

日本語発話資料分析の単位をめぐる問題 — 第二言語習得過程観察の立場から —

田丸淑子・吉岡 薫

国際大学

日本語プログラム

Abstract

This is a research note attempting to clarify the problems surrounding the units of analyses for spoken data of Japanese as a Second Language. Some general points of consideration in selecting the units are mentioned, followed by a survey of the units used in the literature. Then the units in our longitudinal study, S-nodes, T-units and words are discussed. A number of technical difficulties we faced are introduced and suggestive solutions are given.

1 はじめに

日本語の第二言語習得過程を観察する目的で学習者の発話または作文資料を分析しようとする時、前に立ちはだかるのは分析の単位をどうするかという問題である。本稿は、日本語資料の分析の単位をめぐる何が問題になり、先行研究ではどのような単位が用いられたのかを述べ（第二章：田丸）、さらに筆者等が学習者の文構造の長期的変化を分析する過程で遭遇した問題を報告・検討する（第三章：吉岡）ものである。第二言語習得研究の分野での発話分析では、学習者の文の構造的側面の観察や語彙発達の観察、またプラグマティックスなど社会言語学的な視点からの観察が可能であるが、筆者等の今回の分析の焦点は発話の中に現われる文構造の発達にある。そこで、本稿で扱う単位も文構造に関連するもの、その中でも「長さ」と構造の「複雑さ」を数量的にとらえることに关するものに限る。

2 分析の単位

2-1 先行研究における発話資料の分析目的

発話分析を扱った研究を目的別に分類してみる（音声関係、方言の研究は除く）。

1) 母語話者の発話資料分析

- a 話し言葉の特徴を調べる：（国立国語研究所 1955、金久保他 1993等）
- b 発話量を調べる：（江川 1980、川又 1975等）
- c 話し言葉にあらわれる文型の種類を調べる：（国立国語研究所 1960）
- d 特定の言葉や敬語の使用を調べる：（国立国語研究所 1981, 1982, 1983等）
- e 談話／会話分析（語用論を含む）：（ザトラウスキー 1993、メイナード 1993等）

f 幼児の言語習得観察：（大久保 1973、Clancy 1985等）

2) 対母語話者の発話資料分析

a 学習者の構造の習得観察：（Yoshioka, et. al. 1992、石田 1991、迫田 1993、田丸他 1993等。この分野の分析は、数量だけのもの、誤用分析と組み合わせたもの、特定の形態素の習得に限定するもの等がある。また、発話でなく作文の分析は数多い。）

b 学習者のプラグマティクスの観察：（尾崎 1993等）

3) 対非母語話者の発話資料

a フォリナー・トークの分析：（志村 1989、横山 1993等）

また、発話資料分析ではないが語彙調査：（国立国語研究所 1953、1962、1973、1983。これ等の報告は、分析単位を検討する上で参考になる。）

2-2 単位を決める際の基本

資料分析に用いる単位を決めるにあたって、検討を要する点がいくつかある。それは、

1) 妥当性と一般性

a それが分析の目的にかなった単位または単位の組合わせであるかどうか。つまり、何を調べるかによってふさわしい単位は異なる。

b その単位がある程度の一般性をもっているかどうか。これによって他の研究結果との比較が可能になり、結果がより一般性を持つ。

この二点は常に共存するとは限らない。aの為に、既存の単位が不適切であれば、新たな単位を設定したり、既存の単位を修正することが必要になるからである。

2) 日本語習得に限定せず広く第二言語習得という現象を見ようとする場合には、外国語の習得研究で用いられている単位を日本語の資料分析にも用いてその結果を比較し、より一般性のある結論を導こうということになる。その場合は、その単位が外国語の場合に測定したのと同様なものが日本語でも測定し得るか、日本語の構造をふまえて検討する必要がある。

3) これらをふまえた上で、現実的な単位であること。つまり、単位の認定作業（資料を単位に切る作業）があまりに煩雑にならないこと。得られるであろう結果の重要性和単位切りの作業にかかる労力のバランスを考慮に入れる必要がある。

2-3 文の長さに関する単位

文に関する数量的な分析の中で、文の長さは最も一般的な分析項目であり、その調査目的としてあげられる主なものは、次の二点である。

1) 書き言葉と話し言葉の比較。これまでの研究の結果、話し言葉の文は短いことが特徴とされている（江川 1980、メイナード 1993）。

2) 言語能力の発達の観察。第一、第二のいずれの言語習得でも、能力の発達にともなって文が長くなるであろうという前提で調査が行なわれている。第二言語習得では、石田（1991）、田丸他（1993）の分析で、この前提を支持する結果がでている。

ところで、「文が長い」「文が短い」と、ふつう人はこともなげに言うが、それを発話分析で数値で示そうとすると、実に煩雑な作業を経なければならない。まず、何をもって文とするのか、外枠を決めなければならない。次いで、何で計って長い・短いとするのか、

区切る単位を決めなければならない。外枠は書かれた資料では文として比較的問題なく決まるが、発話資料の場合は難題である。区切る単位は書かれた資料、発話資料とも共通する問題で、これも厄介な問題である。このどちらもいかに多くの時間とエネルギーを要するかは、分析経験者が一様に認めている。しかも、結論を出しても、当事者自身が完全に納得いくものではないことが多い。

