを調べた．例えば「出版する」「幸福だ(な)」は各一語とし，「出版」「幸福」が「基本二千語」に入っていることを確認した．「出版する」「幸福だ(な)」は各「publish （英語）」「 publier （仏語）」「publicar （西語）」，「happy （英語）」「 heureux （仏語）」「 feliz （西語）」の一語ずつと対応していると考えられるからである．さらに西欧語の前置詞などと語の区切りをなるべく揃えるために，連語として「おいて」「ついて」「として」「当たって」「という」をたてた．また，「基本二千語」「基本六千語」に接頭辞，接尾辞が含まれているため，「-的」「-年」「各-」「-等」「-上」「-感」「-性」「大-」

注　短単位―― 原則として，和語については最小意味単位２つまでの長さ，字音語については漢字２字まで，外来語については原語１語で構成されている，比較的短い単位．詳しくは国立国語研究所『電子計算機による新聞の語彙調査』［文献７］の「短単位の区切り方」を参照のこと．

「広-」を独立した要素として扱った.

この資料における異なり語数と延べ語数の関係は，以下の通りである.

	異なり語数		延べ語数	
・助詞・助動詞	22語	7.2%	402語	34.0%
・基本二千語まで	148語	48.7%	849語	71.9%
・基本六千語まで	224語	73.7%	1048語	88.7%
・全体	304語	100.0%	1181語	100.0%

これによれば，助詞・助動詞の知識があることを前提とすれば，「基本二千語」までででこの資料の７１．９％の語を，「基本六千語」までで８８．７％の語を読めることになる．また，助詞・助動詞は異なりの２２語（７．２％）で延べの３４．０％，つまり約３分の１をカバーしている.

横軸に異なり語数，縦軸に延べ語数をとり，その関係をグラフにした．作成したのは助詞・助動詞を含む，累積使用率のグラフ（図１.b.）である.「助詞・助動詞」「基本二千語」「基本六千語」「その他」，と語の種類が変わるごとに，新しい孤を描く線となった.このグラフに対し，Y = a X ** b（０＜b＜１．**はべき乗を示す）の式をモデルとして選び，最小二乗法により回帰曲線を算出したところ，

Y = 18.88 X ** 0.366

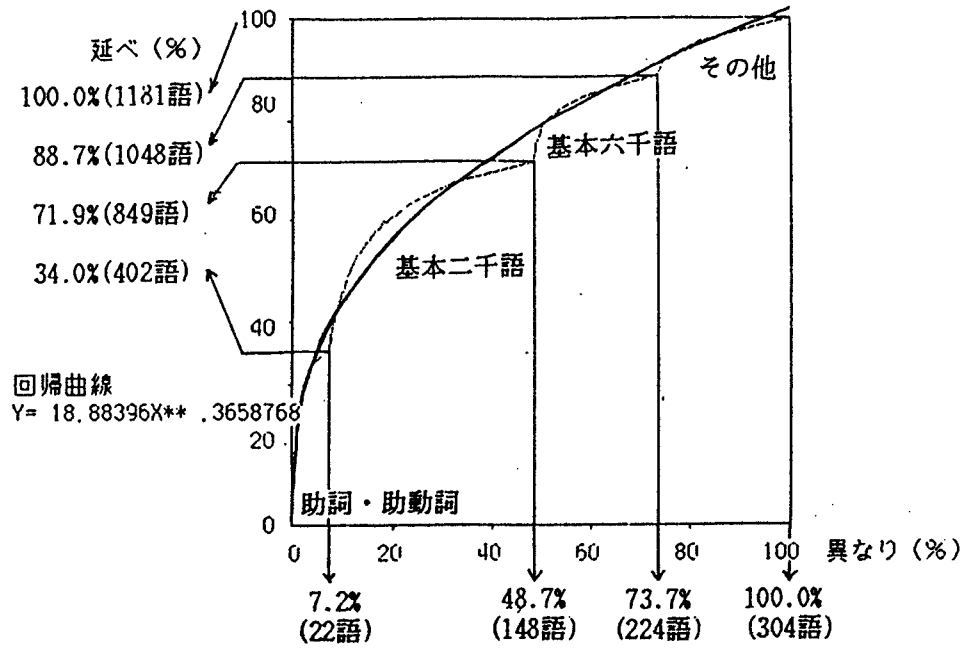


図１.b. 「児童の権利に関する宣言」 二千語 六千語 有効性 含助詞 助動詞 パーセント表示

となった.

助詞・助動詞の影響を除いて考えた場合，上表は以下のように書き換えられる.

	異なり語数		延べ語数	
・基本二千語	126語	44.7%	447語	57.4%
・基本六千語まで	202語	71.6%	646語	82.9%
・全体	282語	100.0%	779語	100.0%

この場合は助詞・助動詞を含めなくても，「基本六千語」までの範囲の２０２語で，この資料の８２．９％の語は読めるということになる.

また，一つの目安として，延べ語数を５０％カバーする点をさがしてみると，助詞・助動詞を含む場合で，異なり語数３５語（１１．５％），助詞・助動詞を除いた場合で，異なり語数７２語（２５．５％）となった．この異なり語数の差は，助詞・助動詞の扱いが影響していると思われる.

2. 英語版「児童の権利に関する宣言」における基本語彙の有効性

英語版の「児童の権利に関する宣言」の一部を図２.a.に示す.

PRINCIPLE 2
The child shall enjoy special protection, and shall be given opportunities and facilities, by law and by other means, to enable him to develop physically, mentally, morally, spiritually and socially in a healthy and normal manner and in conditions of freedom and dignity. In the enactment of laws for this purpose the best interests of the child shall be the paramount consideration.

図２.a. 英語版「児童の権利に関する宣言」
（Declaration of the Rights of the Child）の一部

単語の区切りはスペースとした．冠詞のanはaの項目に，名詞複数形は単数形の見出し語にまとめた．また，形容詞１語で比較級，最上級になっているものは原形の見出し語にまとめた．動詞の三人称単数の活用形，分詞形，過去形は現在形原形にまとめた.

この資料における異なり語数と延べ語数の関係は，以下の通りである.

	異なり語数		延べ語数	
・2000語	221語	73.7%	840語	86.6%
・5000語まで	267語	89.0%	913語	94.1%
・全体	300語	100.0%	970語	100.0%

英語版の延べ語数は９７０語となり，日本語版の延べ語数１１８１語に比較して，17.9%少ない．日本語版の分析で，「ついて」「当って」「という」等という連語を１単位としてたてた事を考えると，連語を分けた場合には更にこの延べ語数の差は広がることになる．日本語版の異なりの「基本二千語」までは，延べの７１．９％しかカバーしないのに対し，英語版の異なり「２０００語」は延べの８６．６％をカバーしている．「５０００語まで」では延べの９４．１％をカバーしている．しかし，実際にはこれは５０００語にその派生語を加えた範囲であることを考慮しなければならない．

横軸に異なり語数，縦軸に延べ語数をとり，その関係を累積使用率のグラフにした（図２.b.）．このグラフを日本語版のグラフと比較すると，英語版（図２.b.）では「２０００語」から「５０００語」，「その他」への境の部分で，日本語版（図１.b.）のような立ち上がりのカーブがなく，直線的なのが特徴である．つまり，日本語版では「基本二千語」以外に，「基本六千語」「その他」の一部の語に，繰り返し用いられる語があるのに対して，英語版では「２０００語」「５０００語」「その他」と語彙の範囲が広がるにつれて，その使

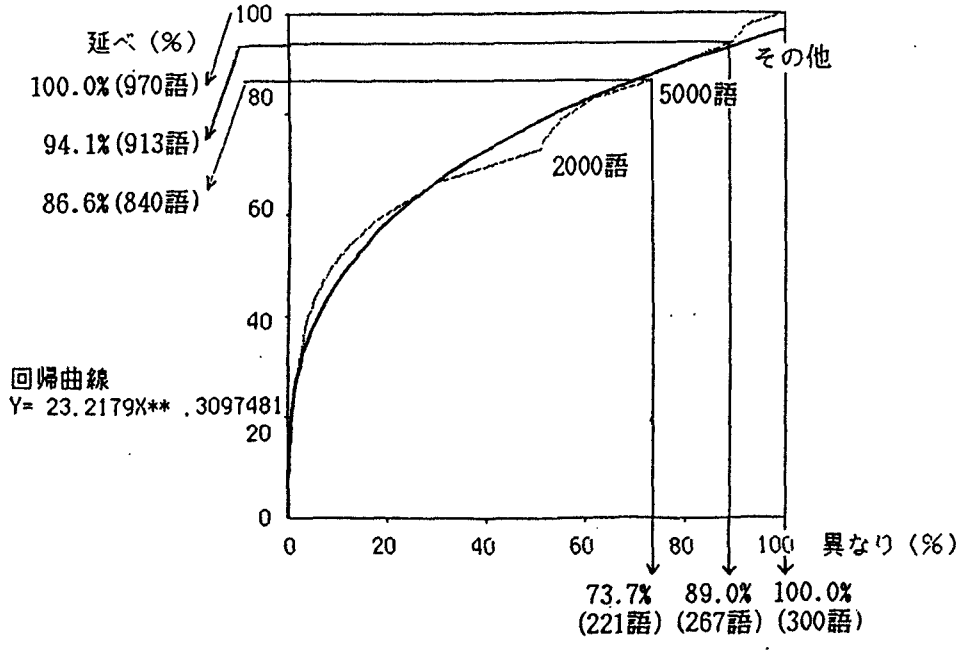


図２.b. 「児童の権利に関する宣言」 英語版　有効性　パーセント表示

用回数も少なくなる傾向があるということである.

このグラフに対し，日本語版と同様，回帰曲線を算出したところ，

Y = 23.22 X ** 0.310

となった．X の係数が日本語版の18.88より大きく，原点からの立ち上がり部分がより急であることを示している.

また，延べ語数を50％カバーする点をさがしてみると，英語版では異なり語数27語（9.0％）と，日本語版（助詞・助動詞を含む）の35語（11.5％）より低い．延べ語数の50％に達するまでの「立ち上がり方」が，日本語版より英語版の方が急であることがこの点からも読み取れる.

3. 仏語版「児童の権利に関する宣言」における基本語彙の有効性

仏語版の「児童の権利に関する宣言」の一部を図3.a.に示す.

Principe 3 L'enfant a droit, dès sa naissance, à un nom et à une nationalité.

図3.a. 仏語版「児童の権利に関する宣言」
（Declaration des droits de l'enfant ）の一部

単語の区切りはスペース，更に母音省略の ' を含む．名詞複数形は単数形の見出し語にまとめた．冠詞の女性形une，la は男性形un，le にまとめた．所有形容詞の女性形も男性形にまとめ，単数形・複数形は分けた．形容詞比較級 meilleur は原形 bonにまとめた．また，動詞の活用,分詞形は原形にまとめた．冠詞との省約形au，aux，du，des のうち，単数形のau，du はa，de に入れ，aux ,des は別項目とした.

この資料における異なり語数と延べ語数の関係は，以下の通りである.

