

日本語教育のための文データベース構築

中尾桂子¹⁾ 本村康哲²⁾

¹⁾ 姫路獨協大学大学院言語教育研究科

²⁾ 姫路獨協大学情報科学センター

あらし

本研究では、小学校教科書の文データベースを作成し、教科学習を考慮した日本語教育に必要とされる学習事項の抽出を行う。外国人児童は、中級以降の日本語学習で、検定教科書に対応できるような日本語能力を習得していかなければならない。そのためには、初級日本語指導で学んだ学習事項と教科書の実文とを関連させる必要がある。そこで、文データベースを利用し、教科毎の文末表現の使用分布や複文の使用形態から、実文中の構文の使用状況を調査する。

Construction Sentence Database for Japanese Education

Keiko Nakao¹⁾, Yasunori Motomura²⁾

¹⁾ Graduate School of Language Education, Himeji-Dokkyo University

²⁾ Information Science Center, Himeji-Dokkyo University

Abstract

In this study, a Japanese sentence database is constructed from elementary school textbooks for extracting a syllabus for intermediate and above level foreign pupils. The sentence database consists of actual sentences, morphemes and parts of speech which are analyzed by the Japanese morphological analysis system *ChaSen*. And the distribution of the basic sentence patterns and the expressions used to combine sentences in each of the general school subjects is researched, to make clear the relationship between the Japanese basic sentence patterns and the actual sentences used in the textbooks.

1. はじめに

平成元年度の入国管理法の改訂以来、国内の義務教育機関で学ぶ外国人児童の数は年々増加してきた[1]。これらの外国人児童が日本人児童と共に学んでいくためには、日本人児童と同等の日本語能力が必要となる。このことを考慮して、図1のように、編入時に適応指導を主とする日本語の指導が初期指導としてなされるようになり、昨今ではさらに、初期指導以降、教科学習の補充を含めた日本語指導の必要性が述べられるようになってきた[2]。

初期指導では、編入後の学校教育に適応させるために、学校生活に必要な語彙や表現の紹介から始まり、コミュニケーション用の日本語指導を行う。教育機関や地域により、教科教育に必要な語彙・表現まで指導されることもある。これは、一

般的な成人に対する初中級日本語教育に相当し、概ね、学校生活に適応できたと担任教師が判断した段階で終了されることが多い。

現在、外国人児童の日本語指導の教材としては、『ひろこさんのにほんご』と『にほんごをまなぼう』シリーズが広く使用されている[2]。『ひろこさんのにほんご』は、成人に対する一般的なコミュニケーション用の指導内容を、外国人児童の生活に即した内容に書き換えて、初級・初中級の日本語教育内容としており、コミュニケーション用の日本語教育の性格が強い。もう一方の『にほんごをまなぼう』シリーズは、文部省から外国人児童に支給される日本語教材で、教科学習に対応する日本語学習教材とされている。こちらは外国人児童への教科学習や教科内容の補充を考慮した日本語指導を目標に構成されている。漢字

の指導に関する指導項目もあるが、主に小学校3年次からの理科、社会、算数の内容をひとまとめにしたものとなっていて、学年に準拠したものではない[2]。したがって、児童によっては、発達段階に合わない未知の概念を未知の日本語で学習することになり、日本語学習教材として、また、学習内容の補助教材としての役割を果たしているかどうかは疑問である。

つまり、現状では検定教科書で使用されている日本語にどの程度対応できるか考慮されないまま、編入当初に指導される話し言葉の運用能力で、書き言葉に対応させている。

学習活動を支える日本語運用能力と、一般的なコミュニケーション用の日本語運用能力とは、内容や運用方法に違いがある。学習活動に対応するためには、コミュニケーションで用いられる基礎的な日本語運用能力とは異なる能力、すなわち、抽象概念や思考内容を理解し表現するための学習用日本語運用能力が必要となる[3]。

会話能力と読み書き能力は異なることから、話し言葉中心の初級日本語教育だけを受けてきた外国人児童には、話し言葉と書き言葉の関連が捉えきれず、学習活動が困難になると考えられる。日本人児童と同様に学習するための適切な学習用日本語の運用訓練がなされているとは言い難く、現状の日本語指導では指導期間が限られていることもあり、教科学習への十分な効果が期待できない。