この単位の問題は、当然のことながら、外国語資料の分析結果との比較の可能性の問題とも深く結びついている（野元 1979）

2-3-1 文に相当する単位

2-3-1-1 問題の所在

まず、発話で文に相当する単位を認定することが難しいのは、書かれた資料の「。」のように、明確に単位の終わりを示すものがないためである。一般的に用いられる判断材料は、1) 沈黙（ポーズ）、2) イントネーション、3) 意味・構文的なひとまとまりであるが、いずれも決定的という訳にはいかず、場合場合で使い分けたり、総合したりして判断することが多いようである。母語話者の発話資料の場合はある種の規範を当てはめられるが、学習者の発話資料の場合は、目標語のイントネーションを習得しているとは限らず、沈黙も意図したものか能力の限界からくるのか不明なので、さらに難しい。

次に、発話の形式、つまり独話か会話かによっても難しさが異なる。会話の場合は、発話交替が単位認定の重要な要素になる。しかし、相づちを発話交替と認めるかどうか、発話交替があっても、場合によってはひとつの「文」相当単位が続いていると考えられるかどうか等の問題が生ずる。独話でも、聴衆の反応（相づち、笑い、驚き等）を受けながらの独話と、聴衆からまったく切り離された独話とは異なる。

さらに大きな問題は、いわゆる話し言葉の構造上の特徴とされているさまざまな現象から生ずる問題である。その例として、

- a 倒置、省略、挿入、重複などがおこり、文の範囲が決めにくい。
- b 語順の倒錯、ことば不足、話線のそれ、中絶、混淆などで形も整わず意図も不明な発話があり、文と認めるかどうか決めにくい。
- c 接続語（～だけど等）の過剰使用による形の上での不完全文があり、文の範囲が決めにくい。

等があげられる（国立国語研究所 1960、金久保他 1993、メイナード 1993）。

以上の諸問題の扱いを通して問われているのは、

- a 作文の場合の文に相当する単位を発話ではどう規定するか。
- b その基本単位をどのように認定するか。
- c 不整かつ意図不明な発話をどう処理するか、である。

2-3-1-2 先行研究で用いられている単位

1) 「文」を使っている例

「談話語の実態」（国立国語研究所 1955）：文の定義は明示されていない。さらにこの中では、文の長さや構造の分析があるが、発話されたすべてのものが分析の対象となっているのか、それとも不整文として扱われて、分析されなかったものが

あるのかについてもまったく言及されていない。

「話しことばの文型I」(国立国語研究所1960):文の定義はされていないが、文、不整文ともに大量の例と不整の理由が示されている。不整文は分析されていない。
石田(1991):文の定義、不整文の扱いともに言及されていない。文はすべて分析し、文法的・非文法的に区別されたいことが資料からうかがえる。

2) 「文」以外を使っている例

志村(1989):「発話」「T-単位」。いずれも Chaudron(1988)、Hunt(1966)の定義によるとしているが、「発話」は文に近いものと考えられる。研究の性質上、不整文は出なかったと考えられる。

金久保他(1993):「発話」。ポーズとイントネーションなどの情報を考慮した上で、ある程度の意味・構文的な整合性をもつ一つのまとまり。原則的には整合性を重視(p.75)。著者によると、不整文として捨てたものはないとのことである。

田丸他(1993):「アタランス」「T-単位」。「アタランス」は1) ひとつのイントネーションのかたまりで、2) ポーズで前後を切り離されていて、3) 意味的にひとつのまとまりを構成している、のうちの少なくとも一つを満たすもの。構造的には文に近い。「T-単位」は「アタランス」中の文構造を示す単位で、付属、または埋め込まれた従属節を含む主節である。不整文として捨てたものはなく、非文法的として扱う。

参考までに、構造分析ではないが、談話／会話分析では、

メイナード(1993):「文」および、その下部単位としてPPU(Pause-bounded Phrasal Unit:ポーズによって区切られた語句)を用いる。文の定義を以下のように明示している(p.98)。

- 1) 通常すくなくとも1つの用言を含む。ただし、あきらかに用言が省略されていると考えられる場合は、用言なしでも「文」とみなす。
- 2) 日本語の場合、用言の連用形で終わる表現も文末の下降調のイントネーションを伴う場合は文末とみなす。
- 3) 従属節と考えられるものも、主節が省略されている、ないしは関連する主節が無いとみられる場合は文とする。
- 4) 倒置文の場合、倒置された部分も、述部に関連しているかぎりその文の一部とする。
- 5) 言いよどみ、又、埋め込み表現は用言を伴うか文末イントネーションを伴うかする以外は文と認めない。

以上をまとめると、「文」を用いた者は、メイナードのように明記しないまでも、かなり発話資料に対応した処置をとっていることが予想される。それに対し、「文」以外の単位を用いた者は、書かれた文との違いを意識して「文」という名称を避けたというのが実情で、両者は非常に近いものであると考えられる。

2-3-2 語またはそれ以下に相当する単位

2-3-2-1 問題の所在

文の長さ、単位あたりの異なり語数、ある語の出現度数、使用率等の分析をする為に、

どんな単位を設定したらよいか。例えば、英語資料の分析で用いる「word」に相当するものとして、日本語では何を充てればよいか。日本語の語とは何か。この問題は、発話資料の場合だけでなく書かれた資料の場合でも、日本語に関する限り避けて通れない。この問題の厄介さは、国立国語研究所の語彙調査が回によって異なった単位を採用し、それをめぐる検討が報告されていることからわかる。問題は、一言で言ってしまえば、どこがつかまっていてどこが離れるのか、独立しているものは何か、という点である。問題の所在として、土屋は、