	異なり語数		延べ語数	
・5000語	254語	83.0%	1061語	92.9%
・全体	306語	100.0%	1142語	100.0%

仏語版では，異なり語数は306語，延べ語数は1142語と，数の上では英語版より日本語版に近い形となった．しかし，仏語版は日本語版より少ない「5000語」で，日本語版より多く，異なりの83.0％，延べの92.9％をカバーしている．つまり，日本語版の

異なりは「その他」の語彙に依存することが多いといえる．

横軸に異なり語数，縦軸に延べ語数をとり，その関係を累積使用率のグラフにした（図３.b.）．この仏語版グラフについても英語版（図２.b.）と同様，「５０００語」から「その他」への境の部分で，日本語版（図１.b.）のような立ち上がりのカーブがなく，直線的なカーブが見られた．従って，日本語版では基本語彙のウェイトが高く，基本度の高と低にへだたりがあるが，英語版，仏語版ではそうした段階的へだたりは見られない．これは，英語版，仏語版の基本語彙の設定方法が日本語のそれと比較して，使用頻度に重点が置かれているためかもしれない．回帰曲線は

Y = 27.57 X ** 0.286

となり，X の係数は日本語版，英語版よりさらに大きくなって，原点からの立ち上がり部分が急になっていることを示している．

また，延べ語数を５０％カバーする点をさがしてみると，仏語版では異なり語数１９語（６.２％）と，日本語版，英語版よりずっと低い．この点からも延べ語数の５０％に達するまでの「立ち上がり方」が，日本語版，英語版より仏語版の方が急であるといえる．

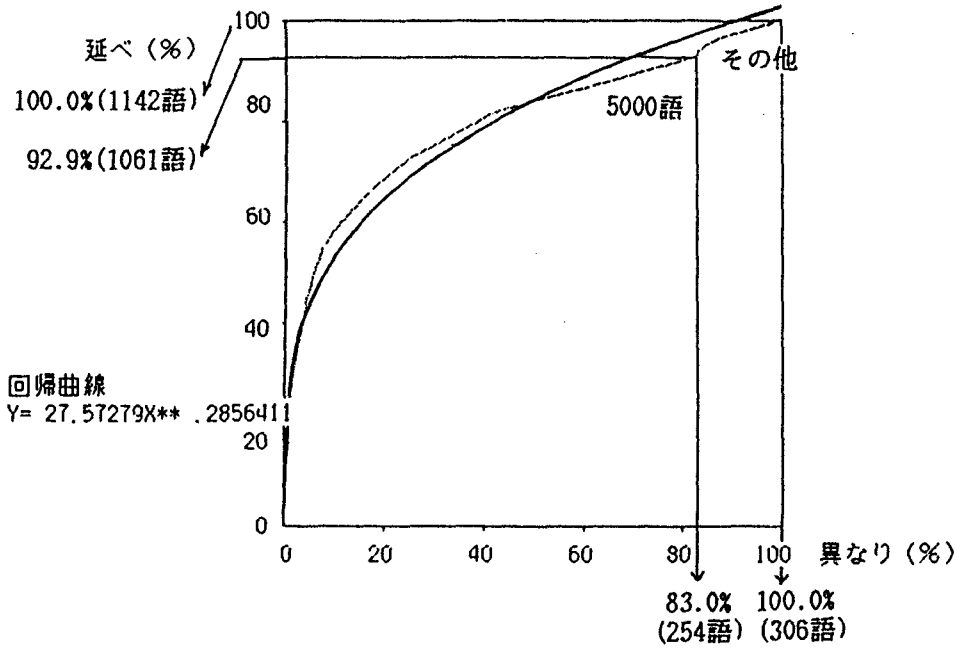


図３.b. 「児童の権利に関する宣言」 仏語版　有効性　パーセント表示

4. 西語版「児童の権利に関する宣言」における基本語彙の有効性

西語版の「児童の権利に関する宣言」の一部を図４.a.に示す.

PRINCIPIO 4
El niño debe gozar de los beneficios de la seguridad social. Tendrá derecho a crecer y desarrollarse en buena salud; con este fin deberán proporcionarse, tanto a él como a su madre, cuidados especiales, incluso atención prenatal y postnatal. El niño tendrá derecho a disfrutar de alimentación, vivienda, recreo y servicios médicos adecuados.

図４.a. 西語版「児童の権利に関する宣言」
（Declaracion de los derechos del niño）の一部

単語の区切りはスペースとした. 名詞複数形は単数形の見出し語にまとめた. 冠詞単数の女性形 una, laは男性形 un, el にまとめた. 中性形 lo も男性形 el にまとめた. 複数女性形の unas, las は男性形 unos, los にまとめた. 形容詞の女性形, 複数形も男性単数形にまとめた. algún, ningún のような形容詞の語尾脱落形は, 原形alguno, ninguno にまとめた. 指示代名詞, 指示形容詞の女性形, 中性形, 複数形はすべて男性単数形にま

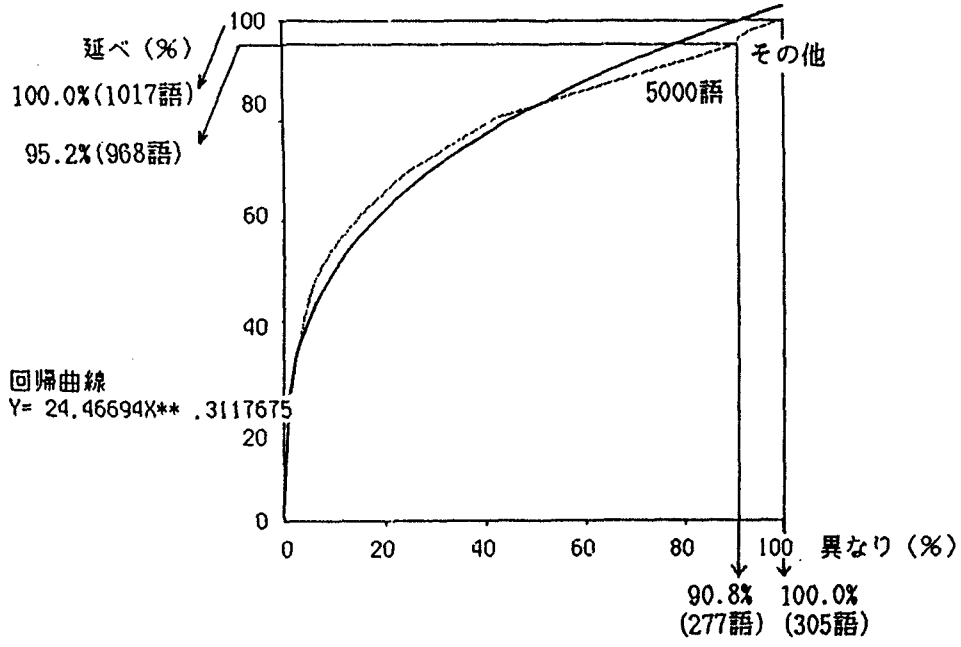


図４.b. 「児童の権利に関する宣言」 西語版 有効性 パーセント表示

とめた．形容詞・副詞で１語で比較級，最上級になっているものは，原形の見出し語にまとめた．また，動詞の活用，分詞形は原形にまとめた．接続詞のうち調音上の変化によるｕ，ｅ はｏ，ｙにまとめた．動詞の不定詞や現在分詞のあとに，separarse のように人称代名詞が付いているものは，separar と se のように分けた．

この資料における異なり語数と延べ語数の関係は，以下の通りである．

	異なり語数		延べ語数	
・5000語	277語	90.8%	968語	95.2%
・全体	305語	100.0%	1017語	100.0%

西語版では，５０００語までで異なりの９０．８％，延べの９５．２％とかなりの範囲をカバーしているのが特徴である．同じ５０００語までの仏語版が，異なりの８３．０％，延べの９２．９％をカバーしているのと比べるとそれがよくわかる．特に，異なりをカバーする率が高いということは，文章全体がやさしく書かれている印象を与えるのではないだろうか．

横軸に異なり語数，縦軸に延べ語数をとり，その関係を累積使用率のグラフにした（図４．b.）．このグラフの場合も，仏語版グラフ（図３．b.）や英語版（図２．b.）と同様，「５０００語」から「その他」への境の部分で，日本語版（図１．b.）のような立ち上がりのカーブがなく，比較的直線的であった．回帰曲線は

Y = 24.47 X ** 0.312

となり，X の係数は英語版と仏語版の間に位置している．

また，延べ語数を５０％カバーする点をさがしてみると，西語版では異なり語数２２語（７．２％）と，やはり英語版と仏語版の間になっている．

5. おわりに

少なくとも，この「児童の権利に関する宣言」を資料として読む場合，日本語の「基本二千語」「基本六千語」より，英語，仏語，西語の「５０００語」の方が有効性が高いと考えられる．これが今回の調査の結果であるが，今後考えなければならない問題もあり，それを簡単に述べて結びとしたい．

第１に，今回の調査で算出した，日本語，英語，仏語，西語の回帰曲線を一つのグラフに重ねてみた（図５．a.）．原点からの曲線の立ち上がり方は，仏語，西語，英語，日本語の順となり，途中から日本語が英語を抜く形になっている．これは，英語の語彙の重要度が，頻度に重きをおいてつけられているのに対し，日本語には「基本六千語」や「その

他」の中にも比較的高頻度の語が混ざっていることによっていると考えられる．仏語版は，回帰曲線全体が日本語版，英語版を常に上回っており，西語版がそれに準じたカーブを見せている．

第２に，中野洋『「星の王子さま」６か国語版の語彙論的研究』［文献13］では，頻度数順の累積使用率の分布曲線（同語異語判別後）において，上から日本語（含助・助動），仏語，英語，独語，日本語（除助・助動）という順になり，次のように述べられている．

> 上位10位ぐらいまでは，フランス語が最も上に分布している．しかし，それ以降は助詞・助動詞を含んだ日本語が上になる．他の言語の格変化は数にあらわれないのに，それにあたる助詞が日本語では生きているためである．日本語から助詞・助動詞を除くと曲線は最も下になる．各国語の前置詞や助動詞の数が計算に入っているのに，それにあたる助詞・助動詞が日本語から除かれるためである．

しかし，今回の「児童の権利に関する宣言」の調査においては，仏語の基本語彙「５０００語」が，助詞・助動詞を含む日本語の「基本六千語」よりも，異なり，延べとも，より多くカバーする結果となった．これについては，今回の調査で，日本語の「名詞＋する」を１語と認定したことも，理由の一つに考えられる．

第３に，今回の調査の結果では，実際の文章を読解する場合，日本語の基本語彙は外国語のものと比較して，より多くの異なり語数を必要とする．これについては，同じ意味分野を表す語に和語・漢語・外来語が併存するという，日本語の語彙構造の特殊性が，一因

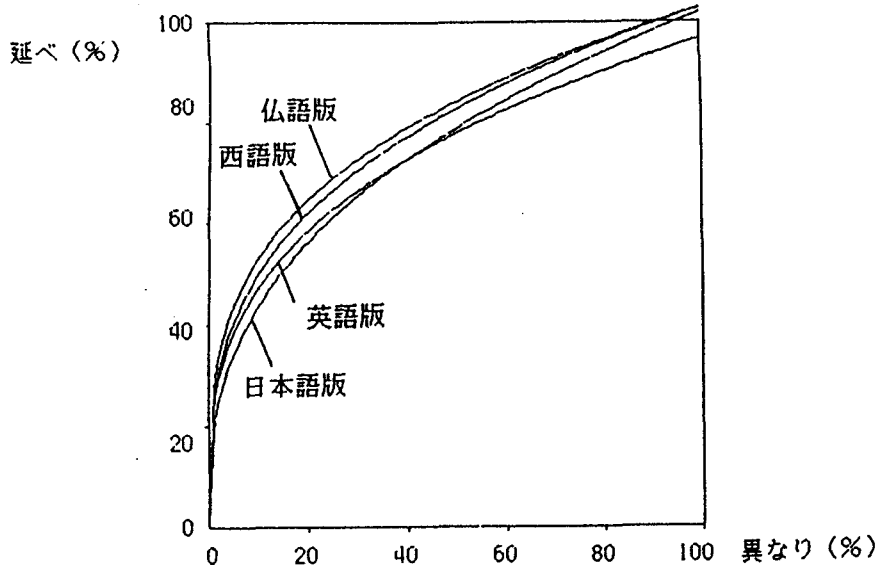


図５.a. 「児童の権利に関する宣言」 各国語版　有効性　回帰曲線

となっていると考えられる．例えば，和語「生まれ」は，基本語彙のひとつであるが，今回の調査の対象のような文章のためには，類義語の漢語「出生」「門地」も，やはり基本語彙のひとつに含めなければならない．諸外国語では，それが，多分，「birth（英語），naissance （仏語），nacimiento （西語）」それぞれ一つですまされるのである．

文献

1．水谷静夫（1958） 基本語彙と語彙調査
『国語教育のための国語講座４』
2．国立国語研究所（1962） 現代雑誌九十種の用語用字 第一分冊
『国立国語研究所報告21』
3．国立国語研究所（1964） 現代雑誌九十種の用語用字 第三分冊
『国立国語研究所報告25』
4．国立国語研究所（1964） 分類語彙表
5．ジョルジュ・マトレ（野村二郎・滑川明彦訳編）（1967）
フランス基本語5000辞典
6．服部四郎（1969） 英語基礎語彙の研究
7．国立国語研究所（1970） 電子計算機による新聞の語彙調査
『国立国語研究所報告37』
8．林四郎（1971） 語彙調査と基本語彙
『国立国語研究所報告39 電子計算機による国語研究Ⅲ』 1－35
9．新英和中辞典 第３版 （1971）
10．岩崎英二郎，早川東三，子安美知子，平尾浩三，鉄野善資（1971）
ドイツ基本語5000辞典
11．高橋正武，瓜谷良平，宮城昇，エンリケ・コントレラス（1972）
スペイン基本語5000辞典
12．ジョン・ロバート・ショー（福田陸太郎訳編）（1972）
ラダー英和基本語5000辞典
13．中野洋（1976） 「星の王子さま」６か国語版の語彙論的研究
『計量国語学 第七十九号』 18－31
14．見坊豪紀（1976） 辞書をつくる 現代の日本語
15．中央教育研究所（1976） 学習基本語彙の基礎調査
16．土居光知（1977） 基礎日本語
『土居光知著作集４』
17．田中章夫（1977） 漢字調査における統計的尺度の問題
『電子計算機による国語研究Ⅷ』
18．中野洋（1980） 『分類語彙表』の語数
『計量国語学 第十二巻第八号』 376－381
19．国立国語研究所（1982） 日本語教育基本語彙七種 比較対照表