適応指導の範疇で指導される初級日本語教育は、生活で一般的に必要とされる会話能力の習得

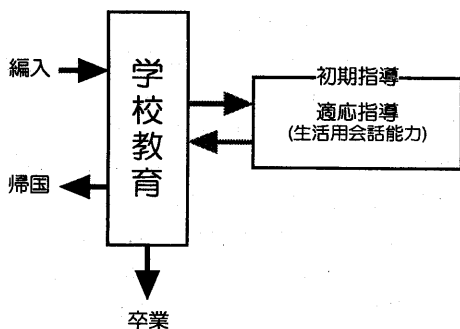


図1：外国人児童の日本語指導体制

を目標としているが、教科学習を円滑に進めるためには、初期の適応指導終了後、教科学習時の学習活動が行えるようにするための日本語指導がさらに必要である。

また、学習活動で用いられる日本語は書き言葉を使用することが多く、授業は検定教科書に添って進められる。したがって、外国人児童が話し言葉だけでなく、教科書の文に対応できるように、学習活動に準じた個々の文法項目の使い方を指導していかなければならない。

そのためには、検定教科書に対応できる日本語を指導できる環境と体制を整えるべきである。まず、重要な構文の教科書の出現状況等、検定教科書の日本語について教科学習に適した中級日本語指導のシラバスを選定する必要がある。また、初級日本語として学習してきた日本語と検定教科書の日本語の関連について、指導者が明確に把握しておく必要がある。

ところが、現段階では、検定教科書の文法項目や使用状況が明確に把握できる資料が存在しない。このため、検定教科書で使用されている日本語が日本語教育とどのように関連しているのかが明らかにされていない。

そこで、本研究では、外国人児童が検定教科書を読解する際に指標となる構文の要素を明らかにするために、検定教科書における構文の出現状況を、文末表現と複文を構成する接続表現から調査する。これにより、日本語教育と検定教科書の日本語との関連を明確にし、さらに、学校教育に対応する日本語指導のシラバス選定の根拠となる基礎資料の作成を行う。

2. 文データベースの構築

今回、文データベース構築の対象とする教科書は、奈良市で平成8年度から使用されている国語、算数、理科、社会、生活の5教科6学年分、合計40冊の検定教科書である。

これらの教科書に掲載されている文のうち、まず、文末に句点を含む文についてテキストデータ化を行う。つぎに、このテキストデータをもとに、1文を1レコードとする文データベースを構築

する。文データベースの構造は、第1フィールドを出版社、第2フィールドを教科、第3フィールドを学年、第4フィールドを教科書毎の行番号、第5フィールドを原文とする。

さらに、第5フィールドの原文を日本語形態素解析システム「茶筌」で処理することによって、形態素に分割された原文、読み、原形、品詞情報が得られる。これらの情報はそれぞれ分解された要素を空白で連結して各フィールドを構成することにし、第6,7,8,9フィールドとする。

ただし、第9フィールドの品詞情報については、一つの形態素について複数個の情報が得られる。茶筌では、形態素の集合を二階層に分類し、それぞれ形態品詞および品詞細分類と呼んでいる。また、動詞、形容詞、形容動詞などのように、後に接続される形態品詞に応じて変化する形態素については、活用型によって分類し、実際にとり得る表層的な個々の形態を活用形と呼んでいる[4]。これらの情報は、茶筌の実行時のオプションを付加することによって選択することができる。ここでは、形態素の品詞細分類(なければ品詞または「未定義語」と表示)、活用型、活用形を出力させる。もし、複数の品詞情報が得られる場合には、品詞細分類、活用型、活用形等の品詞情報を“-”で連結し、さらに、各形態素の品詞情報を空白で連結することによって第9フィールドを構成する。文データベースの例を図2に示す。