- 1) 複合名詞や複合動詞を切るか、一単位としてとるか。
- 2) 助詞や助動詞は、それぞれ一単位としてとるか、連続としてとるか。
- 3) 接頭語や接尾語は、切り離して一単位とするか、切り離さないか。
- 4) 数字はどう扱うか。数字は語か。例えば三百十五はどう取り扱うか。
- 5) その他に書かれた資料では、記号、句読点をどう扱うか。

をあげている（土屋 1980）。

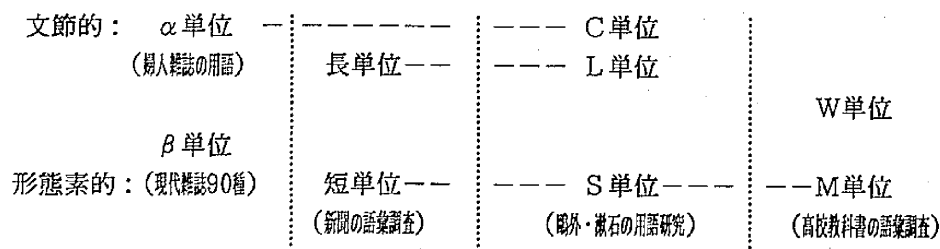
さらに、第二言語習得の研究で、特に能力の発達を観察する場合には単位として区切られた一つ一つが担う「値打ち」の問題もでてくる。例えば「本は」と「読ませられました」はどちらも一文節だが、日本語能力の発達という見地から同じ価値を持つと考えるのは妥当であろうかということである。この、教師としての心理が単位の問題を一層複雑にしている。

2-3-2-2 先行研究で用いられている単位

文の長さは、最も一般的な分析項目であるが、用いられている単位は文節が大勢を占めている。文節の場合は殆どがいわゆる学校文法に依拠しており、必ず1個の詞が含まれ、0個ないし1個以上の辞が含まれるとする。文節を単位として用いる場合は、それが妥当かどうかは別として単位の認定作業は比較的問題が少ないといえる。時に、文節と併用して音節を用いる場合もある。

日本語の発話分析としては古典的と言える国立国語研究所の「談話語の実態」（1955年）では文の長さは文節数・語数・音節数の三種の単位で計られている。他に文節を用いているものは、石田（1991）、金久保他（1993）、メイナード（1993）等がある。

発話の分析用ではないが、国立国語研究所の語彙調査等では、目的にあわせて下記のように数組の単位を認定している（国立国語研究所 1953、1962、1970、1983、齋岡 1976、1980、米田 1974）。ここでは文節レベル、形態素レベル、その中間の三レベルがある。中間レベルのとり方は、少しずつ異なるようである。その関係は下図のようだといえよう。



中間のレベルにある四単位は次のように説明されている。

長単位・C単位：助詞・助動詞およびその連続は1単位とする。つまり、助詞一語は1単位だが、連続の場合はまとまりで1単位。

W単位：文節をこえない。活用語（動詞・形容詞およびそれらの終止連体形以外の活用形にいわゆる助動詞のついたもの）は、末尾が終止連体形であるもの、命令形であるもの、および中止用法・修飾用法である連用形であるものが、1W単位となる。この自立形態の後ろにつく助辞（助詞および助動詞類）は1要素ごとに1W単位とする。

β 単位：現代語として意味を担っている最小の言語単位（一語の成分と考えられる場合の接辞・助詞・助動詞は最小単位としない。）

なお、鷗外と志賀直哉の文の長さの分析では、文体の影響を受けないという理由で、C単位が用いられている（米田1974）。

文節以外の単位を用いているのは、志村（1989）と田丸他（1993）の「語」である。志村は「語」の定義は明示していない。田丸他の「語（W）」は次章に述べるが、文節と形態素の中間をとったものであり、単位認定作業が非常に難しかった。発話分析で形態素を単位としたものはまだ目にしていないが、形態素を単位として用いると前述の問題点である接頭語・接尾語、複合名詞等の問題と直面しなければならなくなるであろう。

2-4 文の構造に関する単位

発話の文の特徴をとらえることを目的とする分析では、「長さ」とともに文の構造の分析が行なわれるのがふつうであるが、「長さ」の場合とは異なり、分析目的によってその切り口はさまざまである。例えば、先行研究の分析の目的は以下のように大別される。

- 1) 文の構造の型を分類しようとするもの
「話しことばの文型」の類
- 2) ある構文、文の要素、また形態素の現われ方をみようとするもの
主語の有無、文末表現、受け身の有無、連体修飾の構造等
- 3) 構造の複雑さをとらえようとするもの
重文・複文、文成分間の関係等

ここでは、第二言語習得の分野にかかわってくる2)と3)に限って述べる。

2-4-1 問題の所在

目的によって様々な切り口の可能な構造の分析では、単位の設定や認定の問題は個別的で共通の問題としては扱いにくい。ここで問題になるのは単位の切り方ではなく、ある単位が妥当かどうか、そしてそれを用いた分析結果が第二言語習得の上でどんな意味を持つかということであろう。

構造の分析での見方は今のところ以下の二つがある。

- 1) ある構造、形態素などを特定し、その出現数または割合を見る。
 - 2) 文の構造を「複雑さの度合い」という尺度で数量化してとらえる。
- 1) の場合は作業そのものは比較的単純であるが、分析結果の意味づけが難しい。つまり、ある項目の出現が学習者の習得上何を意味するのか、仮説に基づいておさえられていない