『日本語教育指導参考書９』
20. 寺田富男（1982） 基本語・基礎語
『日本語教育事典』 295－299
21. 水谷静夫（1983） 語彙
『朝倉日本語新講座２』
22. 阪本一郎（1984） 新教育基本語彙
23. 阪本一郎（1984） 私の基本語彙論
『日本語学 第三巻第二号』 11－15
24. 林四郎（1984） 私の基本語彙論
『日本語学 第三巻第二号』 16－22
25. 森岡健二（1984） 私の「基本語彙論」観
『日本語学 第三巻第二号』 23－27
26. 田中章夫（1984） 基本語彙と基本語
『日本語学 第三巻第二号』 28－38
27. 国立国語研究所（1984） 日本語教育のための基本語彙調査
『国立国語研究所報告78』
28. 国立国語研究所日本語教育センター（1984） フランス語基本語彙七種比較対照表
29. 岩井勇児，鈴木真雄（1985） 教師のための統計法入門［第２版］
30. 安倍齊（1985） 応用数理統計学入門
31. 国立国語研究所（1986） 日独仏西基本語彙対照表
32. 細川英雄（1987） 日本語教育基本語彙と国語教科書低学年語彙
――動詞を中心とした二種比較対照作業の試み――
『金沢大学大学教育開放センター紀要 第７号』
33. 田中章夫（1988） 国語語彙論
34. Michael West (1961) The New Method English Dictionary Fourth Edition
35. E. L. Thorndike, Clarence L. Barnhart (1968)
Thorndike･Barnhart High School Dictionary Fifth Edition
36. Michael West (1976) The New Method English Dictionary Fifth Edition
37. John Robert Shaw, Sara Janet Shaw (1990)
The New Horizon Ladder Dictionary of the English Language
Revised and enlarged edition
38. G. Gougenheim, R. Michéa, P. Rivenc, A. Sauvageot (1956)
L'Elaboration du Français Elémentaire
39. Georges Gougenheim (1958)
Dictionnaire Fondamental de la Langue Française
40. Georges Matoré (1963) Dictionnaire du vocabulaire essentiel
(les 5000 mots fondamentaux)

（1992年7月31日受付）

***** descriptors and abstracts *****

COMPARISON OF EFFECTIVENESS OF VARIOUS BASIC VOCABULARIES

SANADA Haruko (Gakushuin University)

Descriptors: basic vocabulary; Japanese; English; French; Spanish; effectiveness

Abstract:

This paper deals with an analysis of effectiveness of basic vocabulary.

The effectiveness of Japanese, English, French, and Spanish basic vocabularies is compared using "Declaration of the right of the child" adopted by the Unite Nations in 1959 in Japanese, English, French, and Spanish versions.

As samples of basic vocabularies, the following materials are used:

Japanese: 'Basic 6,000 words' including 'Basic 2,000 words' listed in *A Study of the Fundamental Vocabulary for Japanese Language Teaching*, The National Language Research Institute, Research Report 78 (Tokyo: The National Language Research Institute, 1984)

English: about 5,000 words listed in *The New Horizon Ladder Dictionary of the English Language*, J. R. Shaw, translated by R. Fukuda (Tokyo: Hakusuisha, 1972).

French: about 5,000 words listed in *Dictionnaire du Vocabulaire essentiel*, G. Matoré, translated by J. Nomura and A. Namekawa (Tokyo: Hakusuisha, 1967).

Spanish: about 5,000 words listed in *Diccionario de Vocabulario Fundamental del Español*, M. Takahashi and others (Tokyo: Hakusuisha, 1972).

For each version a graph is drawn of the number of running words versus the number of different words. A graph of regression curves for Japanese, English, French, and Spanish is also drawn. The values of the ordinates (the degree of accumulative frequency) near the origin are highest for French, followed by Spanish, English, and Japanese. The Japanese curve outruns the English one. This is because the frequency is weighted for the selection of the English basic vocabulary, while in Japanese 'Basic 6,000 words' and 'Others' include high-frequency words. The curves for French and Spanish lie above those for English and Japanese.

In this study it is observed that for an understanding of the sample text the Japanese basic vocabulary is not sufficient compared with foreign ones. This is probably due to the fact that Japanese has a unique structure of the vocabulary in which original Japanese words, loan words from China, and loan words from Europe constitute the same semantic field.

（３）文体の自動変換　デアル体への変換

（『計量国語学』第17巻第３号　1989年　収録）

計量国語学第十七巻第三号〔Mathematical Linguistics, vol.17 no.3〕1989年

調査報告

文体の自動変換――デアル体への変換

真田　治子（日本アイ・ビー・エム㈱　学習院大学大学院聴講生）

ディスクリプタ：　文体変換　文体　常体　敬体
デス・マス体　デアル体　活用分析

1. デアル体変換の目的

筆者は各種「文体の自動変換システム」の作成を目指し，既に「常体→敬体システム」を試作した．

今回は，様々な文体で分担執筆されている文章の編集等では「デアル体への自動変換」が有用であろうと考え，システムの作成を試みた．

2. 文体変換プログラムの概要

2.1. プログラムの機能　今回作成した文体変換プログラムの主な機能は次の通りである．

・ダ体，デス・マス体の述部をデアル体に変換する

文中，文末を問わず該当の述部はすべて変換する．

・引用文は変換対象外とする

・大きな辞書ファイルを用いず，一部に確率的判断を含む

パソコンのレベルのプログラムを目指したので，動詞・形容詞の活用の分析の所に一部確率的判断が含まれている．

2.2. 稼動環境　このプログラムはＢＡＳＩＣで作成した．また，入出力はＭＳ－ＤＯＳのテキストファイルの書式を使用した．前処理として国立国語研究所の一貫処理プログラムを使用し，単位切り・読み仮名付け・品詞認定・活用形認定を行なった．

前処理と変換プログラムとの関係を図2.2.に示す．

ベタ打ちした漢字仮名混じり文のファイル（2.2.a.）を一貫処理プログラム（2.2.b.）にかけてデータファイル（2.2.c.）を得る．この時前処理の誤りは人手で修正しておく．変換プログラム（2.2.d.）はこのファイルを入力とし，文体変換をしてファイル（2.2.e.）に出力する．

2.3. 入出力文の例　変換プログラムの入出力の具体例を図2.3.a.（入力）と図2.3.b.（出力）に示す．これらは各図2.2.（前処理と変換プログラムとの関連図）のファイル2.2.c.と2.2.e.に相当する．文章が１文字１行の形で出力され，各文字に字種，読み，単位切り，品詞，活用形の情報が付加される．

SANADA Haruko（IBM Japan, Ltd. ／Gakushuin University）—Automatic Sentence Style Conversion in Japanese：Conversion to Dearu-Style

図2.2.　前処理との関連図

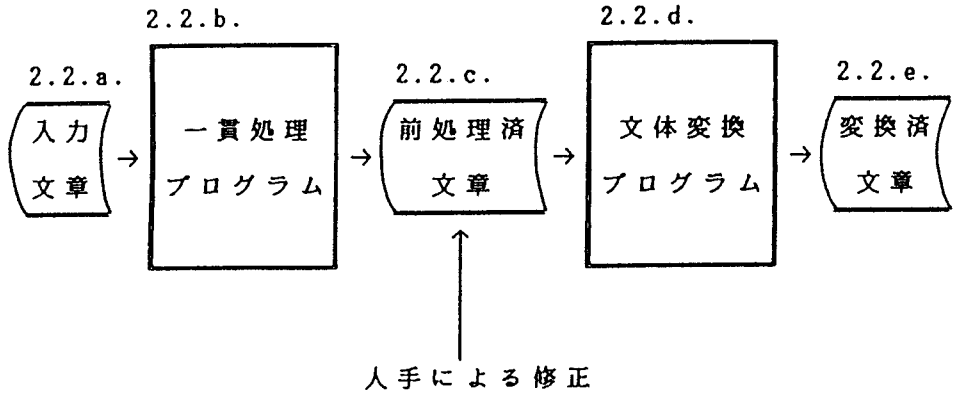


図2.3.a.　入力（=2.2.c.）

文字	字種	読み	単位	品詞	活用
天	L	てん			
気	L	き	1	1	
の	H		1	R	
く	H				
ず	H				
れ	H		1	1	
を	H		1	R	
起	L	お			
こ	H				
す	H		1	E	+
低	L	てい			
気	L	き			
圧	L	あつ	1	1	
が	H		1	R	
近	L	ちか			
づ	H				
い	H		1	E	9
て	H		1	Q	9
い	H				
る	H		1	E	+
こ	H				
と	H		1	1	
が	H		1	R	
多	L	おお			
い	H		1	M	+
の	H		1	R	
で	H				
す	H		1	P	H
。	E		1	Y	

図2.3.b.　出力（=2.2.e.）

文字	字種	読み	単位	品詞	活用
天	I				
気	I				
の	H		1	R	
く	H				
ず	H				
れ	H		1	1	
を	H		1	R	
起	L	お			
こ	H				
す	H		1	E	+
低	L	てい			
気	L	き			
圧	L	あつ	1	1	
が	H		1	R	
近	L	ちか			
づ	H				
い	H		1	E	9
て	H		1	Q	9
い	H				
る	H		1	E	+
こ	H				
と	H		1	1	
が	H		1	R	
多	L	おお			
い	H		1	M	+
の	H		1	R	
で	H				
ある	H		1	Q	9
。	E		1	E	H
			1	Y	

＜凡例＞

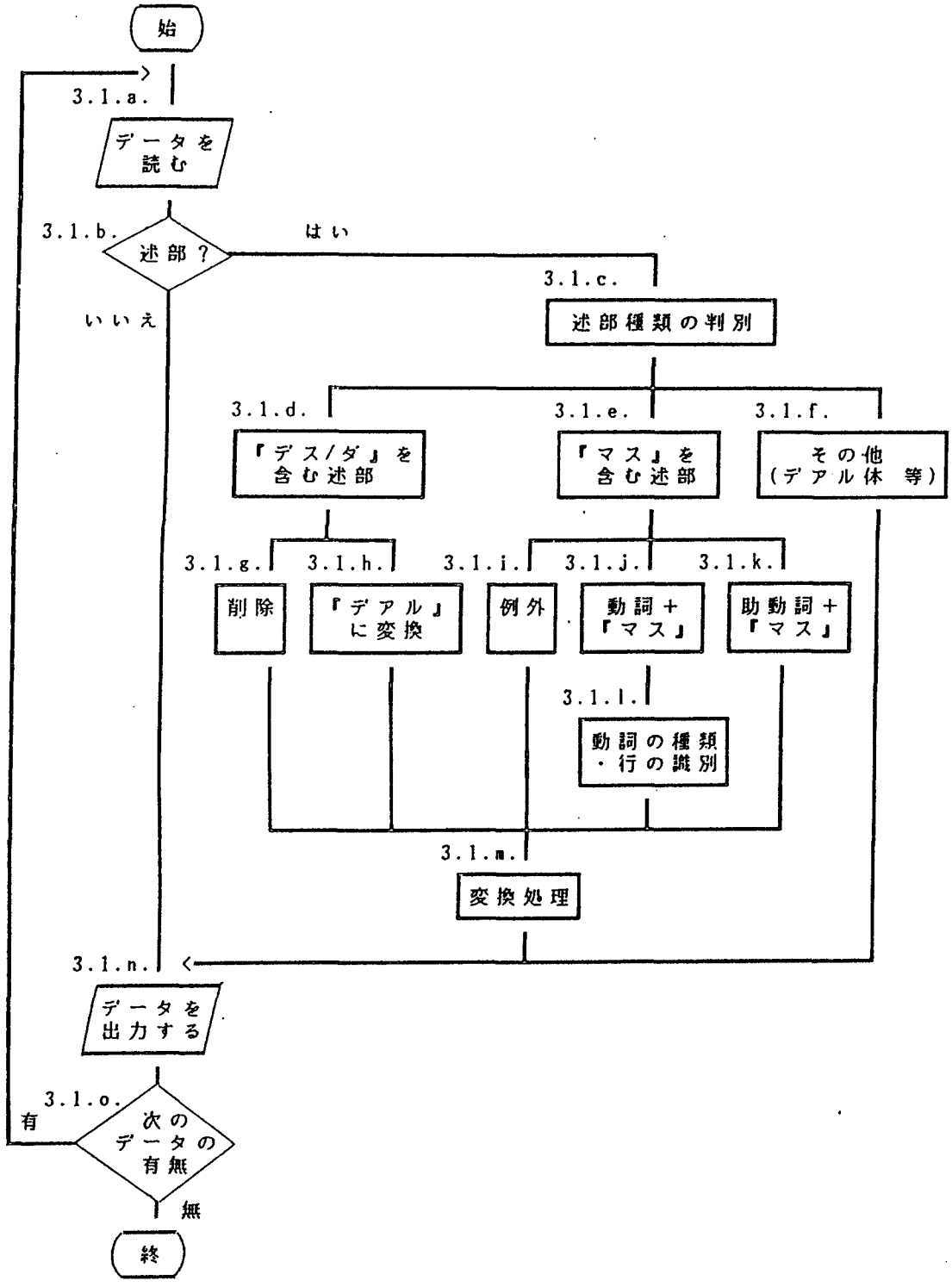
字種　H：かな
　　L：漢字

単位　0：続く
　　1：切れる

品詞　1：名詞
　　R：助詞
　　M：形容詞
　　E：動詞
　　Q：助助動詞
　　P：助動詞

活用　+：終止連体
　　9：連用
　　H：終止

図3.1.　変換処理全体の流れ



3. 変換処理の概要

3.1. 変換処理全体の流れ　変換処理全体の流れを図3.1.に示す.

