Language Testing に関する覚書

北 市 陽 一

一 言語と言語の学習

我々の学習対象である言語には，expression の面と，content の面とがあると考えられる。この両面ともその言語に個有の体系があり，両者の相互関係も，それぞれの言語によって異なる。[1] 前者の expression の面には，話された時に直接観察出来る音で代表され，示差的特徴による音素体系がある。一方，後者の content の面には，文化の内容を背景として，文法組織（形態素論及び統語論を含む）及び語いが含まれる。我々が何か伝達する時には，両面が常に相関関係にあると考えられ，そして，それが一連の習慣となっている。

母国語の習得にあたっては，これら一連の習慣が自動的に働らき，content さえ与えられば，expression が無意識に出てくる。[2] 併し，外国語の場合は，その言語に特有の習慣を自動的に運用出来ように，意識的に練習しなくてはならない。すでに習得した母国語に於ける習慣が新しい習慣を習得する時に，若し両者が同じであれば facilitation が起り，異なる場合には interference が起こる。それで最も理想的な外国語学習の目標は，interference

(1) この定義は普遍的なものではないが，こゝでは Charles F. Hockett, *A Manual of Phonology*. (Baltimore, 1955) pp. 2—18 の考えに従つた。

(2) John B. Carrol, "Language Development in Children" in *Psycholinguistics: A Book of Readings*, ed by Sol Socporta and Jarvis R. Bastian. (New York, 1961) pp. 331—345.

をなくし学習者が bilingualism の状態になることである。[1]

言語の習得が上に述べた習慣を形成することであるが，その習慣を運用面での技能で分けると，その言語の文化の理解を背景として，hearing, speaking, reading, writing の 4 つに普通分類されている。これらの技能の習得にあたっては，古くから種々の方法が考案され，又実施されて来ている。併しながら，組織的に外国語を教える場合には，言語学の方法にもとづいて，母国語の習慣と対象となる外国語の習慣とを対照分析して，その相異点を見出し，次に，それを基礎にして教材を編集することは言をまたない。[2] 又，実施教授面に於ては，学習心理学の理論及び実験結果の力を借りて，学習者の学習過程を外国語学習に取りいれる事も又大切なことである。

二　Language Testing

語学テストの目的とするところは，言うまでもなく，(1) 学習者の適性を検査する (aptitude tests とか prognostic tests いわれるもの) (2) 学習者が指導目標に対して，どの程度習得したかを検査する (achievement tests) そして (3) 学習者の一般的な技能をそれまでの学習内容如何にかかわらず，検査する (proficiency tests) がある。(1) の aptitude tests は我が国ではあまりやられては居ないが，(2) の achievement tests は所謂「中三テスト」として行われ，又教室で行われているものもこの範疇に入れられよう。(3) の proficiency test は留学試験がこれに該当する。大学の入学試験もこれに属するが，多少とも achievement tests の性質を帯びている。

テストの作成にあたっては，学習の際に考慮に入れた母国語と，対象とな

(1) W. F. Leopold, "A Child's Learning of Two Languages," in *Report of the Fifth Annual Round Table meeting on Linguistics and Language Teaching*, ed by Hug J. Mueller. (Washington, D. C., 1954) pp. 19—30; Nelson Book, *Language and Language Learning*. (New York, 1960) pp. 39—41, p. 203.

(2) 特に Robert Lado, *Linguistcs across Cultwes: Applied Linguistcs for Language Teaches*. (*Aun Arlor*, 1958).

る外国語の対照分析の結果が利用されねばならない。以下でテスト作成の際に基準となるものを考えて見よう。[1]

(1) 目的を明確にする。テスト出来うる分野は下図にある様な12が考えられる。[2]

Skill / Aspect		Understanding			
		Hearing (Recognition)	Speaknig (Production)	Reading (Recognition)	Writing (Productoin)
Expression					
Content	Grammatical Structure				
	Vocabulary				

実施にあたって，上の12の分野を個別にテストする必要のないことは当然である。作成された問題の中に，2つ以上の分野にわたる事は大いにありうるが，目的としては1つの点にしぼるべきである。又，テストを行う時期は学習期間とも関連してくる。

(2) 妥当性。目的が明確に決定されたら，それにテストの内容が妥当であるかどうかを基準として考慮するべきである。hearingのテストとして作成された問題が，他の技能をテストする結果になれば，その問題の妥当性は失なわれる。更に，能力の違う者から同じ様な結果が生れたりすることもありうる。

(3) 客観性。問題が複雑であったり，又その結果が出題者の目的とかなり

(1) この章と次の章にわたつては，Robert Lado の著作を参照。特に *Language Testing*; *The Construction and Use of Foreign Language Tests.* (London, 1961): Robert M. W. Travers, "The Construction of Paper-and-Pencil Tests and the Assignment of Grades," in Edncational Measurement. (New York, 1955) pp. 175—206.