ればならないだろう。2) の場合は「複雑さの度合い」を表すために、どんな単位を、どのように組み合わせて計算するかが問題になる。

2-4-2 先行研究に用いられた単位

文の複雑さという点に限って先行研究を見てみると二つの分析方法があることがわかる。

- 1) 文中の述語と他の成分とのかかわり方に基づく分析(国立国語研究所 1955)。
- 2) 重文的・複文(埋め込み)的成分の有無に基づく分析(志村 1989、石田 1991、金久保他 1993、田丸他 1993)。

1) の分析は、文中で最も間接的な成分が述語に対して第何次の関係にあるかということに着目し、発話の文を「次」の多少ということで分類している。結果は全文中に何次の文がどのくらい現われているか、その割合ということで示されている。この方法は「談話語の実態」(1955)で母語話者の発話分析に用いられている。構造の簡易な短文の場合は比較的問題がないだろうが、文が長くなった場合や複雑な構造の場合、さらに学習者の誤用を含んだ文の場合には「次」を決める作業が非常に難しくなることが予想される。

2) の分析では、発話中に占める単文、重文、複文の割合を調べるというのが最も一般的な方法である。石田(1991)が学習者の中間言語の発達を分析する目的で使用している。ここから一歩進め、とくに複文に注目して、複雑さの度合いを数量的に示そうとする試みもいくつかある。まず、母語話者の発話分析では、金久保他(1993:79)が「述語相当句」という単位を設定している。これは、「発話の一部分で述語に相当する働きをする文節」と定義されており、述語相当句の総数と、「一発話平均述語相当句数(総述語相当句数を従属節を持つ発話数で割った数値)」とで、講義の日本語の特徴をとらえようとしている。

第二言語習得関連の分野で同様の分析を行なっているのが、志村(1989)と田丸他(1993)である。ここでは、複文構造が文の構造を複雑にする重要な要素であると考えて、埋め込み構造について「S句」という単位を立てている。そしてT-単位あたりの「S句」の数という形で文の複雑さを数量で表そうとしている。ここでいう「S句」は上記の「述語相当句」と非常に近いものだと考えられる。複雑さの度合いを数量化できるかという問が当然でてくと思うが、この方法の利点は、複数の資料間(例えば縦断的研究で初期の発話資料とある時間が経過した後の資料との間)での比較がわかりやすいことと、一つの文の中に複数の「S句」(述語相当句)があるような文でも、それぞれの句の価値が生きることである。

3 筆者等の場合

ここで筆者等の研究(Yoshioka, et. al.1992、田丸他 1993)で使用した分析単位についてその定義を述べ、遭遇した問題点をまとめる。筆者等のデータは学習者の発話を2年3か月間観察、記録した資料の中から、6名の被験者に絵の描写をしてもらったタスクを抜き出したものである。なお、分析は他のタスクを混じえて現在も継続中であり、ここに述べる単位は中間報告的な意味合いがあることを付け加えておく。

3-1 定義

前章と多少重なるが、分析単位のもっとも大きなものから順に述べることにする。

- 1) 「アタランス」。Crookes とRulon (1985)によるアタランスとは以下の特徴のうち少なくとも一つを有するスピーチの連続である。
 - a ひとつのイントネーションのかたまりで
 - b ポーズで前後と切り離されていて
 - c 意味的に一つの単位を構成している (Crookes とRulon 1985:9)
- 2) 「T-単位」。アタランス中の文構造を示す単位としてはT-単位を使用した。T-単位とは「付属、又は埋め込まれた従属節を含む主節」と定義される (Crookes 1990:184)。元来第一言語での作文の構文の複雑さを示す単位として考案され、単位あたりの語数を調べるための枠組みとして用いられてきていたが、これが後に第二言語の発話資料の分析にも使用されるようになったものである。T-単位を日本語の分析に初めて使用したHarringtonは日本語のT-単位を"nuclear sentence"とし、その最小のもの例としてa) 動詞の終止形、b) 形容詞の終止形、c) 名詞+だ、の三つの例を示している (Harrington 1983:52-53)。これらが連結してより複雑な文を形成するのだが、この単位の定義上、埋め込まれた従属節は考慮しないので、重文的な連結のみに焦点が当たる。即ち、1アタランスあたりのT-単位数を見ることでアタランスの構造が重文のような形で「横」へ伸びているかどうかを把握することができる。
- 3) 「S句」。文構造を複雑にする要素として従属による「重層的」な広がり、つまり埋め込みの構造があるが、この埋め込みの度合いを判断する手段として「S句」を使用した。金久保他 (1993:79) の言う「述語相当句」を1Sとした。
- 4) 「語」。最小分析単位は語であるが、この語の定義は先に述べたように先行研究でも問題が多いもので苦慮した。Yoshioka, et. al. (1992)、田丸他 (1993) では独自の判断を試みたが、現在進行中の分析では、文節である「C単位」(齋岡 1976) とより小さく形態素に近い「M単位」(齋岡 1980) を併用することにした。

3-2 問題

3-2-1 アタランスの問題

筆者等はアタランスの定義をCrookes とRulon (1985) に求めた。彼らの定義の基本になっているのは英語の母語話者の談話分析であるが、筆者等の被験者は学習者であるという点で様々な問題が生じた。上に述べた定義中の三つの特徴を順に検討するが、まず「ポーズで切り離されたかたまり」という点が問題になった。母語話者と比べると、学習者の場合はその習得段階の下のものほどポーズの長さも頻度も多くなる。またポーズの性質も母語話者の場合と必ずしも同じではない。というのは母語話者のポーズは意味の区切りで一息つくことが多いのに対し、学習者の場合は語彙を思い出そうとしたり、活用を確認したり、あるいは自信のない表現の誤用を避けようとするときの迷いなどによってもポーズをとるのであって、同じ機能を果たしているとはいえないのである。それで学習者のポーズでデータを区切ってしまうと、学習者の意図した(意味のつながった)文を分断してしまうことになる場合がある。これは習得初期に比較的高頻な問題であった。