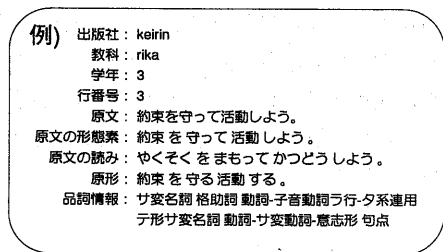


図2：文データベースの例

3. 文末表現と接続表現の分類調査

多くの場合、述部は文末に位置しており、構文の核をなしている。この性質を利用すれば、文末表現によって文を文型ごとに分類することが可能である。

文の構造は、接続表現を含むか否かによって、複文と単文に分類することができる。複文は述部が2つ以上で構成されている。単文は、述部を1つだけ含み、接続表現を含まない[5]。このため、接続表現の有無によって複文/単文を判定できる。ただし、今回は動詞性の形態素に後接される接続表現についてのみ対象とする。また、動詞述語文が名詞を修飾する連体修飾節は接続表現を持たないため、接続する品詞の並び方で判定することにする。

文末表現、単文と複文、初級の知識で認識できる範囲、そして、複文を構成する接続表現[6]の1文中の出現頻度について調査し、検定教科書で使われる文構造の概要を明らかにする。

3.1. 調査方法

図3に処理手順を示す。まず、国際交流基金の「日本語能力試験出題基準」[7]から文末表現と接続表現を抜粋整理し、文末表現リストと接続表現リストを作成する。

リストは、第1フィールドをリスト番号、第2フィールドを原形、第3フィールドを品詞情報とし、パターンを正規表現で記述したものとなっている。図4に文末表現リストの例を示す。

これらのリストの第2、第3フィールドを文デー

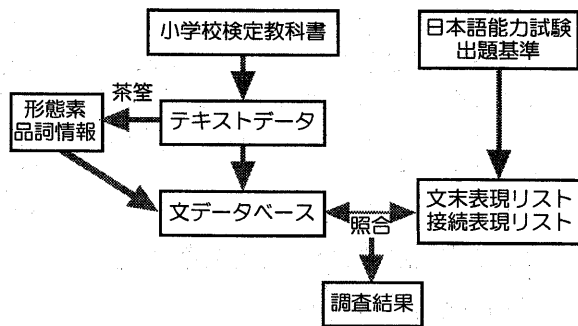


図3：処理手順

データベースの第8、第9フィールドと照合し、適合したリスト番号を文データベースに付加して行く。

照合が終了した後、文末表現と接続表現の実文中における出現頻度、また、それらの出現状況について調査する。調査項目は、文末表現、単文・複文、初級構文・初級外構文、接続表現である。

3.1.1. 文末表現

文データベースと文末表現リストを照合し、文末表現リストに適合する場合には初級の文末表現であるとし、文データベースの第10フィールド目にリスト番号を付加する。適合しない場合には初級外文末表現を含む文であるとし、0を付加する。その後、各教科と各学年について、文末表現リスト中の各パターンの出現頻度を計測する。これによって、各初級文末表現と初級外文末表現の出現頻度とその割合を各教科と各学年毎に調査する。

文末表現リストの基礎となる『日本語能力試験出題基準』には初級文末表現が86種掲載されているが、接続関係を考慮して下位分類すると172種類ある。そこで、大きく使用の傾向を捉えるために、構造文型や話者の意図が含まれる表現文型などを形態や意味機能により8種の文末表現グループ(以下、グループ)に分類して整理を行う。

グループ1はいわゆる動詞述語文、名詞述語文、形容詞述語文などの基本的な構造文型である。

グループ2は動詞述語文の中でも、形態的に動詞1語で述部を構成していない動詞述語文のグループである。いわゆる補助動詞の述語文や複合動詞の述語文である。

グループ3は助動詞文である。いわゆる助動詞ではなく、茶筌でいう助動詞で、「ようだ」「そうだ」「らしい」「つもりだ」等の文を指す。

グループ4は引用助詞を使用する文で、「思う」「言う」が引用助詞「と」や引用の「ように」等と共起する場合の文である。

グループ5は「なる」「する」の表現である。

グループ6は“やりもらい”の授受表現で、「さしあげる」などは敬語になるが、構文上は授受表現のグループに入るので敬語としては分類しなかった。

グループ7は尊敬語や謙譲語のグループである。

グループ8は、「～てください」や「～てもいいです」等、1～7以外の表現文型である。

3.1.2. 接続表現

文末表現と同じように、文データベースと接続表現リストを照合する。接続表現リストは初級接続表現のリストと初級外接続表現リストの2種類を用意し、それぞれを接続表現リスト1と接続表現リスト2とする。照合の結果、適合する場合にはそれぞれのリスト番号を文データベースの第11、12フィールドに付加する。ただし、接続表現が文中に複数含まれる場合には、リスト番号を“-”で連結してから付加する。適合しない場合には、0を付加する。その後、各教科と各学年毎について、文末表現リスト中の各パターンの出現頻度を計測する。