(2) John B. Carrol, "Fundamental Considerations in Testing for English Language Proficiency of Foreign Stndents," in *Testing: the English Proficiency of Foreign Students.* (Washington, D. C., 1961) p. 3 にある図を基礎とした。

離れたりすれば，採点が客観的に行なわれなくなる。この基準は，数人で多数の同一問題の採点する場合には，欠くことの出当ないものである。更にその採点があまり複雑であっても客観性が失なわれる。

(4) 信頼性。単に同じ分野に多くの問題を出すことが，客観的で且能率が良いとはいえない。50題の採点と100題の採点との結果が同様であれば，少ない問題でも，充分に評価出来るわけである。又，試験を受ける側から見ても，長時間を要すると，自己の能力を全般にわたって，充分発揮出来ない結果となる。

(5) 施行可能であること。特に音を取扱う場合，肉声にして，機械を用いるにしろ，忠実度が高くなくては，効果をあげえない。

(6) 利用度の高いこと。妥当性や信頼性等，すでに挙げた事項に関連するが，結果が指導上その他の方面で，充分利用出来るものであれば望ましい。

三　従来のテスト方法

前章で述べたような基準を念頭に於て，これまで行なわれてきた方法を検討してみよう。

(1) 英文和訳

英文和訳は英語学習の直接目的とは離れた非常に高度な技術と考えられる。英語をたくみに話す者は実際には日本語を英語になさず，英語で考えて英語で話しをしている。この事から，英文和訳の技能に欠けたものが，英語の4つの技能に欠けているとは言いえない。

併しながら，訳する技能がテストさるべき場合，例えば，通訳や飜訳に従事する場合ならば，これはその目的にかなうテストと言いうる。対象となる英語には色々の文体，小説，論文，新聞記事，劇等があり，テストの目的により選択が行なわれなければならない。それも，数個の文章を全部訳させる時には，採点の問題が加入してくるし，若し，かりに All-or-nothing の方法

で採点しても，長時間を要するし，1箇所の間違いの為に全部が間違いとされ，客観性はあるにしても，妥当性や信頼性に欠ける。これを匡正するために，問題点となる文構造及び語いの部分だけを空所にして置き，他の部分は解答欄にあらかじめ与えておくことで，採点の客観性を増すことが出来る。併しながら，英文に対する日本語での表現能力を，この方法では充分評価することは出来ない。

英文和訳は，作成が比較的容易であるから簡単に実施出来る。もし色々な文体の文章を出題したら他のテストに対する参考資料としては，主観性が入るにしても価値はあろう。

(2) 和文英訳

前述の方法と同様な問題を含んではいるが，他に，第一に，弁えられた日本語の理解が問題となる。そうなると，問題点は英語になく，日本語の理解に移ってしまう。次に解答として書かれた英語には，文法組織，語いの問題の他に，スペリング，句読点の問題が同時に検討されねばならなくなり，結果としては増々，採点を複雑化する。

(3) 自由作文

自由作文には作文内容及び文体が主題により異なり又，個人の思考能力や日本語に於ける作文能力にも関連をもっている。これを採点評価することは，なかなかの難事である。この点を除いても，個人が使用する文法構造や語いに相異があり，客観性にとぼしい。又使用する文法組織や語いも，学習の際に，問題となったようなものを避けることになる傾向が生じ，妥当性や信頼性が失われる。簡単な作文力をテストする方法としては，絵等を提示して，解答に，試験者が文法構造なり，語いの一部を与えて，受験者に残りの部分を書かせる方法は考えられる。併し，これとても，充分な表現力を見ることは出来ない。

(4) 客観テストといわれる空所補充，誤文訂正等の問題がこの範疇に入れられ，近年かなり多く利用されている。客観テストは，与えられ選択肢の内から正解を選ぶ recognition test と自分から作る recall test（これは解答全体を自分から作るのではないから partial production test とも呼ばれている）とに分けられる。[1]

両者とも客観性はあるが，recognition test では，二者択一の場合に，あて推量が出来る可能性があり，選択肢を多くしなくてはならぬ。この為には問題の作成がかなり難かしいし，充分に検討しないと，不自然な解答肢が出来て，実際に選択が行なわれなかったり，又解答肢間に明瞭な差別がなくなる危険がある。[2] 又試験を受ける側の表現力は直接測定出来ない。一方，recall test では，当て推量での解答は出来にくく，選択肢を作る手数は要しない。併し，二つ以上の正解が出来得ることもある点，注意さるべきである。又やゝもすると，機械的な記憶を要求する危険がある。両者とも一長一短ではあるが，客観性があることから，問題の抽出する際に，問題点となるものを充分研究すれば，その他の基準に合致したものを作成しうる。

(5) 書き取り

上述のテストに於ては，実際に音を聞くテストは行なわれない。その面をテストするには適切なものとして，屢々用いられて来ている。併し書き取りは何を目的としているのであろうか。単に耳で聞いた英語を書くのであれば，スペリングと句読点のテストにしかならず，単に英語を文字に移すことが，その英文の理解とはならない。又，逆に，一語一語を文字に移すことが出来なくても，その英文の概略を理解していることがありうる。特に，文章

(1) Recognition と recall method の違いについては B. R. Bugelski, *The Psychology of Learning.* (Buffalo, 1956) pp. 317—318.