「一つのイントネーションのかたまりとなっている」という条件は習得初期にはポーズ

ほどには問題にならなかった。というのは、初期にはそれほど長い文が現われなかったため、これはおそらく時間をかけて文を考えて組み合わせたりするより、たとえ一語文になってもとにかくわかることから発話してしまうというストラテジーを使う学習者が多いためと思われる。

ポーズやイントネーションを基準とできないとなるとデータ区分の際に「意味的に一つの単位をなしている」という条件を頼りにするしかないが、この場合は、ポーズの場合と逆に習得が進んでかなり滑らかに続けて話せるようになってきてからが問題になった。というのは、T-単位の項でも述べるように意味のねじれなどを含みながら終わらない文が増えてきたからである。

このほか技術的に検討を要したケースとして以下の事項がある。

- 1) 以下の例の、<>内のような挿入語句の取り扱い。

学習者： バングラデシュはあの、<先生知っているでしょう>、あの、インドの東です。

これはストラテジー、あるいは単にL1も含めた個人の癖と言えるものとも考えられ、特に短い挿入（「先生」など）が目立つ学習者もあった。筆者等はこの前後でアタランスを切って三つのアタランスを立てるよりは、挿入語句を挿入のあった文の最後に移して別個のアタランスとして計算した。しかし、こうすると、挿入をせずに代わりに文を付け加えた場合とアタランスの数値は同じになってしまう。挿入があってもなお文を続けていけるということは評価する手段があつてしかるべきだと考えるが、現段階では考慮中である。

- 2) 学習者の発話が終わらないうちにインタビュアーの発言が入り、それを越えて前の話の内容を学習者が引き継いで続けた場合、つまりターンの区切りを越えてアタランスを認めるかどうかについて。インタビュアーが相づちを打っているような場合はともかく、以下のように割り込んでしまう場合が問題であった。

学習者： だから国の人々の心の力と

インタビュアー： いいことを言うね。

学習者： 心の力と意識とか意識と知識も大変良く大変いいです。

筆者等は、今回の研究の焦点が対話という視点からの談話分析ではなく学習者の発話中の文構造を検討するということであるから、上記のような場合はターンを越えてアタランスを認めることにした。

- 3) 発話データの特徴である言い直し、繰り返しなどをどう取るかということも非常に大きな問題である。また、それらが非文である場合、外国語である場合、などそれぞれに頭が痛い問題となる。前に述べたように、何を調べるのかによって対応が変わってくることであるが、筆者等の場合、あくまで文構造の変化に注目するという観点から下した判断は以下のとおりである。アタランス単位で検討しているので、文末の言い直しや繰り返しは削り、学習者の最終的なチョイスのみを採用する。つまり単なる繰り返しは「わたし、わたし、わたし」、間違いの訂正やより好ましいものへの言い換えは「女の子、女の子」「坂、slope」などとする。ただし、強調のためにあえて繰

り返したことが明らかな場合は残す。文単位の繰り返しの場合は「雪はあまり降りません、全然降りません。」のように残す。なお、インタビュアーの質問の全部、あるいは一部をそのまま繰り返す場合もある。下に例を挙げるが、これは答える時間を稼ぐための戦略的なものとも考えられる。

インタビューア： あなたの国はどんな国ですか。

学習者： わたしの国はどんな国。わたしの国はバングラデシュです。

このような例は言ってみればおうむ返しのようなもので学習者の中でどこまでプロセスしたものか保証はないのだが、これは削りなかった。

- 4) 語彙に不安があって尾高イントネーションで外国語を言うことで日本語訳を求めている場合は以下のように扱った。まず例を挙げる。

学習者： わたしの国は南のAsia?

インタビューア： アジア。

学習者： アジア。

この場合の問題は学習者が自分の国の場所を聞いているのではなく、明らかにアジアという語彙を確認しているのである。つまり「わたしの」と言い始めたときと「Asia」と言い終わった時では内容が変わってしまっていることになる。即ち「わたしの国は（...です。）」「Asiaは日本語で何ですか。」と言う二つのアタランスからなっていて、初めのアタランスは終わらないままになっていると考える。

- 5) 「はい、はい、はい。」などと返答の際に繰り返す場合はそれらがポーズを間に含むかどうかによってそれぞれを独立したアタランスとするかどうか判断せざるを得なかった。

3-2-2 T-単位に関する問題

T-単位がアタランスあたりに増えることにつながる構造は多岐にわたり、それらが全て単純に習得過程を示す道標になれば話は簡単だが実際はそうではない。筆者等のデータでもT-単位数は時間を追って伸びを示すのだが、必ずしもその数値の伸び即ち母語話者の日本語により近い中間言語というわけではなかった。他の日本語発話データを分析した研究でも、T-単位が数えられなくなるなど機能しなくなるケースが見られるようだ。