まずデータを読む（3.1.a.）. これが述部であれば述部種類の判別処理（3.1.c.）へ行く. 述部でなければそのまま出力処理（3.1.n.）に進む.

述部の種類は

- 『デス，ダ』を含む述部（3.1.d.）
- 『マス』を含む述部（3.1.e.）
- その他（例えばデアル体の述部）（3.1.f.）

の3つに判別される. 「その他」の述部はそのまま出力処理（3.1.n.）に進む.

「『デス，ダ』を含む述部（3.1.d.）」は更に

- 『デス，ダ』を削除する述部（「赤いです」→「赤い」等）（3.1.g.）
- 『デアル』に変換する述部（「本です」→「本である」等）（3.1.h.）

の2つに判別され，変換処理（3.1.m.）が行なわれる.

「『マス』を含む述部」（3.1.e.）」は更に

- 『マス』の直前の語が動詞である述部（「歩きます」→「歩く」等）（3.1.j.）
- 『マス』の直前の語が助動詞である述部（「書かせます」→「書かせる」等）（3.1.k.）
- 特殊な変換を要する述部（「ありません」→「なかった」等）（3.1.i.）

の3つに判別される.

このうち，「『マス』の直前の語が助動詞である述部（3.1.k.）」と「特殊な変換を要する述部（3.1.i.）」はすぐ変換処理（3.1.m.）へ行くが，「『マス』の直前の語が動詞である述部（3.1.j.）」は動詞の種類と行の識別処理（3.1.l.）を経由して変換処理（3.1.m.）へ進む. ここでは『マス』に接続している連用形の動詞を，『マス』の次の語によって音便連用形，終止形，未然形等に活用させるための前処理として，種類と行を識別する. これについては次で述べる.

変換処理（3.1.m.）で変換された述部は出力される（3.1.n.）. 次のデータがあれば（3.1.o.）また最初（3.1.a.）に戻る.

3.2. 動詞（連用形）の識別処理の流れ　この処理は図3.1.（変換処理全体の流れ）のうち，動詞の種類と行の識別処理（3.1.l.）に相当する. 流れを図3.2.に示す.

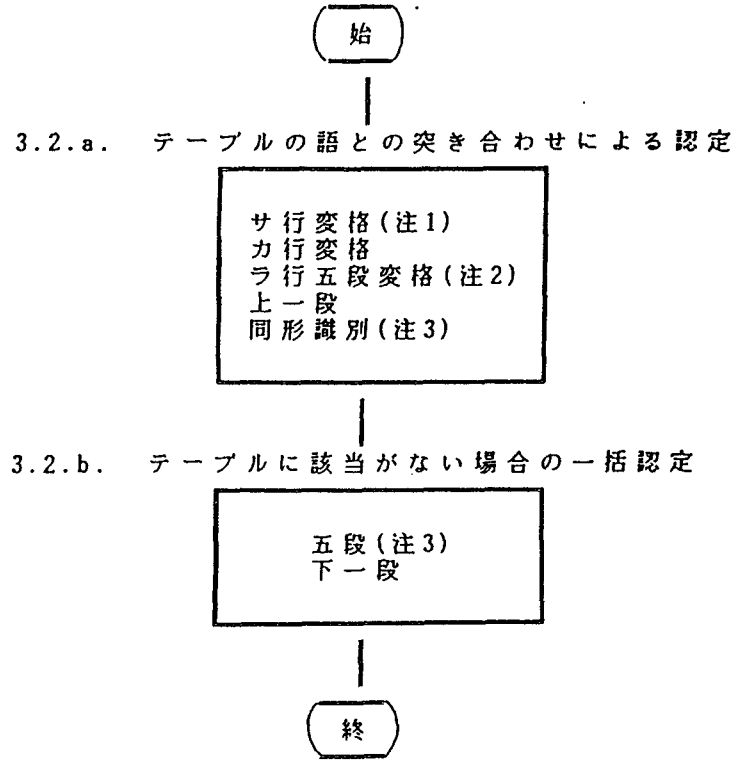
ここで処理される動詞は助動詞『マス』に接続しているため，すべて連用形である.

まず，上一段・変格活用動詞のテーブルと突き合わせ（3.2.a.），該当がない場合は語末音により，五段または下一段と一括認定を行なう（3.2.b.）.

また表記上は同形でどちらの種類か識別できない場合は，同形識別用テーブルと突き合わせ，強制的にいずれか1つに認定するようにした.

例えば「起きます」や「来ます」という表記の場合は上一段・変格活用動詞のテーブルとの突き合わせで，各カ行上一段，カ行変格活用と認定される.「歩きます」「受けます」の場合はテーブルに該当がなく，語末音が各イ段音，エ段音であるので，カ行五段，カ行下一段と一括認定される.「きます」という表記の場合には「行く，来る」の「来ます」（カ行変格活用）か「服を着る」の「着ます」（カ行上一段）か識別できないので確率判断により，強制的にカ行変格と認定する.

図3.2.　動詞（連用形）の識別処理（=3.1.1.）の流れ



（注1）前処理の段階で、
　　複合サ変動詞は名詞＋サ変としてある
（注2）"なさる"、"くださる"の類
（注3）確率的判断を含む

この処理では，同形識別，五段動詞の認定に確率判断（一番頻度の高い所に落ち着く方法）が含まれている．

ラ行五段変格とは，「なさいます」「くださいます」等連用形活用語尾が「ーい」となる類の動詞である．

また前処理の段階で，複合サ変動詞は名詞＋サ変と処理してある．

3.3.　変換処理の流れ　　この処理は図3.1.（変換処理全体の流れ）のうち，変換処理（3.1.m.）に相当する．

以上のようにして行，種類を認定した動詞や，変換の必要のある形容詞，助動詞は次のようにして変換処理を行なう．

品詞別目的別に変換テーブルを持っており，適切なテーブルを読みこんで削除情報と追加情報を得る．この情報を用いて入力データを加工する．

例えば「書きます」という述部はカ行五段動詞連用形に『マス』が付いた語であると認定されてこの処理に入る．削除情報は語末の「きます」，追加情報は「く」であることがテーブルからわかるので「書く」に変換される．

図4.1.　テスト結果の例

4.1.a.　入力文（1）

　金魚ばちの中で、金魚が、無心に泳いでいます。そこへ、金魚のえさとして、赤えびの一きれを投げてやると、水底にいた金魚まで、それを見つけて集まってきます。金魚は、赤えびの赤い色に目をひかれたのでしょうか。それとも、投げこんだときの、かすかな音を聞きつけたのでしょうか。それとも、えびのにおいをかぎ分けたのでしょうか。
　いったい、魚には、色がわかるのでしょうか。魚の中には、ある種の深海魚のように目のないのもありますが、ふつうは、頭の左右に、一対の目をもっています。これをよく調べてみると、目に必要な部分はそろっているし、色に感じる円すい体という部分もあるので、物を見る感覚はあると思われます。しかし、構造だけでは、その働きは、よくわかりません。魚が実際に物を見分け、色に感じることができるかどうかを調べるために、科学者たちは、次のような実験をしました。
（中略）
　そのころ、ドイツに、ラドクリッフという生物学者がいました。その人は、友だちといっしょに、「魚は、音が聞こえるだろうか。」という研究をしていました。ますのたくさんいる川で、てっぽうをうって、その音で、ますがおどろくかどうかという実験をしたのですが、その結果、「ますは、音が聞こえない。」という説を発表しました。
（以下略）

4.1.c.　4.1.b.

4.1.d.　入力文（2）

　ラジオやテレビで報道される天気予報は、たくさんの観測資料にもとづいて、せんもん家が天気の変化を予測したものです。わたしたちは、この天気予報によって、明日の予定を立てたり、持ち物を用意したりしています。
　むかしの人々は、身近にいる鳥や虫などの生き物の様子や、空や風の様子などを手がかりにして、天気の変化をあらかじめ知ろうとしました。ですから、生き物や空や風の様子と天気との関係についてのことわざが、たくさん生まれました。そういうことわざは、今日でも、各地に伝えられています。
（以下略）

4.1.e.

4.1.f.　入力文（3）

「一」
　或日の事でございます。御釈迦様は極楽の蓮池のふちを、独りでぶらぶら御歩きになっていらっしゃいました。池の中に咲いている蓮の花は、みんな玉のようにまっ白で、そのまん中にある金色の蕊からは、何とも云えない好い匂が、絶間なくあたりへ溢れております。極楽は丁度朝なのでございましょう。
（以下略）

4.1.g.　入力文（4）

　主人は、羽織・はかまを着けて、茶わんをりっぱな箱の中に収めて、それをかかえて参上いたしました。

4.1.h.

4.1.i.　入力文（5）

　お役人は、殿さまの前に、茶わんをささげて、持ってまいりました。

4.1.j.

4.1.k.　入力文（6）

　季節は、もう秋の末で寒うございましたから、熱いお汁が身体をあたためて、たいへんうもうございましたが、茶わんは厚いから、けっして手が焼けるようなことがありませんでした。

4.1.m.　4.1.l.

4.1.n.　出力文（1）

金魚ばちの中で、金魚が、無心に泳いでいる。そこへ、金魚のえさとして、赤えびの一きれを投げてやると、水底にいた金魚まで、それを見つけて集まってくる。金魚は、赤えびの赤い色に目をひかれたのであろうか。それとも、投げこんだときの、かすかな音を聞きつけたのであろうか。それとも、えびのにおいをかぎ分けたのであろうか。いったい、魚には、色がわかるのであろうか。魚の中には、ある種の深海魚のように目のないのもあるが、ふつうは、頭の左右に、一対の目をもっている。これをよく調べてみると、目に必要な部分はそろっているし、色に感じる円すい体という部分もあるので、物を見る感覚はあると思われる。しかし、構造だけでは、その働きは、よくわからない。魚が実際に物を見分け、色に感じることができるかどうかを調べるために、科学者たちは、次のような実験をした。
（中略）
そのころ、ドイツに、ラドクリッフという生物学者がいた。その人は、友だちといっしょに、「魚は、音が聞こえるだろうか。」という研究をしていた。ますのたくさんいる川で、てっぽうをうって、その音で、ますがおどろくかどうかという実験をしたのであるが、その結果、「ますは、音が聞こえない。」という説を発表した。
（以下略）
—4.1.o.　—4.1.p.

4.1.q.　出力文（2）

ラジオやテレビで報道される天気予報は、たくさんの観測資料にもとづいて、せんもん家が天気の変化を予測したものである。わたしたちは、この天気予報によって、明日の予定を立てたり、持ち物を用意したりしている。　むかしの人々は、身近にいる鳥や虫などの生き物の様子や、空や風の様子などを手がかりにして、天気の変化をあらかじめ知ろうとした。ですから、生き物や空や風の様子と天気との関係についてのことわざが、たくさん生まれた。そういうことわざは、今日でも、各地に伝えられている。
（以下略）
—4.1.r.

4.1.s.　出力文（3）

「一」　或日の事である。御釈迦様は極楽の蓮池のふちを、独りでぶらぶら御歩きになっていらっしゃった。池の中に咲いている蓮の花は、みんな玉のようにまっ白で、そのまん中にある金色の蕊からは、何とも云えない好い匂が、絶間なくあたりへ溢れている。極楽は丁度朝なのであろう。
（以下略）

4.1.t.　出力文（4）

主人は、羽織・はかまを着けて、茶わんをりっぱな箱の中に収めて、それをかかえて参上した。
—4.1.u.

4.1.v.　出力文（5）

お役人は、殿さまの前に、茶わんをささげて、持ってきた。
—4.1.v.

4.1.x.　出力文（6）

季節は、もう秋の末で寒かったから、熱いお汁が身体をあたためて、たいへんうもかったが、茶わんは厚いから、けっして手が焼けるようなことがなかった。
—4.1.z.　—4.1.y.