さらに、初級構文と初級外構文について出現頻度を計測する。初級構文とは、初級文末表現の単文と、さらに、初級文末表現に初級接続表現が1回含まれている複文であるとする。したがって、初級外構文としては、初級文末表現以外の文末表現を持つ単文と文末表現に関係なく初級外の接続表現が使用されている複文となる。ただし、初級接続表現が複数使用されている複文については、初級構文学習時に取り立てて指導されないことが多いため、初級外と考える。

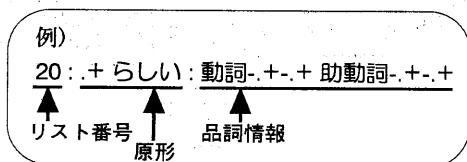


図4：文末表現リストの例

4. 調査結果

4.1. 初級文末表現と初級外文末表現

各教科と各学年における文末表現グループの占める割合をそれぞれ図5(a)および(b)に示す。全体の傾向として、グループ1の占める割合が顕著であった。グループ2を含めると、基本的な文である動詞述語文、名詞述語文、形容詞述語文は、各教科で使用される文のほとんどを占めていた。全文中における初級文末表現の占める割合は93.3%であった。これは、教科書の文のほとんどが基本的な文末表現で構成されていることを示している。

使用されている文末表現を詳しく見ると、「～てあげる」「～ほうがいい」「～つもりだ」「～はずだ」「～はずがない」「～い(形容詞述語文)と思う」等の25種類については、顕著な使用が見られなかった。

一方、初級外の文末表現と判断されたものを詳しく見ると、基礎的な構文の応用と考えられる「倒置文」「従属節を主節と入れ換えた倒置的な文「～は～からです」、節に後続する文が省略された省略文(遠足に行ったら・・・)等が多く、

これらの基礎的な構文を応用した構文の方が初級外の表現文型より多い頻度で使用されていた。初級外の文末表現としては、「～まい」「～だっけ」「～たのなんのって」「～やしない」「～たか!(驚き)」等の15種類程度の表現文型で、出現頻度も全文中10回程度と極少数であり、そのほとんどが倒置文や名詞1語文であった。

図5(a)で各教科毎にグループの割合を見ると、算数で使用される初級文末表現の割合が98.4%で最も多かった。また、グループ8の割合が14.9%と他教科より多いが、この内訳を詳しく見ると、表現文型「～なさい」の出現頻度が顕著であり、使用文末表現の種類は他教科よりむしろ少なかった。

生活科については、表現文型の占める割合が多くなると予想された国語よりもグループ8の割合が多かった。また、他教科よりも初級文末表現を使用している割合が少なかった。さらに、理科、社会よりもグループ5の占める割合が多く、補助動詞文のグループ2や助動詞文のグループ3の使用が少ないことが分かった。

その他、グループ3の占める割合が、理科は

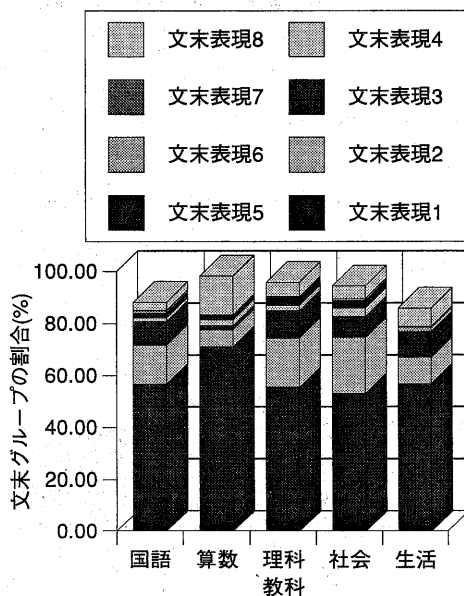


図5(a)：文末表現グループの占める割合(教科別)