まずT-単位数の伸びが母語話者の日本語に近づく方向を示すと思われる例を挙げる。

a) 動詞の「て」形による接続

学習者： でも私は勉強で忙しくてまだ手紙を出しません。

b) 接続語句による接続

学習者： 例えば油も入れるし、そして肉も入れるし、それが他の名前です。

次にT-単位数の伸びが日本語の中間言語の「発達」を示す尺度にならない例を挙げる。まずc)、d)、e)は文が「横」につながっていくそのつながり方に問題があるもので、T-単位をどこで切ったらよいか判断を迷う例である。

c) 文が終わらないもの

学習者： それから会社の近くはだいたいいいー何でも(あの)大切な店とか(あの)ところとかでも会社の周り。

d) 脈絡無く続くもの

学習者： 意識は色々な例えば(あの)社会とか政治と経済とか文化についてとか国際的な何でも何でも現在の近代の何でも色々なことについて意識がありますとかこういうふうに意識とか筆者等の私私達の状況とか自分の状況自分の国の人々の状況自分の社会の状況。

e) 文が「ねじれる」もの

学習者： それからもう一つは(あの)もっと大事なところは例えば私達は(あの)はじめて生活をそんなところで(あの)始めて始めました。

このようにだらだらと文が続いていく中にもさらに二つのタイプが見られる。第一に思い付きで次々と種々の情報を続けていくものがある。例えば上のd。これに対して、意味的には区切りのある塊で話しているながら文を切らず「ですけれども」などの接続語句によって続ける場合がある。こういった例ではT-単位数は伸びても母語話者の日本語に近づいているとはいいがたい。

f) 学習者： 私の会社は例えばでっかい会社じゃないから何かコンサルティングの会社何処でもそんなに大きい工場としては工場と比べたら人間(あの)人数的なちょっと低いですけれどもコンサルティングの会社でだいたい大きいですけれども三と四番目ですけれども。

このように日本語の上達と直接結びつかないようなT-単位の増加が見られたのでT-単位が発話の中の文の「発達度」を計る尺度としてどの程度有効なのかを考えざるを得なかった。一つ確かなことはT-単位は一文中の部分の成分を考察するものであって、T-単位の集合体としての文全体が論理的につながっているかという視野を持っていない。そこに一つの限界があると思う。

ただしこれらの問題例の場合は母語話者も含む話し言葉の特徴であり、話し言葉には「単純な並列や接続でだらだらと長く続く文」が少なくなく、このような長い文が「とかく整わない、欠陥の多い文になりがち」(大石 1990:47)であるという指摘が多い。筆者

等の被験者達は日本滞在が長くなるにつれ日本人のこのような話し方に影響されている可能性は大きい。と言うことは、単位が元来書き言葉の分析用に作られた単位だから話し言葉の分析に使うには無理があるということなのだろうか。しかし同じ話し言葉でも、筆者等のデータでは学習初期からはほぼ中級前期のインタビューデータまではT-単位は機能していたと思われるのである。即ちT-単位の数値の伸びは学習者が習得過程を目的言語に向かって移行していることを反映しているようであった。これはどういうことであろうか。

この理由は恐らくある習得段階までの被験者の発話中に現われる文の性質によると思われる。筆者等の被験者が初級から中級前期頃の文は短く、構造が単純で文中のT-単位同士のかかわりが少なかったこと、また文が教科書的で比較的きちんと終わっていたことなどが特徴的である。そのため書かれた文との違いがあまりなかったのである。ただ、筆者等の被験者の場合、習得段階という要素に加えて、その生活環境が日本でありながら英語で学習や日常のコミュニケーションができる特殊な環境であったことも大きく関係していると考えられる。つまり彼らが受ける日本語のインプットは非常に限られていて、教科書や日本語の授業のなかの日本語が大半を占めていたため、発話の中にも教科書的な文構造が出現したのであろう。しかしながら卒業してからは、日本滞在が長くなったことに加えて各被験者とも日本企業に就職し、教科書以外の日本語、それも話し言葉に接する時間が飛躍的に伸びた。当然彼らの発話中の文構造も変化して、書き言葉との距離が開いてくる。ここにT-単位数が機能しなくなる第二の原因があると考えられる。逆に言えば、正規の授業を受けずにいわゆるストリート・ラーナーとして生活の中で日本語を身に付けていく学習者の場合は教科書的なきちんと終わる文の比率が少ないだろうから、T-単位数がもっと早い段階から機能しなくなる可能性がある。

ところで筆者等は被験者が究極の目的とする日本語の母語話者の場合を、比較のために検討した。被験者と同じタスクを行なって話してもらい、同じT-単位で結果を分析した。母語話者であるから話し言葉と書き言葉の違いはスタイルの違いであって言語能力に起因するものではない。そして話し言葉と書き言葉には違いがあることは上に述べてきたとおりである。にもかかわらずT-単位の使用に問題は起こらなかった。この理由はおそらく与えられたタスクが自由会話ではなく、絵の描写だったことがあると思われる。つまり相手の言うことに反応したり、相づちを打ったり、相手の話しているときに割り込んだり、と言う即興的な発話はなく、自分の描写したいものを見て考えて発話の計画を立てて話す、ということが可能なタスクだった。そのため、いわゆる話し言葉より書き言葉に近い文が起こる可能性が高かったと言えるであろう。また、絵自体の内容によって起こりうる発話内容が規定されたということも考えられる。

ここでまとめると、T-単位を発話の分析の尺度として使用する限度は次の二点が考えられる。

- 1) 学習者の場合は言語習得段階のあるレベルまでに限られる。つまりあまり文が複雑になる以前。これはT-単位の視野が文全体まで及ばないため。
- 2) タスクの性質による。ある程度発話内容を考える時間的余裕のあるタスクの方がよい。その反対の即興的な会話であれば、日本人でもあるレベル以上の学習者でも同様にT-単位は機能しなくなる可能性がある。