4. テスト結果

4.1. テスト結果の例　デス・マス体の小学生向け説明文やゴザイマス体の童話等５編を使用して変換テストを行なった．テスト結果の例を図4.1.に示す．

入力文 (1)（4.1.a.）はデス・マス体の文章で，変換結果は出力文 (1)（4.1.n.）である．文中のデス・マス体（4.1.b.と4.1.c.）はデアル体に変換されている（4.1.p.と4.1.o.）．

入力文 (2)（4.1.d.）もデス・マス体の文章で，変換結果は出力文 (2)（4.1.q.）である．接続詞「ですから」（4.1.e.）は今回は変換しなかったため，デアル体の中に残っている（4.1.r.）．

入力文 (3)（4.1.f.）はゴザイマス体の文章で，変換結果は出力文 (3)（4.1.s.）である．

入力文 (4)（4.1.g.），入力文 (5)（4.1.i.），入力文 (6)（4.1.k.）の文章は各，出力文 (4)（4.1.t.），出力文 (5)（4.1.v.），出力文 (6)（4.1.x.）に変換された．

「いたしました」（4.1.h.）は「した」（4.1.u.），「まいりました」（4.1.j.）は「きた」（4.1.w.）と変換し，デス・マスを削除するだけでなく，動詞そのものを文体に合わせて差し替えた．「参る」という動詞は「行く」と「来る」の両方の意味があるが，ここでは「来る」に一括変換してある．

「寒うございました」（4.1.l.）と「うもうございました」（4.1.m.）は形容詞に「ございます」が付いた形である．「寒い」は正しく変換されたが（4.1.y.），「うまい」では誤変換が起こっている（4.1.z.）．ク活用の形容詞＋「ございます」を変換する場合，形容詞の語末の「オ段音＋う」は「ア段音またはオ段音＋活用語尾」にしなければならないが，今回は「オ段音＋活用語尾」に一括変換したためこのような誤変換が発生した．これについては後で詳しく述べる．

4.2. テスト集計結果　文末数は387，そのうち引用文であるため変換対象外としたのが73，残り314が変換対象で，うち通過数は314，失敗数は０であった．

また文中のデス・マスで変換したものは36，うち失敗数は１であった．これは4.1.z.のケースである．

5. 今回のプログラムで特に考慮した点

5.1. 文中の述部の変換　今回は文末だけでなく，文中の述部も変換対象とした．

・文頭からの述部探索

当初は，変換すべき文中の述部を抽出する際，文を最初から１語ずつ読んで，述部が成立するかどうか検査する方法をとった．しかし，この方法では「〜ませんでした」という語に誤変換が生じた．例えば「書きませんでした」という語は「書かなかった」と変換すべきであるが，「書きません」まで読むと述部が成立してしまうため，「書きません」を「書かない」，「でした」を「であった」と各変換してしまった．

・基準となる語からの遡行探索

このような誤変換を避けるため，自立語・助詞・読点等を基準として前へ遡って述部を探索する方法を用いた．例えば「書きませんでしたが，」という表現では，まず接続助詞の「が」を見つけ，そこから前へもどって「書きませんでした」という述部を変換した．

5.2. 丁寧語の変換　「ございます」「いたします」等の丁寧語は『デス・マス』を変換するのと共に，動詞そのものを変えて普通の表現にした．「～でございます」は「～である」，「いたします」は「する」，「おります」は「いる」とした．「参ります」は状況によって「行く」または「来る」の意味になるが，今回のシステムでは一括変換で「来る」としてある．注1

6. 今回のプログラムで予想される誤変換，非変換のケース

6.1. 誤変換

・同音異活用語の確率判断

「来ます」という表記はカ行変格，「着ます」という表記はカ行上一段と判断するが，「きます」と仮名で表記してあるものはカ行変格と確率的に判断する．

この場合，どのような助詞が述部の直前に現れるかを調べることによって，ある程度判断の精度を上げることが可能であろうと考えている．例えば「服を着る」の「着る」は他動詞，「行く，来る」の「来る」は自動詞なので，「～をきます」という表記であればカ行上一段の「着る」であると判断できる．注2

しかしこの方法もどの動詞にも適用できるわけではない．例えば「おります」という表記の述部は「元気で居ります」（ラ行五段・丁寧語），「枝を手で折ります」（ラ行五段），「東京駅で降ります」（ラ行上一段）等が考えられ，それぞれ「いる」「おる」「おりる」と変換しなければならない．この３つの動詞の直前に出現しうる助詞としては，「居ります」には「で・て・に」，「折ります」には「で・て・を・に」，「降ります」には「で・て・を・に・から」等が考えられる．従って，助詞「から」の場合だけはラ行上一段の「降りる」と識別することが出来る注3が，他の助詞の場合は判断が困難である．

・テーブルにない上一段動詞の識別

動詞のテーブルは国立国語研究所『電子計算機による新聞の語彙調査Ⅱ』(1971)の「動詞の表」から作成した．従って表にない動詞の場合は一括認定を行なうので誤変換の可能性がある．

・ク活用音便連用形の形容詞の識別

また「美しゅうございます」のように形容詞に『ゴザイマス』がついた述部では，音便連用形を終止形等に活用させる必要がある．この時「くろうございます」のようにク活用の形容詞で語幹の最後がオ段音のものは，「くろ（黒）い」のようにそのままでいい場合と，「くら（暗）い」のようにア段音にする必要がある場合とがある．今回はすべてオ段音のままで一括認定を行なったので4.1.z.「うもかった」のような誤変換の可能性がある．

注１　「存じます」の場合には状況によって「わかる」または「思う」の意味になるが，今回のシステムでは一括変換で「思う」とした．

注２　[文献16]によれば格の結合価を「中核」の範囲に限定すれば，このような例は識別できるが，「周辺」に範囲を広げると「この道をきます（来る）」のように識別できないものが出てくるという．

注３　注２と同様の理由によって「三時からおります（いる）」「根元からおります（折る）」のように識別できないものが出てくる．

6.2. 非変換

・挨拶

「おはようございます」「いただきます」等の挨拶は引用文と同等の主観的叙述と考え，前処理で感動詞扱いにして非変換とした．

・デアル体の述部

客観的叙述であるか否かに係わらず，デアル体はそのまま出力する．

7. 今回の変換プログラムについての課題

7.1. 接続詞「ですが」等の処理　今回の変換プログラムでは述部だけを処理対象としたが，述部以外についても，接続詞のうち「ですが」「ですから」等は「だが」「だから」と変換することを考えている．

7.2. 語レベルでの調和　今回の変換プログラムではゴザイマス体もデアル体へ変換したが，述部以外の表現については変換対象としなかったので，変換後の述部と丁寧な表現の名詞等が不調和をおこす可能性がある．この点について考慮を加え，例えば「本日はお日柄がようございます」は「本日はお日柄がよい」ではなく，「今日は日がよい」と変換する．「立派な御令息がお二人もございます」は「立派な御令息がお二人もある」ではなく，「立派な息子が二人もある」と変換する．「一昨日調べましたら，明後年は十三回忌でございます」の「一昨日」「明後年」は「おととい」「再来年」と言い換える，等があげられる．

昨年報告した「常体から敬体への変換システム」[文献15]と併せて，これで一応相互変換ができるようになった．注4

謝辞

この研究にあたって一貫処理プログラムの使用を許可してくださった中野洋先生に御礼を申し上げたい．

文献

1. 斎藤秀紀（1968）　漢字かな混り文のエントロピー
『計量国語学』　43／44号　39−45
2. 江川清（1969）　「活用形処理」の自動化に関する一方式
『国立国語研究所報告34　電子計算機による国語研究Ⅱ』　55−79
3. 田中章夫（1969）　漢字かなまじり文を全文カナ書き・ローマ字書きにするシステムについて
『国立国語研究所報告34　電子計算機による国語研究Ⅱ』　107−138
4. 中野洋（1971）　品詞認定の自動化

注4　常体から敬体への変換では文中にデス・マス体を挿入することはしなかった．従って，文中にデス・マス体を含む文章（図4.1.−4.1.b.，4.1.c.参照）をこのシステムで変換し，更に敬体への変換システムを用いて再度変換しても，原文の復元にはならない．

『国立国語研究所報告39 電子計算機による国語研究Ⅱ』 98－120
5. 田中章夫（1971） 新聞語彙調査の同音語と同形語
『国立国語研究所報告39 電子計算機による国語研究Ⅱ』 121－145
6. 国立国語研究所（1971） 電子計算機による新聞の語彙調査Ⅱ
『国立国語研究所報告38』
7. 斎藤秀紀（1971） 漢字かな混り文の文字列
『国立国語研究所ＬＤＰ月報別冊８』 55－86
8. 鈴木一彦，林巨樹編（1972） 動詞
『品詞別 日本文法講座３』
9. 鶴岡昭夫（1973） 文語形・口語形活用語の代表形の変換処理について
『国立国語研究所報告49 電子計算機による国語研究Ⅴ』 121－140
10. 中野洋（1978） 言語処理における一貫処理の研究
『国立国語研究所報告61 電子計算機による国語研究Ⅸ』 17－40
11. 電子技術総合研究所編（1980） 『新編 日本品詞列集成』
12. 水谷静夫，石綿敏雄，荻野孝野，賀来直子，草薙裕（1983） 文法と意味Ⅰ
『朝倉日本語新講座３』
13. 林四郎，荻野綱男，田中幸子，樺島忠夫（1983） 運用Ⅰ
『朝倉日本語新講座５』
14. 田中章夫（1987） 日本語の機械処理
『大阪外国語大学 昭和61年度特定研究 研究成果論文集』 263－284
15. 真田治子（1988） 文体の自動変換 ―ダ体からデス・マス体へ
『計量国語学』 16巻７号 313－326
16. 丸山直子（1989） 格助詞と格と結合価
『計量国語学会 第三十三回大会 研究発表要旨』及び配布資料

（この報告は第三十三回大会での研究発表に加筆したものである．）

（1989年10月20日受付）

***** descriptors and abstracts *****

Report
AUTOMATIC SENTENCE STYLE CONVERSION IN JAPANESE:
CONVERSION TO DEARU-STYLE
SANADA Haruko (IBM Japan, Ltd. / Gakushuin University)
Descriptors: style conversion; style; common style; polite style;
Desu/Masu-style; Dearu-style; conjugation of verb;
Abstract:
This paper deals with an automatic conversion system which converts any sentence style to the dearu-style in Japanese. The following are the characteristics of the system.
(1) It is available on personal computers.

(2) It converts clause-final and sentence-final predicates in all sentences except conversational sentences.
(3) Verb-conversion incorporates a probability logic according to circumstances.
The system was developed with the following considerations in mind:
(1) Conversion of clause-final predicates
(2) Conversion of polite verbs
Two problems still remain. One is the conversion of conjunctions such as desukara, desuga, etc. The other concerns the differenciation of homonymous verb forms; this may be facilitated by examining accompanying postpositions.

（４）文体の自動変換　ダ体からデス・マス体へ

（『計量国語学』第16巻第７号　1988年　収録）

計量国語学第十六巻第七号〔Mathematical Linguistics, vol. 16 no. 7〕1988年

文体の自動変換——ダ体からデス・マス体へ

真田　治子　（日本アイ・ビー・エム㈱　学習院大学大学院委託生）

ディスクリプタ：　文体変換　文体　常体　敬体
ダ体　デス・マス体　活用分析

1. 文体変換の必要性

最近，一般の人々が計算機に触れる機会が増えているが，そのような場でのマン・マシン・インターフェースとしての日本語の自動生成が次第に重要になってくると考えられる．例として音声応答システムで計算機が発する会話文の生成などがあろう．また，手紙文を外国語翻訳システムで翻訳する場合も，出力された常体文を敬体文へ変換することが必要である．

そこで今回はダ体からデス・マス体への文末の自動変換について報告する．

2. 文体変換プログラムの概要

2.1. プログラムの機能　今回作成した文体変換プログラムの主な機能は次の通りである．

・ダ体の文末をデス・マス体に変換する

句点の前の２語を変換対象とした．

・客観的叙述の文末のみ変換する

引用文は変換対象外とした．また，今回は終止形の文末のみ変換する．これについては後で詳しく述べる．

・大きな辞書ファイルを用いず，一部に確率的判断を含む

パソコンのレベルのプログラムを目指したので，動詞の活用の分析の箇所に一部確率的判断が含まれている．また辞書ファイルは約30個で容量は合計約28Kバイトとなった．

2.2　稼動環境　このプログラムはBASICで作成した．ステップ数は約1200となった．また，入出力はMS-DOSのテキストファイルの書式を使用した．前処理として国立国語研究所の一貫処理プログラムを使用し，単位切り・読み仮名付け・品詞認定・活用形認定を行なった．

前処理と変換プログラムとの関係を図2.2.に示す．

ベタ打ちした入力文のファイル（2.2.a.）を一貫処理プログラム（2.2.b.）にかけて出力ファイル（2.2.c.）を得る．この時前処理の誤りは人手で修正しておく．変換プログラム（2.2.d.）はこのファイルを入力とし，文体変換をしてファイル（2.2.e.）に出力する．

2.3.' 入出力文の例　変換プログラムの入出力の具体例を 図2.3.a.（入力）と 図

SANADA Haruko（IBM Japan, Ltd.／Gakushuin University）——An automatic conversion of Japanese sentence style：from Da-style to Desu／Masu-style