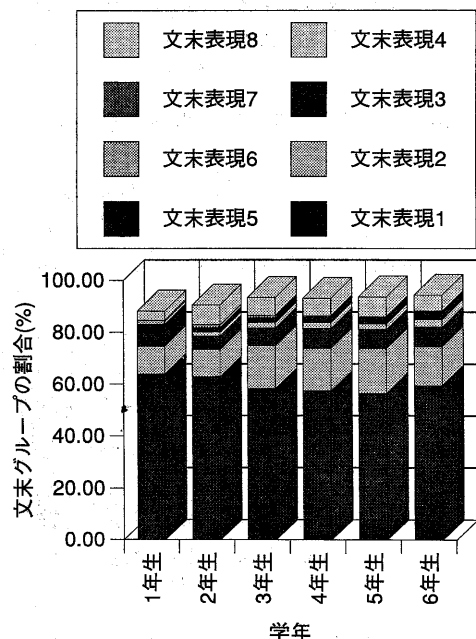


図5(b)：文末表現グループの占める割合(学年別)

10.6%、国語は9.2%、社会は8.0%と比較的多かった。グループ4は、社会、理科、算数でのみ使用されていた。グループ6、7については、生活で若干使用が見られるが、概ね少なかった。

図5(b)で各学年毎にグループの割合を見ると、グループ1の出現頻度が学年を追う毎に若干ながら減少していく傾向にある。また、低学年ではグループ2の占める割合は低いが、中高学年では高くなっていた。低学年ではグループ6の使用が見られるのに、中学年以降は見られない。1年生ではグループ8の頻度が他学年より若干低かった。

4.2. 単文と複文

各教科と各学年における単文/複文の占める割合をそれぞれ図6(a)および(b)に示す。全体的な傾向として、複文の方が多く、全体の75%を占めていた。

図6(a)で教科別に複文の占める割合を見ると、理科と社会ではそれぞれ83.4%と84.4%で最も多く、次いで、国語で67.2%、算数で66.0%、生活で57.0%であった。

図6(b)で学年別に見ると、学年が上がるに連れて複文の占める割合が増加していくが、逆に単文の占める割合は減少していくことが示されている。

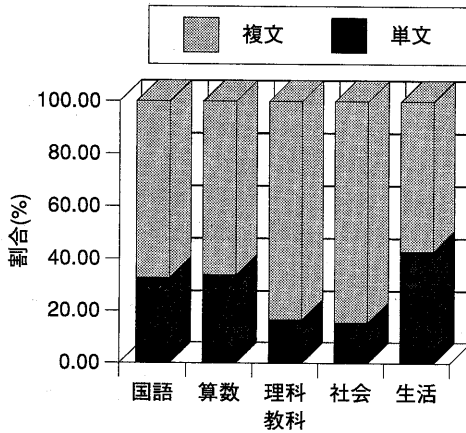


図6(a)：初級の単文・複文の割合(教科別)

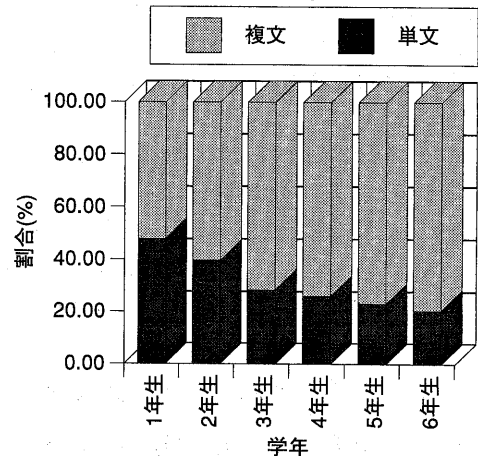


図6(b)：初級の単文・複文の割合(学年別)