3-2-3 S句の問題

本論の2-1で分析単位の一般性が必要なことについて触れた。同一言語を扱った研究間での比較、検討ができることは当然考慮されるべきだが、言語習得過程の普遍的特性を探ろうとすると他言語データとの比較が可能でなければならない。筆者等の場合、S句に関しては英語データとの比較の限界を考えることになった。これはS句を形成する表現の検討中により明らかになったことだが、以下のように英語ではS句以外で表現される表現であっても、同内容の日本語ではS句を使うものが多いのである。

そういう店	a store like that
食べられる茸	edible mushroom
林檎を買うつもり	intending to buy an apple
病院に入った後	after entering a hospital
東京に行くため	in order to go to Tokyo
青いシャツを着ている人	a man in the blue shirt
わたしが会社に入ったとき	When I entered the company

ということは、S句の起こりうる環境／条件自体が言語によって異なるということだから単純に言語間での数値のみを比較することは意味がないということになる。また田丸他(1993)で述べたようにS句＝重層構造＝文の並列より難しいはず＝並列より後から出現するだろう、というような教師の予測が外れるのも日本語のS句の守備範囲の広さから来ているのだと考えられる。S句、即ち英語の関係詞節、という訳ではないからである。

3-2-4 語の問題

上に述べたように膨大な数の語を扱った語彙研究の中で様々な語の定義がなされてきたが、筆者等は齋岡(1980)の語判断を基準にした。その理由は、筆者等として全て納得できるものだったからというより、かなり細かく品詞別・和語・熟語・数など例が出ていて参考にしやすかったことがある。というのもC単位、あるいは文節に相当する単位に関してはあまり問題がなかったが、形態素にあたるM単位は非常に判断が難しく、何らかの参考基準が必要と思われたのである。

ここで起こってきた問題はまず、齋岡の対象が書き言葉であって筆者等のような話し言葉ではないことである(かといって、上に述べたように話し言葉で語を単位にした志村(1989)にはその定義がされていない)。言うことはデータ書き起こしの際に起こる間違いの可能性、学習者の発音のエラーの記載の問題(そのままにするか正しいものを書くのか)、などは筆者等が解決しなければならないということであった。

また齋岡は母語話者の書いたものを対象としていたので文法上のエラーを考慮する必要がなかったが、学習者のエラーの処置はどうするか(「暖かいだから」など)や、同じエラーでも語彙のエラーと判断するか発音のエラーと取るかといった判断にも時間を割かなければならなかった。

さらに、齋岡の場合は目的が語彙調査であったために細かい単位が必要であったと考えられるが、文構造の発達を主眼とする筆者等としては妥当性に問題を感じる切り方もあった。例えばそのM単位で「1994」が「せんーきゅうーひゃくーきゅうーじゅうーよん(6語)」、「人々」は2語、など。

羅岡にはかなり細かく語の切り方をまとめているのだが、それでも筆者等のデータに現われた語句で判断に苦しむものが多かった。多少例を挙げ、その筆者等なりの結論を書いておく。以下スペースはC単位の、ハイフンはM単位の区切りである。また、用いられた\$などの記号は分析にコンピュータプログラムを使用することを考慮してある。

1) 動詞の語尾をどのように分けるのか。

=形態素別にする。例：あるーかーも しれーませーん

2) どこまでをサ変動詞と取って、どこから助詞の欠落した名詞+「する」と判断するか（「アルバイトする」など）。

=統一を図るため全てサ変とする。例：パーティーする

3) 文頭、文中の「だから」はそれぞれーM（単位）なのか二Mとするのか。

=文中、文頭共に二M。（「だーから」）

4) 学習者が、かたまりとして覚えていると思われる語句を分けるほうが妥当なのか、あるいは逆なのか（「女の人」など）、また、どこまでをかたまりとしてとらえているのだろうか、そもそも学習者がどうとらえているかなどということを研究者側から言えるのか。学習者自身にも明確な意識はないかもしれないし、はじめかたまりとして覚えたものが次第に分割されて部分部分が意識化されていくことはよく言われることである。

=分ける。例：女ーのー人

5) 可能形と受け身形をどのように差別化するか（これは筆者等がデータをコンピュータプログラムを使用して分析する関係で生じた問題である）。

=アルファベットを使用する。例：taber-are-ru 受け身、taberar-e-ru 可能

6) 文末の「でしょう」を短く「でしよ」と言った場合の記載方法はどうかすれば両方に誘いかけの意味を残せるか。

=アルファベットと記号を使用する。例： （でしょう）ある-desho-o、
（でしよ）ある-desh\$o-o

7) 助動詞をどう扱うか。

=例：妹ーの ようーです、妹ーの ようーな 人ーです

結局このような基本的な基準で個々に検討を重ねざるを得ないものが多く、これが研究を進めていく上で時間もエネルギーも要する作業なのだが、先行研究では紙面の都合もあるのかこういったデータの扱いについて言及してあるものが少なく、非常に残念であった。

4 まとめ

研究の基礎となる分析単位は、先行研究で使用され、試された分析単位でなければ行き当たりばったりで統一性を欠くものになるおそれがある。しかしながら、先に述べたように、研究目的が同じでない先行研究の単位を応用すると結局修正を加えなければならなくなる。筆者等の場合であれば、文の構造にあまり影響がない接頭辞や接尾辞などはあまり細かく切り方を検討する必要がなかったのかもしれない。結局、研究目的にあわせて分析の単位は変える必要がでてくるということなのであろうが、ショードロンが言うように、(Chaudron 1988: 75)分析過程を報告している研究がほとんど無いので研究間の比較が難し