図2.2.　前処理との関連図

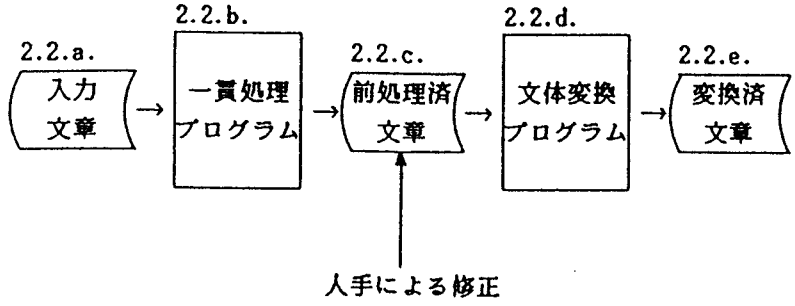


図2.3.a.　入力（＝2.2.c.）

文字	字種	読み	単位	品詞	活用
ミ	I				
ヤ	I				
マ	I				
オ	I				
ダ	I				
マ	I				
キ	I		1	1	
の	H		1	R	
気	し	き			
品	し	ひん	1	1	
の	H		1	R	
あ	H				
る	H		1	E	+
青	し	あお			
紫	し	むらさき	1	1	
の	H		1	R	
花	し	はな	1	1	
が	H		1	R	
風	し	かぜ	1	1	
と	H		1	R	
遊	し	あそ			
ん	H		1	E	9
で	H		1	Q	9
い	H				
る	H		1	E	+
。	E		1	Y	

図2.3.b.　出力（＝2.2.e.）

文字	字種	読み	単位	品詞	活用
ミ	I				
ヤ	I				
マ	I				
オ	I				
ダ	I				
マ	I				
キ	I		1	1	
の	H		1	R	
気	し	き			
品	し	ひん	1	1	
の	H		1	R	
あ	H				
る	H		1	E	+
青	し	あお			
紫	し	むらさき	1	1	
の	H		1	R	
花	し	はな	1	1	
が	H		1	R	
風	し	かぜ	1	1	
と	H		1	R	
遊	し	あそ		1	
ん	H		1	E	9
で	H		1	Q	9
い	H				
ま	H		0		
す	H		1	P	H
。	E		1	Y	

＜凡例＞

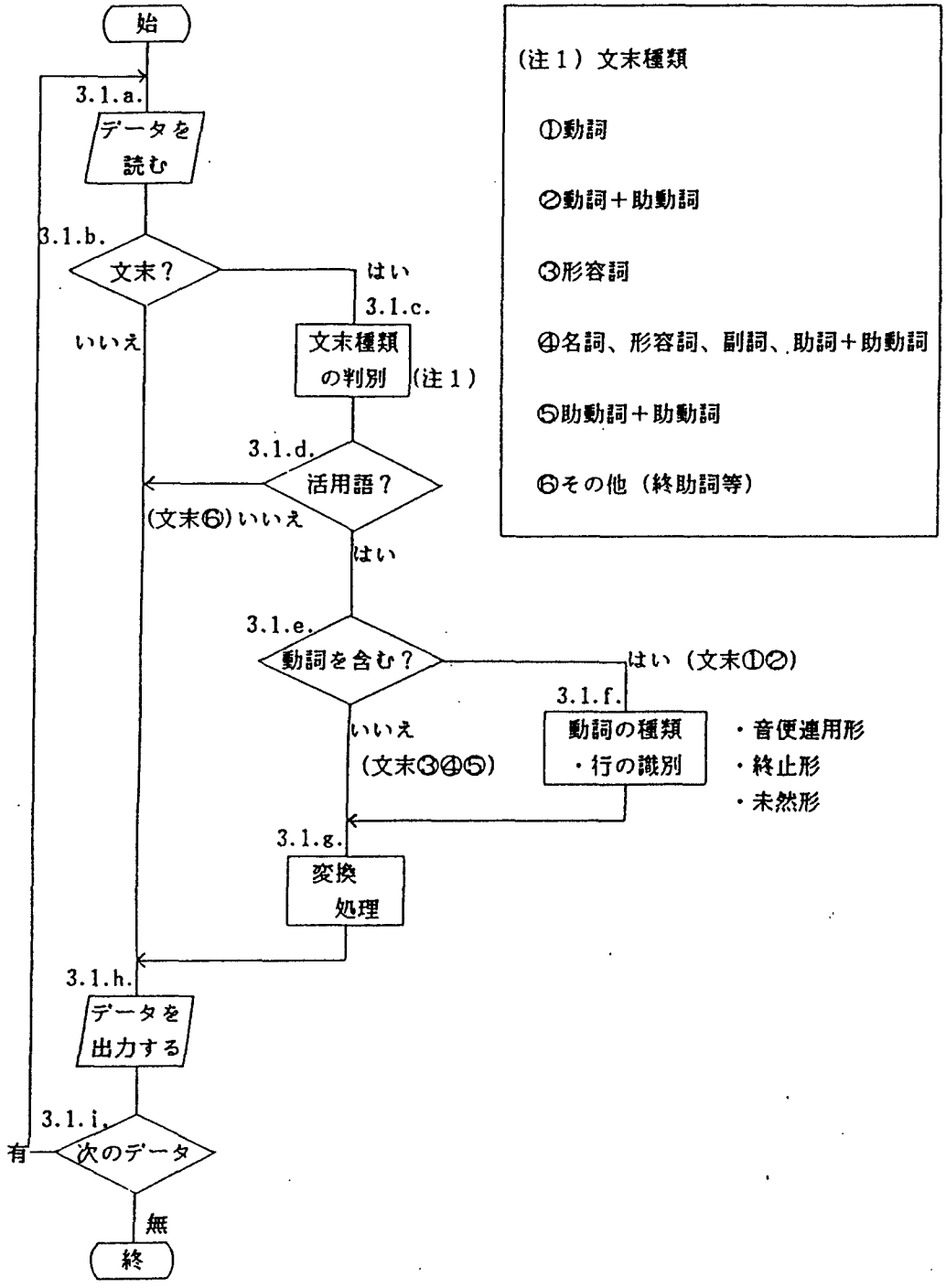
字種I:カナ
　H:かな
　L:漢字

単位0:続く
　1:切れる

品詞1:名詞
　R:助詞
　E:動詞
　Q:助助動詞
　P:助動詞

活用+:終止連体
　9:連用
　H:終止

図3.1.　変換処理全体の流れ



2.3.b.（出力）に示す．これらは各，図2.2.（前処理と変換プログラムとの関連図）のファイル2.2.c.と2.2.e.に相当する．

文章が１文字１行の形で出力され，各文字に字種，読み，単位切り，品詞，活用形の情報が付加される．

3. 変換処理の概要

3.1. 変換処理全体の流れ　変換処理全体の流れを 図3.1.に示す．

まずデータを読む（3.1.a.）．これが文末であれば文末種類の判別処理（3.1.c.）へ行く．文末でなければそのまま出力される．

文末の種類は

(1) 動詞
(2) 動詞 ＋ 助動詞
(3) 形容詞
(4) 名詞，形容詞，副詞，助詞 ＋ 助動詞
(5) 助動詞 ＋ 助動詞
(6) その他（例えば終助詞で終わる文末）

の６つに判別される．(6) その他 の文末は活用語でないのでそのまま出力処理（3.1.h.）に進む．また動詞を含まない文末 (3) (4) (5) はすぐ変換処理（3.1.g.）へ行くが，動詞を含む文末 (1) (2) は，動詞の種類と行の識別処理（3.1.f.）を経由して，変換処理（3.1.g.）へ行く．

動詞の種類と行の識別処理（3.1.f.）では入力された動詞をすべて連用形にするための前処理として識別を行なう．入力された動詞の活用形によって，音便連用形用，終止形用，未然形用の３通りに分かれるが，これについては後述する．

活用語である文末 (1)～(5) は変換処理（3.1.g.）で変換後，出力される（3.1.h.）．次のデータがあれは（3.1.i.），また最初（3.1.a.）に戻る．

3.2. 動詞（音便連用形）の識別処理の流れ　この処理は 図3.1.（変換処理全体の流れ）のうち，動詞の種類と行の識別処理（3.1.f.）の１つに相当する．

図3.2.を参照されたい．

連用形のうち明らかに音便形でないものは変換不要（3.2.a.）と認定される．該当するのはサ行五段，ア行以外の上一段，下一段，サ行変格，カ行変格である．

音便形の可能性のあるものは形によって，ウ音便，イ音便，撥音便，促音便処理に分けられる．語末がこれら４つのどれにも該当しないもの，例えば文語動詞等は「その他」グループに入れられ，変換されない．

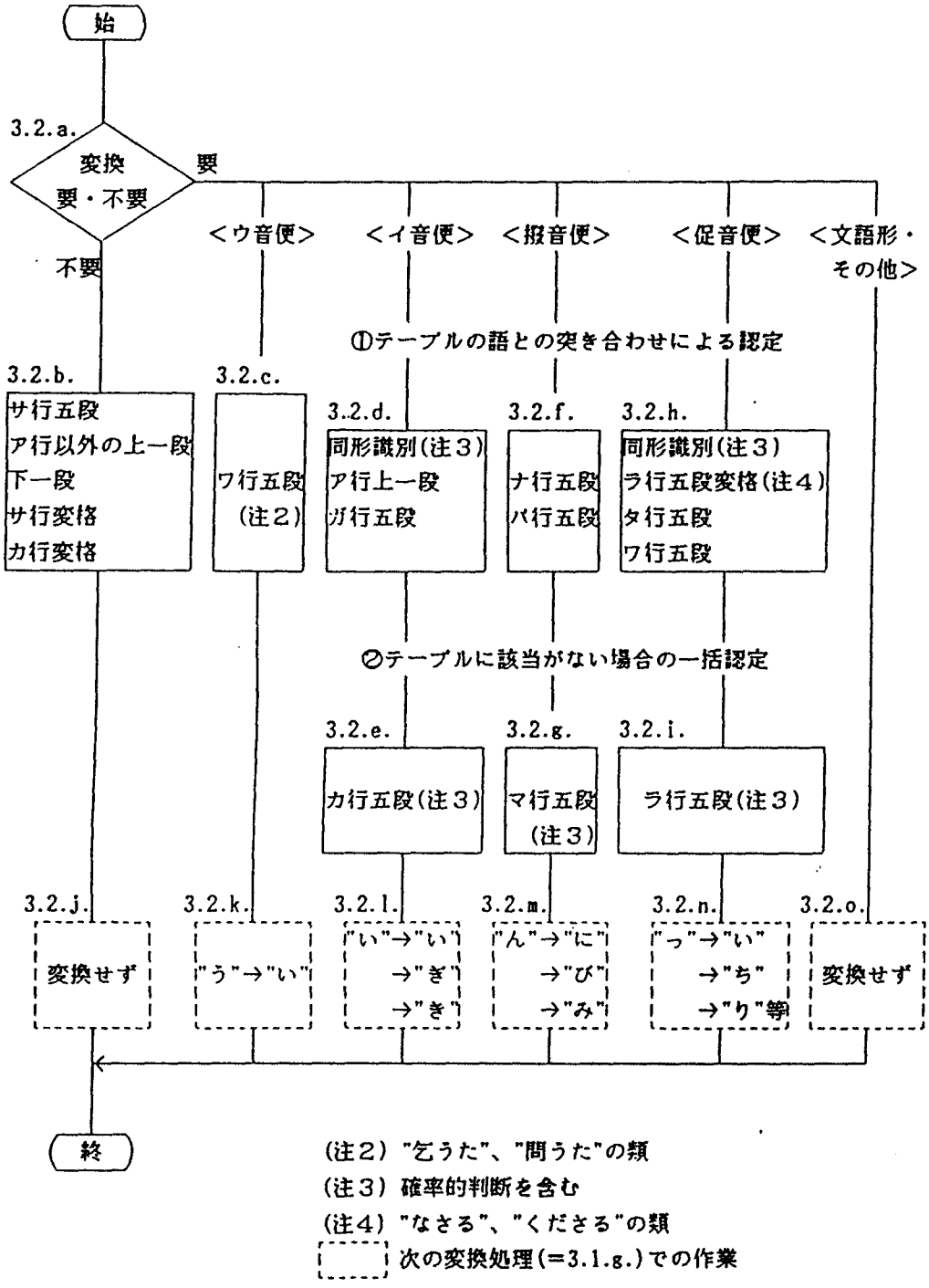
ウ音便というのは「乞うた」「問うた」の類の動詞で，これらはすべてワ行五段と認定される．このグループの動詞は，次の変換処理で語末の「う」が「い」に変換される．