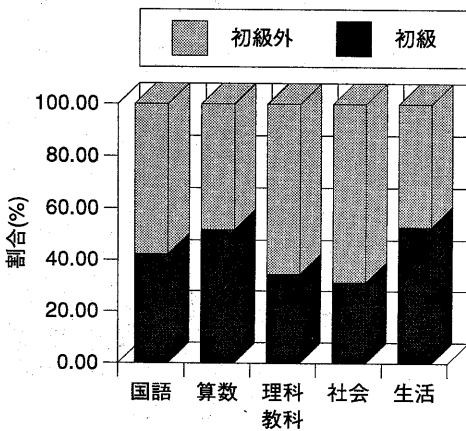


図7(a)：初級/初級外の文末表現(教科別)

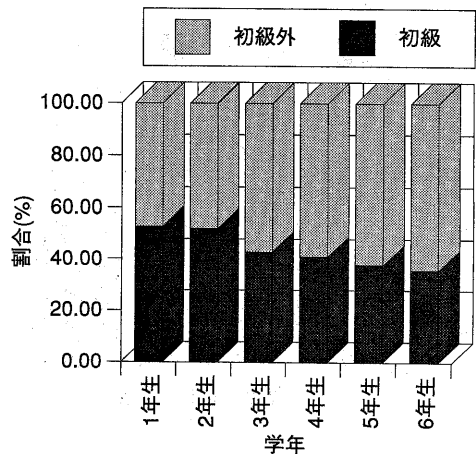


図7(b)：初級/初級外の文末表現(学年別)

4.3. 初級構文と初級外構文

各教科と各学年における初級構文/初級外構文の占める割合を図7(a),(b)に示す。

図7(a)で教科別に見ると、初級構文の占める割合は教科によって特徴があり、理科と社会では初級構文の占める割合は、それぞれ34.5%と31.4%と比較的低いが、生活と算数ではそれぞれ52.0%と53.1%で半分以上を占めていた。

図7(b)で学年別に見ると、学年が上がるに連れて初級構文の占める割合が減少していくが、逆に初級外構文の占める割合が増加していくことが示されている。

4.4. 複文の構造

各教科と各学年における初級複文の占める割合をそれぞれ図7(a)および(b)に示す。

初級複文は、初級の接続表現が1回出現する複文(1/0)である。これは複文全体の20%程度の割合であったが、教科別にも、学年別にも、使用に偏りが見られなかった。初級の接続表現が複数回出現する複文(2/0)と初級の接続表現が複数回と初級外の接続表現が1回出現する複文(2/1)については、初級複文の発展的応用となる文である。これら3つの複文形態を全てあわせると、全複文中の63.4%を占めていた。

図7(a)で教科別の複文構造を見ると、算数と生活以外は、初級の接続表現が複数回使用される複

文の占める割合が多かった。

図7(b)で学年別に見ると、1年生は、1/0の方が2/0, 2/1より多かった。また、初級接続表現を含む複文の割合が3年生で急に増加しているが、以後の学年では、概ね同様の割合であった。2/1は、学年が上がるに連れて徐々に増加していた。

また、個々の接続表現の組み合わせパターンには特徴は見られなかったが、文末表現と同様に、接続表現の使用においても教科毎に偏りがあった。しかし、動詞のテ形、形式名詞コト、条件文のト、ヨウニなど、よく使用されている物は限定されていた。さらに、よく使用される接続表現でも、1文中に1度しか出ないものと何度でも出現するものがあった。

5. 考察

全文中の初級文末表現の割合と初級外とされる文末表現の内容や種類を調査した結果、初級で学習する文末表現で検定教科書には使用されていないものがあったが、文末表現は初級で学習する項目で構文上はほぼ理解可能だと考えられる。また、教科別、学年別で使用に偏りが見られる文末表現や接続表現の使用状況については、指導比重を教科別、学年別に考慮すべきであることがわかった。

一般には、高頻度で使用される文を類型化した物が基本文型であるとされ、多くの一般日本語教

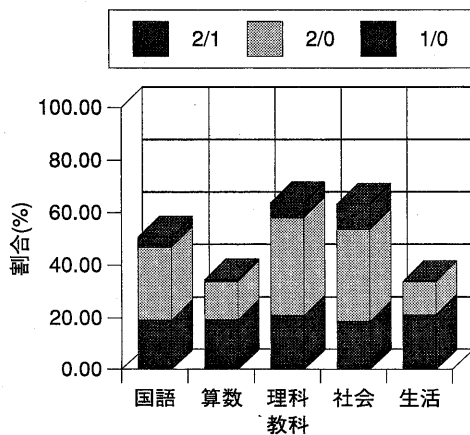


図8(a)：初級複文の占める割合(教科別)