い。そこで新しく研究を試みるものは独自の単位を開発する。この悪循環から逃れるのは難しそうである。

筆者等の遭遇した諸問題はそれぞれ何らかの結論を出さなければならず、その度にかかりの時間を費やして研究者間で議論をした。これらの様々な問題は事前に予測できそうではなかったものであり、だからこそ本論が同様の研究を試みる研究者の一助になればと期待するものである。他の同様の研究でもおそらくこのような分析過程で多大の労力を費やしていることと思われるが、それが何らかの形で発表されることを望みたい。

（筆者等の共同研究者として木村静子、池田裕両氏が以上の分析単位に関する諸問題に取り組んできていることを申し添えておく。）

参考文献

- Chaudron, Craig (1988) Second Language Classrooms: Research on teaching and learning. Cambridge University Press.
- Clancy, Patricia (1985) "The acquisition of Japanese," In Slobin, D. I. (ed.), Crosslinguistic study of language acquisition: Volume 1: The data. pp. 373-524
- Crookes, Graham and K. Rulon (1985) "Incorporation of corrective feedback in native speaker/non-native speaker conversation," Technical Report No.3. Center for Second Language Classroom Research, Social Science Research Institute. University of Hawai'i.
- Crookes, Graham (1990) "The utterance, and other basic units for second language discourse analysis," Applied Linguistics 11, pp. 183-199
- 江川 清 (1980) 「現代人の話し言葉」『「ことば」シリーズ12 話し言葉』文化庁 pp. 52-62
- Harrington, Michael (1986) "The T-unit as a measure of JSL oral proficiency," Bulletin of the ICU Summer Institute in Linguistics 19. pp. 49-56
- 石田敏子 (1991) 「フランス語話者の日本語習得過程」『日本語教育』75号 pp. 64-77
- 金久保紀子他 (1993) 講義の日本語における理科系・文科系の特徴『日本語教育』80号 pp. 74-90
- 川又瑠璃子 (1975) 「一日の言語生活 —主婦のことばの一日—」『新・日本語講座 5 日本人の言語生活』汐文社 pp. 1-21
- 国立国語研究所 (1953) 『現代語の語彙調査 婦人雑誌の用語』秀英出版
- 国立国語研究所 (1955) 『談話語の実態』秀英出版
- 国立国語研究所 (1960) 『話しことばの文型I』秀英出版
- 国立国語研究所 (1962) 『現代雑誌九十種の用語用字 第一分冊』秀英出版
- 国立国語研究所 (1970) 『電子計算機による新聞の語彙調査 (II)』秀英出版
- 国立国語研究所 (1981) 『大都市の言語生活 分析編』三省堂
- 国立国語研究所 (1982) 『企業の中の敬語』三省堂
- 国立国語研究所 (1983) 『敬語と敬語意識』三省堂
- 国立国語研究所 (1983) 『高校教科書の語彙調査』秀英出版
- メイト・K・泉子 (1993) 『会話分析』くろしお出版
- 野元菊雄 (1979) 「日本語の話しことばと書きことば」『「ことば」シリーズ1 日本語の特色』文化庁
- 尾崎明人 (1993) 「接触場面の訂正ストラテジー —「聞き返し」の発話交換をめぐって—」『日本語教育』81号 pp. 19-30
- 大石初太郎 (1980) 「話し言葉とは何か」『「ことば」シリーズ12 話し言葉』文化庁 pp. 36-51

- 大久保愛 (1973) 『国立国語研究所報告 50 幼児の文構造の発達 - 3歳～6歳児の場合-』
- 迫田久美子 (1993) 「話し言葉におけるコ・ソ・アの中間言語研究」『日本語教育』81号 pp. 67-80
- 志村明彦 (1989) 「日本語の Foreigner Talk と日本語教育」『日本語教育』68号 pp. 204-215
- 田丸淑子他 (1993) 「学習者の発話に見られる文構造の長期的観察」『日本語教育』81号 pp. 43-54
- 土屋信一 (1980) 「言語情報処理における語の把握」『国立国語研究所報告 67 電子計算機による国語研究X』 pp. 52-64
- 齋岡昭夫 (1976) 「国語研究のための索引作成システム - 新システム基本思想を中心として-」『国立国語研究所報告 59 電子計算機による国語研究VIII』 pp. 1-17
- 齋岡昭夫 (1980) 「高校教科書の用語調査の言語単位について」『国立国語研究所報告 67 電子計算機による国語研究X』 pp. 20-51
- 横山杉子 (1993) 「日本語における、『日本人の日本人に対する断り』と『日本人のアメリカ人に対する断り』の比較 - 社会言語学のレベルのフォリナートーカー」『日本語教育』81号 pp. 141-151
- 米田正人 (1974) 「"文の長さ"の統計学的一考察 - 一時系列的解釈-」『国立国語研究所報告 54 電子計算機による国語研究VII』 pp. 130-139
- Yoshioka, Kaoru, et. al. (1992) " Longitudinal development of Japanese sentence structures: A case of oral production by six learners," In Hayes Tom and Kaoru Yoshioka (eds.), The Language Programs of the International University of Japan Working Papers 3. pp. 11-25
- ザトラウスキー・ポリー (1993) 『日本語の談話の構造分析 - 勧誘のストラテジーの考察-』くろしお出版