イ音便，撥音便，促音便処理はどれも

(1) テーブルとの突き合わせ
(2) テーブルに該当がない場合の一括認定

の２段階になっている．また表記上は同形でどちらの種類か識別できない場合は，同形識別用テーブルと突き合わせ，強制的にいずれか１つに認定するようにした．今回は，可能

図3.2.　動詞（音便連用形）の識別処理（＝3.1.f.）の流れ



性のある動詞の種類のうち，国立国語研究所「電子計算機による新聞の語彙調査（Ⅱ）」（1971）の「動詞の表」の中で代表形度数が１番多いものに決まるよう，テーブルを作成した．

イ音便では同形識別，ア行上一段，ガ行五段の３つのテーブルを使用する．これらに該当がない時はカ行五段であると一括認定を行なう．例えば「老いた」「泳いだ」の場合には各テーブルと突き合わせてア行上一段，ガ行五段と認定する．「書いた」の場合はテーブルに該当がないのでカ行五段となる．「おいた」という表記の場合には「老いた」（ア行上一段）か「置いた」（カ行五段）が識別できないので強制的にカ行五段と認定する．これは前出の「動詞の表」で代表形度数が「置く」４４，「老いる」１であることから決定した．イ音便処理で認定された動詞はその種類により，次の変換処理で語末の「い」を「い」「き」「ぎ」に変換する．

撥音便では，ナ行五段とバ行五段の２つのテーブルを使用する．テーブルに該当がない場合はマ行五段と一括認定される．撥音便では次の変換処理で語末の「ん」が「に」「び」「み」に変換される．

促音便では同形識別，タ行五段，ワ行五段，ラ行五段変格の４つのテーブルを使用する．ラ行五段変格といっているのは「なさる」「くださる」の類の動詞である．テーブルとの突き合わせの後，該当がない場合には，ラ行五段と一括認定される．促音便の同形識別の例では「かった」という表記が挙げられる．この場合「勝った」（タ行五段），「刈った」（ラ行五段），「買った」（ワ行五段）等の可能性があるが，これも「動詞の表」の代表形度数より，ワ行五段と認定した．以上のようにして認定された促音便動詞は次の変換処理で語末の「っ」が「い」「ち」「り」「き」（「行く」という動詞の場合のみ）に変換される．

音便連用形動詞の識別処理では

- イ音便の同形識別
- イ音便の一括認定
- 撥音便の一括認定
- 促音便の同形識別
- 促音便の一括認定

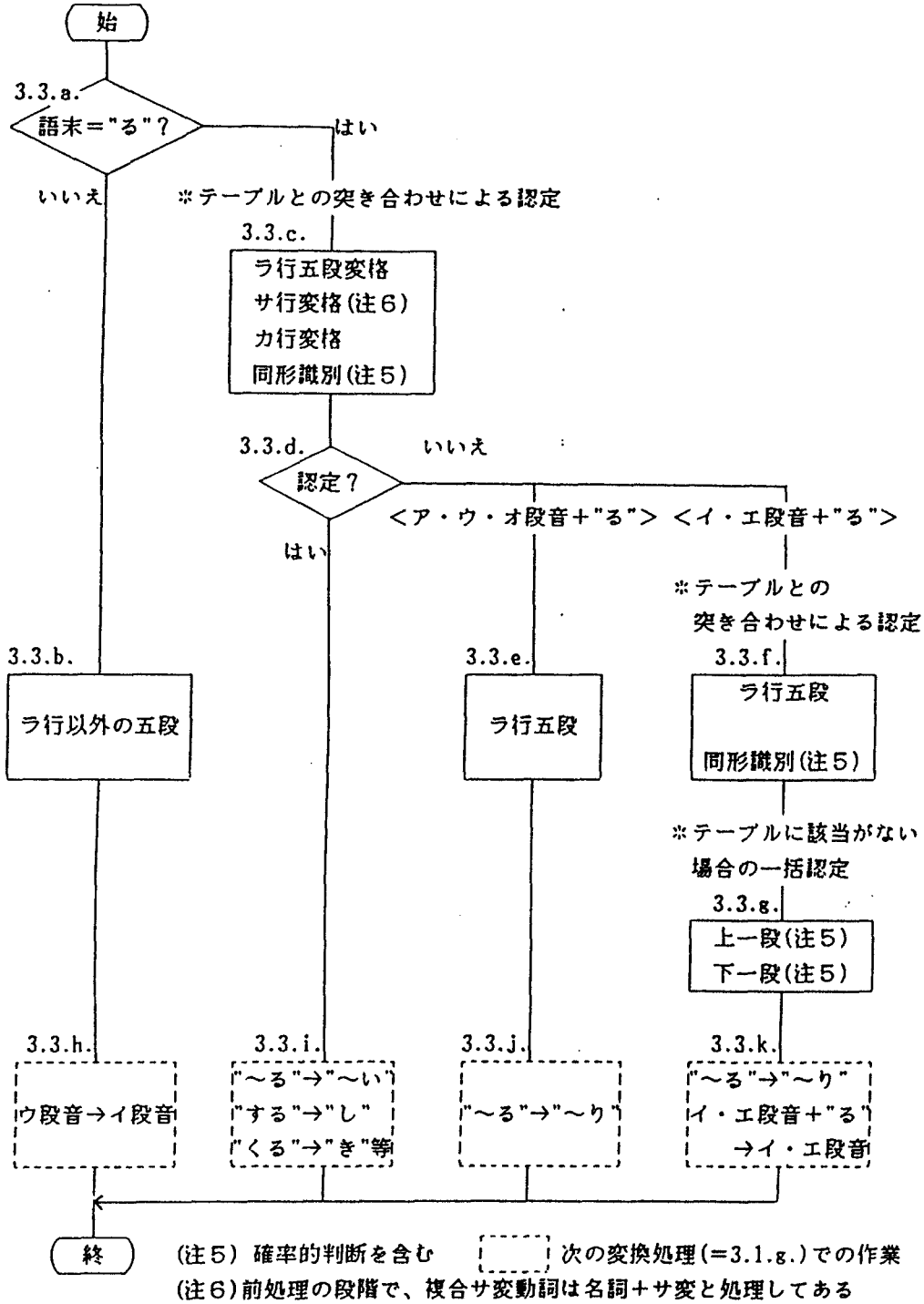
が確率的判断（一番頻度の多い所に落ち着く方法）を含む認定となる．

3.3.　動詞（終止形）の識別処理の流れ　　この処理は 図 3.1.（変換処理全体の流れ）のうち，動詞の種類と行の識別処理（3.1.f.）の１つに相当する．

図3.3.を参照されたい．

終止形では語末が「る」以外のものはラ行以外の五段と認定される（3.3.a.～3.3.b.）．語末が「る」のものはテーブルとの突き合わせ（3.3.c.）によって，サ行変格，カ行変格，前述のラ行変格でないか，探索を行なう．この時，前処理の段階で複合サ変動詞は名詞 ＋ サ変と処理してある．テーブルに該当があった場合には，次の変換処理で，サ行変格の「する」は「し」に，カ行変格の「くる」は「き」に，ラ行変格の語末の「る」は「い」に各変換される．この場合，「する」という表記はサ行変格の「する」とラ行五段の「刷る」と同形，「くる」はカ行変格の「来る」とラ行五段の「繰る」と同形であるが，「動詞の表」によりサ行変格，カ行変格と強制的に認定した．

図3.3.　動詞（終止形）の識別処理（＝3.1.f.）の流れ



サ変，カ変，ラ変のテーブルに該当がなかった場合，まず，ア・ウ・オ段音 ＋ 「る」はラ行五段と認定する．これは次の変換処理では語末の「る」が「り」に変換される．

次に，イ・エ段音 ＋ 「る」は，同形識別とラ行五段の２つのテーブルと突き合わせる．同形識別の例では「へる」という表記を「減る」（ラ行五段）か「経る」（ハ行下一段）か識別するというケースがある．この場合も「動詞の表」の代表形度数によりラ行五段と認定した．同形識別とラ行五段のテーブルに該当がない場合には，イ段音 ＋ 「る」なら上一段，エ段音 ＋ 「る」なら下一段と一括認定を行なう．以上のようにして認定されたイ・エ段音 ＋ 「る」は，次の変換処理では，ラ行五段であれば語末の「る」を「り」に変換し，上一段，下一段であれば語末の「る」を削除する．

終止形動詞の識別処理では

- サ行変格の認定
- カ行変格の認定（同形の場合）
- イ・エ段音 ＋ 「る」の同形識別
- 上一段，下一段の一括認定

に確率的判断が含まれる．

3.4.　動詞（未然形）の識別処理の流れ　　この処理は 図 3.1.（変換処理全体の流れ）のうち，動詞の種類と行の識別処理（ 3.1.f .）の１つに相当する．

図 3.4.を参照されたい．

未然形では入力データが「せ」または「ぜ」ならサ変（ 3.4.a .），語末がア・オ段音ならカ変，五段，サ変の「さ」のグループ（ 3.4.b .），イ・エ段音なら上一段，サ変の「し」，下一段のグループ（ 3.4.c .）と認定する．

このような識別処理は変換を目的としたもので，正しく動詞の種類を認定するためのものではない．同じサ変でも「さ」と「し」は各五段動詞のグループ，上一段のグループとして変換処理される．

変換処理では「せ」「ぜ」のサ変グループは「し」「じ」に，五段動詞のグループは語末のア・オ段音はイ段音に変換される．上一段・下一段グループの語末のイ・エ段音は変換されない．

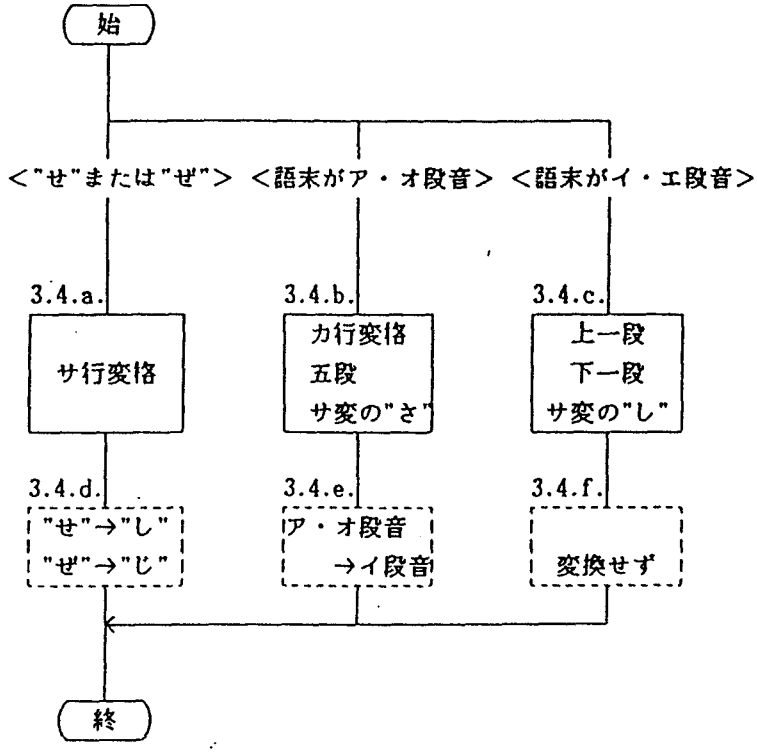
未然形動詞の識別処理では確率的判断はない．

3.5.　変換処理の流れ　　この処理は 図 3.1.（変換処理全体の流れ）のうち，変換処理（ 3.1.g .）に相当する．行，種類を認定した動詞や，変換の必要のある形容詞，助動詞は変換テーブルを用いて変換処理を行なう．このテーブルは，削除情報と追加情報で構成されている．

例えば「書く」という文末はカ行五段動詞終止形であると認定されてこの処理に入る．削除情報は語末の「く」，追加情報は「きます」であることがテーブルからわかるので「書きます」に変換される．

変換テーブルはなるべく文末の種類や品詞等に合わせて，目的を限定し，細分化して持つことにした．複数の変換目的を兼ねた共用テーブルは作成しなかった．これは，テーブルを修正した時に，その影響範囲を確実に把握するためである．これにより，ある誤変換を修正しようとテーブルを書き換えたためにそのテーブルを用いて従来正しく動いていた処理で誤りが出るのを防ぐようにした．

図3.4.　動詞（未然形）の識別処理（＝3.1.f.）の流れ



4.　テスト結果

4.1.　テスト結果の例　　「天声人語」を使用して変換テストを行なった．変換結果の例を 図４.１.a.と 図４.１.b.に示す．

まず，図４.１.a.を参照されたい．

・例４.１.c.（原文）

部分的にベルをやめ，客の反応を観察しながら廃止の部分をふやしていった．

・例４.１.d.（原文）

夕方の混雑時，ベルの鳴らぬ駅の様子を見た同僚がいった．

上記２つの文末のうち，例４.１.c.は「いきました」，例４.１.d.は「いいました」と変換されるべきところであるが，同形識別の確率的判断により両方とも「いいました」と