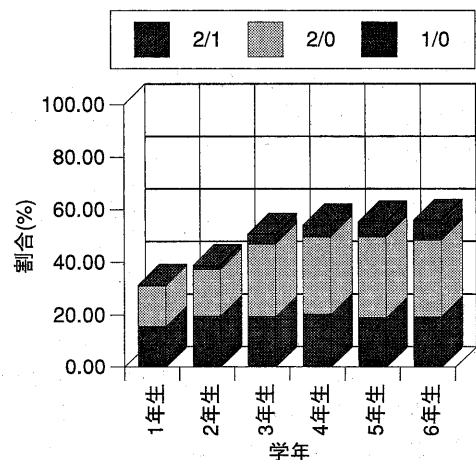


図8(b)：初級複文の占める割合(学年別)

育において文の類型から基本文形が設定されている。同様に、検定教科書の日本語を目標に日本語教育を行う際には、各教科の教科書で頻繁に出現する文型が検定教科書の基本文型と考えられる。したがって、学習用日本語教育のためには、各教科の基本文型を選定しておく必要があり、それをもとに、教科別に指導すべき構文シラバスを組むことが有効な手段である。

本研究により明らかになった文末表現や接続表現の結果を参考にして学習用日本語のシラバスとカリキュラムを構成するためには、今後、学習用の日本語教育と一般日本語教育とをより明確にする必要がある。したがって、さらに調査を進め、各教科毎の基本文型、基本的な複文構造[6]、基本的な文法機能語や語彙、慣用語句等を抽出し、類型化を計るべきである。このため、教科書の語彙調査によって文法機能語を抽出し、文末で述部を形成する文法項目、複文の指標となる接続表現、その他の文法項目を抜粋することが必要である。

最後に、言語教育としての日本語教育を念頭に置いた場合、初級日本語教育の次には、中級日本語教育として、話し言葉中心の日本語教育と書き言葉中心の日本語教育を分けて行う期間を設ける必要がある。書き言葉では使用されない表現でも、実生活では自己表現のために必要である表現も多いからである。

6. むすび

外国人児童のための中級以降の日本語教育に利用するための文データベースを作成した。日本語形態素解析システム『茶釜』によって得られる品詞情報が付加されたデータベースを整備し、これに初級日本語文型のパターンを照合した結果、初級日本語と教科書の実文中の構文の使用状況が把握できた。

今後の課題としては、五段動詞の可能形、受け身形、使役形について、また、自動詞・他動詞についても抽出できるようにするため、茶釜の辞書記述の書き換える必要がある。

また、複文の構造について、さらに詳細に調査

する必要がある。

今後は、構築した教科書文データベースを公開し、外国人児童の日本語教育に携わる方々が活用できるようにすることを予定している。これは、編入時期や学年、日本語の習得状況や原学級の学習進度に合わせて、それぞれの外国人児童が持つ問題に対応する支援体制の一環を担うと考える。

参考文献

- [1] 総務庁行政監察局, “教育の国際化を目指して”, 大蔵省印刷局, 1997.
- [2] 外国人子女の日本語指導に関する調査研究協力者会議, “外国人子女の日本語指導に関する調査研究<最終報告書>” 東京外国語大学, 1998.
- [3] 西原鈴子, “外国人児童のための日本語教育のあり方”, 日本語学 Vol15-2, 明治書院, 1996.
- [4] 松本裕治 他, “日本語形態素解析システム『茶釜』version1.5 使用説明書”, NAIST Technical Report NAIST-IS-TR97007, Jul. 1997.
- [5] 益岡隆志, “新日本語文法選書 2 複文”, くろしお出版, 1997.
- [6] 益岡隆志, 田窪行則, “基礎日本語文法—改訂版—” くろしお出版, 1992.
- [7] 国際交流基金・財団法人日本国際教育協会, “日本語能力試験出題基準”, 凡人社, 1996.