図4.1.a.　テスト結果の例（1）

東日本旅客鉄道・千葉駅の板倉義和駅長たちは、この八日から、列車が出発する時のベルを全廃しました。勇気ある決断であります。ベルをやめたら苦情がでるかもしれません。乗り遅れの人がでるかもしれません。初日、駅長は不安と緊張でいっぱいでした。午前六時から、ホームに立ちました。苦情はすべて自分が引き受けるつもりでした。だが、覚悟していた苦情はひとつもありませんでした。駆け込み乗車も減りました。「構内が静かになった。いいことだ」という手紙や電話がきました。千葉駅は日に約千二百本の電車、列車が発着します。朝の出勤時は、間断なく予告ベルが鳴り響くことになります。近接する住宅地からの苦情がありました。乗降客の三大苦情も、１ベルがうるさい２放送がうるさい３トイレが臭い、というものでした。まず、ベルの全廃にとりかかりました。だがこれは、日本の鉄道関係者にとっては一種の「革命」です。周到な準備を重ねました。部分的にベルをやめ、客の反応を観察しながら廃止の部分をふやしていいました。夕方の混雑時、ベルの鳴らぬ駅の様子を見た同僚がいいました。「初めは間の抜けたような奇妙な感じだった。そのうち、どことなくのんびりした雰囲気が新鮮でよいものに思えてきた。あわただしいベルが鳴るのが当然、という自分の感覚の方がおかしかったと気づいた」。七年前、当時の国鉄が二、三の駅で「駅構内の放送なし」の実験をしたことがあります。あの時も、大きな駅特有のせかせかした雰囲気がうすらいで、まずまずの評判でした。放送やベルをどこまで減らせるか、一工夫も二工夫もほしいところです。「日本人は個人的には非常に敏感な、デリケートな耳の持ち主だと思う」と詩人の田村隆一さんがある会合で発言しています。「しかし社会的な音に対しては、歴史が浅いので感受性が鈍い」とも語っています。静かさが売りものの静養の地に、のべつ幕なしに音楽が流されることもあります。

例4.1.e.

変換された（4.1.e.）．これについては「〜ていく」という形の語を認めてテーブルに載せることによって判断の精度をあげることができる．

次に図4.1.b.を参照されたい．

・例4.1.f.（原文）

大空にすがって生きようとするのは木の芽であり，木の芽をみる己のことであろうか．

・例4.1.g.（原文）

アカヤシオやカタクリの咲く一帯を保護地域にはできないものか．

上記の２例は終助詞で終わる文末で，変換処理後，原文のまま出力された（4.1.h.〜4.1.i.）．一般に命令，禁止，質問等の表現は「行くな」「欲しいかい」のように主観

図4.1.b.　テスト結果の例（2）

数日前、秩父の山を歩きました。沢ぞいの草地にアズマイチゲの白い花が咲いていました。古峰神社という小さな社のある山の頂にのぼると、何本ものアカヤシオがありました。岩場に根を張って斜めに伸び、やわらかな、淡い紅色の花を咲かせています。花の好みは人さまざまだが、樹木博士の小林義雄さんは「ツツジの女王コンテストになったら容姿、色合いのすぐれたアカヤシオにやはり票が集まるだろう」と書いています。それほど、この花には人をひきつけるものがあります。秩父の山の雑木林はまだ枯れ葉色をまとっていました。離れてみると、アカヤシオの淡い紅やアブラチャンの花の黄が枯れ葉色の世界にしみこんで、みごとな調和をみせています。イヌブナなどの千万の木の芽が光っています。「大空にすがりたし木の芽さかんなる」（渡辺水巴）。大空にすがって生きようとするのは木の芽であり、木の芽をみる己のことであろうか。
夜、このあたりのアカヤシオを盗みに来て、岩場から足をふみはずして死んだ人がいる、と土地の人がいいました。岩にしがみつくようにしてイワウチワが咲いているのがみえます。アカヤシオの花をさらに淡くした色の花です。「このイワウチワも、昔は岩場に花を敷きつめて咲いていたものですが、盗掘でこんなに減りました」。同行してくれた秩父市大田小学校長の守屋忠之さんがいいました。私たちは足をのばしてカタクリの大集落をみてから帰途につきました。草地に咲く姿を先ほど見たばかりのアズマイチゲの群落がそっくり姿を消しているのに、驚かされました。石の陰の一輪だけが命拾いをし、あとは全部、盗掘にあったらしいです。「チチブドウダンやシロヤシオなどの立派な木が盗み去られたこともあります。節分草もカタクリもイワウチワも、年々秩父の山から消えてゆきます。残念です」と守屋さんはいいます。山の神の怒りを軽くみてはいけません。アカヤシオやカタクリの咲く一帯を保護地域にはできないものか。

例4.1.h.

例4.1.i.

的な表現であり，動態的，現場的叙述であるといえる．しかし，文体変換では客観的叙述が主たる対象となるように思われる．そのため，今回は命令，禁止，質問のような動態的，現場的，主観的叙述は操作しないという方針をとった．その結果，例４.1.h.と例４.1.i.は「～でありましょうか」，「～ものでしょうか」の形にはならなかった．

4.2.　テスト集計結果　　このようにして８日分の「天声人語」についてテストを行なった．

文末数は200，そのうち引用文であるため変換対象外としたのが３１，残り169が変換対象で，うち通過数は168，失敗数は１であった．この失敗の１件は前述の 例４.1.c.の「いった」の誤変換である．通過率は99.4％（168／169）であった．

5. 今回のプログラムで予想される誤変換，非変換のケース

5.1. 誤変換

・同音異活用語の確率判断

「行った」という表記はカ行五段，「言った」という表記はワ行五段と判断するが，前述のテスト例のように仮名で「いった」と表記してあるものはワ行五段と確率的に判断する．

・テーブルにない音便連用形／終止形動詞の識別

テーブルは国立国語研究所「電子計算機による新聞の語彙調査（Ⅱ）」（1971）の「動詞の表」から作成した．従って表にない動詞の場合は一括認定を行なうので誤変換の可能性がある．例として「いざなう」，「さまよう」等の語が考えられる．これらはワ行五段であるが，ラ行五段に一括認定される．

5.2. 非変換

・句点以外の記号（感嘆符等）で終わる文末

例えば，「優勝だ！」という文にはダ体と感嘆符を合わせて使うことによって書き手の強い感情を表しており，変換すべきではないと考える．

・命令形で終わる文末

「手をあげろ」，「お恵みあれ」など命令形で終わる文末も句点以外の記号（感嘆符等）で終わる文末と同様の理由で変換しなかった．

・終助詞で終わる文末

終助詞については「な」，「よ」等は変換すべきでないと考えた．但し「か」についてはそれほど現場的な表現ではないと思われるので，変換してもよいのではないかと考えている．また客観的叙述であるか否かに係らず，「です」，「ます」の文体はそのまま出力する．

6. 今後の課題

6.1. 今回の変換プログラムについての課題

・終助詞「か」で終わる文末の処理

前述のようにこれについてはまだ検討が必要である．

・助動詞「ない」の変換

形容詞では「ない」は「ありません」，「なかった」は「ありませんでした」に変換した．一方，助動詞では「～ない」は「ません」に変え，「～なかった」は「です」を付けて「～なかったです」という形にした．例えば「行かなかった」などは「行きませんでした」とした方が自然かもしれないが，今回のプログラムでは句点の前の２語だけを変換するようにしたため，このような結果になった．句点の前の３語を採ればどちらの形にも変換できる．

・「～てる」，「～とく」，「～ちゃう」の変換

これらの文末は現在は前処理の仕方によって変換されずに出力されるか，確率判断で誤変換される可能性がある．これは助動詞の１種類とみなしてテーブルに加えることにより変換が可能になる．

6.2.　文体変換処理における課題

・変換の観点

変換の観点は今回行なった文末の変換の他に次のものが考えられる.

ひとつは「文中の終止部の制限付き変換」で，次の例６.１.ａ.～６.１.ｂ.のようにすべての終止部を強制的に変換するのではなく，日本語として自然なものだけを変換しなくてはならない.

・例６.２.ａ.

枝にとまっています鶯が鳴きました.（×）

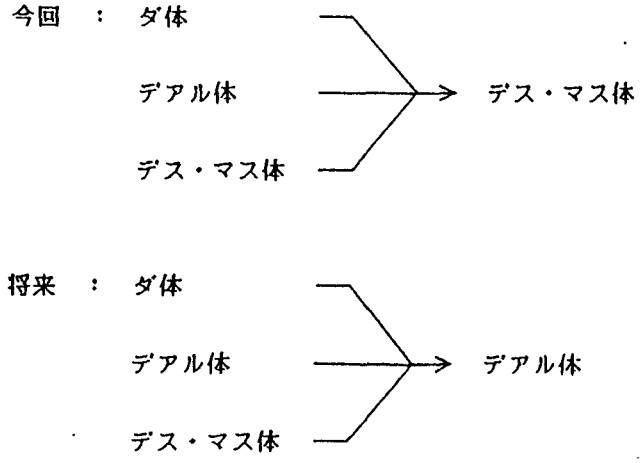
・例６.２.ｂ.

天気になりましたので散歩にでかけました.（○）

もうひとつは「文体に応じた語レベルの調和」を行なうことで，例えば，ダ体で「みんな」となっているものをデス・マス体で「みなさん」と言い換える，或は「あさって」を「明後日」と言い換える，などが考えられる.

・各種文体の統一変換（図６.２.ｃ.）

図6.2.c.　各種文体の統一変換



今回はダ体・デアル体からデス・マス体へ変換することを目標としたが，デス・マス体が入力の場合にはそのまま出力されるので，デス・マス体への統一変換システムであると考えることもできる.

各種の文体変換のうち，次に必要性が高いのはすべての文体を，現在最も一般的な書き言葉の文体であるデアル体に変換することであろうと考えている．これが実現できれば例えば色々な文体で分担執筆した文章を集めて報告書やマニュアルを編集する時などに役立つと思われる.

謝辞

この研究にあたって一貫処理プログラムの使用を許可してくださった中野洋先生に御礼を申し上げたいと思う.

文献

１．斎藤秀紀　（1968）　漢字かな混り文のエントロピー
『計量国語学』　43／44号　39－45
２．江川清　（1969）　「活用形処理」の自動化に関する一方式
『国立国語研究所報告34　電子計算機による国語研究Ⅱ』　55－79
３．田中章夫　（1969）　漢字かなまじり文を全文カナ書き・ローマ字書きにするシステムについて
『国立国語研究所報告34　電子計算機による国語研究Ⅱ』　107－138
４．中野洋　（1971）　品詞認定の自動化
『国立国語研究所報告39　電子計算機による国語研究Ⅰ』　98－120
５．田中章夫　（1971）　新聞語彙調査の同音語と同形語
『国立国語研究所報告39　電子計算機による国語研究Ⅰ』　121－145
６．国立国語研究所　（1971）　電子計算機による新聞の語彙調査Ⅱ
『国立国語研究所報告38』
７．斎藤秀紀　（1971）　漢字かな混り文の文字列
『国立国語研究所LDP月報別冊８』　55－86
８．鈴木一彦，林巨樹　（1972）　動詞
『品詞別　日本文法講座３』
９．鶴岡昭夫　（1973）　文語形・口語形活用語の代表形の変換処理について
『国立国語研究所報告49　電子計算機による国語研究Ⅴ』　121－140
10．中野洋　（1978）　言語処理における一貫処理の研究
『国立国語研究所報告61　電子計算機による国語研究Ⅸ』　17－40
11．電子技術総合研究所編　（1980）　『新編　日本品詞列集成』
12．田中章夫　（1987）　日本語の機械処理
『大阪外国語大学　昭和61年度特定研究　研究成果論文集』　263－284

（この報告は第三十二回大会での研究発表に加筆したものである．）

（1988年11月２日受付）

***** descriptors and abstracts *****

Report
AN AUTOMATIC CONVERSION OF JAPANESE SENTENCE STYLE:
FROM DA-STYLE TO DESU/MASU-STYLE
SANADA Haruko (IBM Japan, Ltd. / Gakushuin University)
Descriptors: style conversion; style; common style; polite style;
Da-style; Desu/Masu-style; conjugation of verb
Abstract:
The automatic generation of Japanese sentences is becoming very important as a subject of man-machine-interface. This paper is about an automatic conversion system used to convert Da-style to Desu/Masu-style. The following are the characters of the system.
(1) It is available on the personal computers; therefore, its dictionary files are very small.
(2) It converts only the end of the objective description.
(3) It converts verbs using a probability logic according to circumstances.
The automatic conversion system has two problems. One is about the subject of conversion, and the other is about the style of sentences in aim. There are two move targets for the former problem besides the conversion of the end of sentences, which is reported in this paper. These two are as follows:
(1) The conversion of predicates with limitation.
(2) The conversion of words to unify the word level used in a sentence.
As for the latter problem, it is necessary to consider the conversion of varied sentences to unified sentences, especially to Dearu-stle.