(2) Scholastic Aptitude Test に用いられた客観テストの批評としては，Banesh Hoffmann, The Tyranny of Testing. (New York, 1962) pp. 223 がある。しかし，本書は外国語のテストとは，直接にはつながらない。

が長い場合，速記の速度テストと同類になってしまう恐れがあり，外国語学習の目的以外である。

その目的が aural perception であっても，語いは文内容から類推が出来うるし，語いのある部分，特に最後の部分が聴えなくとも，前後関係から判断出来うるので，語いのテストとしては不十分である。[1] 又，間違いが聴いた時のものか，又スペリングのものか判断しにくい。更に，多くの場合，英文が 3 回読まれ，その内 1 回は普通の速度よりゆっくり読まれる。これ結果，自然な速度では weak form で読まれたものが，遅い速度で読まれる為に，strong form になったりすることから，自然な英語のリズムが失なわれることになる。語順とか語尾変化といった文法組織の問題点は，既に英文で与えられているので，それらの点をテストすることにはならない。この形式のテストは今まで述べたテストと違い，試験官が読む時間を制御することが出来，短い文であれば，教室内の aural perception の練習には利用価値があろう。

その利点としては，chorus reading や language laboratory practice の様に，クラス全体を参加させられうるし，又，学習者の注意を集中させられうる。次に，教師が問題作成に手数，時間をあまり要しない。勿論，採点にはかなりの時間を要するが，自由作文に要する程のことはない。採点の結果から普段気付かなかった問題点を発見することも出来る。更に，学習者の側から見れば，スペリングや句読点の外に自分は何処を聞き落したか知ることが出来る。[2] 最後に，耳の訓練といっても，language laboratory を用いる必要もない。

---

(1) 例えば *He comes* に於ける，*-s* が聞えない場合とか，*two books* に於ける *two* を仮に，聞き落した場合とかである。又，voiceless sibilant は機械を通ずると，かなり，聞きとりにくい。

(2) J. Sawyer and S. Silver, "Dictation in Language Learning," reprinted in *Selected Articles from Language Learning No.* 2. (Ann Arbor, 1963) pp. 249—258.

四　望ましいテスト方法

さて，書き取りに含まれる問題を個別に取扱うことにより，そのテストの客観性を増し，採点も容易になり，その他の基準をも満足出来うる。

(一) 音とそれを表わすスペリングとの関係をテストする際に，まず考慮すべきは，音を聞き分けることは hearing の問題であり，後者は writing の問題であることである。[1)]

後者を主として，前者を従と考えるなら，語いを耳で聞かせて，スペリングを書き取らせる item dictation を行なえば，評価出来うる。[2)] その content の理解度を含めるには，内容を与え，更に語いの一部，例えば，最初の文字を与えて，その語いを完成させるか。又は，その様なものを 3 個与えて，選択させ，そして完成させる方法が考えられる。これは partial production test であるが，問題の語いを答とするような疑問文を与えることも出来る。又例えば，絵を見せて，書かせるならば，full production test となろう。しかし，これはごく限られた語いにしか利用出来ない。

単に，recognition をテストするには，multiple choice 方法を用いることで評価出来る。併し，選択肢の語いが，問題と考えている語いより，むづかしい時には，選択肢の語いのテストになる危険がある。

次に，句読点に関しては，hearing の面からテストすることは良策とは言えない。なぜなら，音調や語順がすでに与えられているからである。それよりも，recognition test 又は recall test で，必要な句読点を選択させるか，又は句読点を附けさせることで評価出来る。

---

(1) phoneme と graphic representation の関係については Robert A. Hall, Jr., *Sound and Spelleig in English.* (Philadelphia, 1961) pp. 16—26.

(2) production の語いと recognition の語いとの差，必要な語いの範囲については，Charles C. Fries & A. Aileen Trauer, *English Word Lists* 英語教育シリーズ (Tokyo, 1958)

(二) Expression の面，即ち，Sound, stress, terminal juncture, intonation を hearing と speaking の側からテストするには，それぞれ，別個のテストが考察されることにより，本当の意味の評価が得られよう。[1] 併し，実際にテストを行う場合，Expression の面と相関している content の面を完全に無視出来ないことが起こる。sound の弁別を hearing の側からテストする場合には，内容で判断したり，又は語いのスペリングや語順等で類推が出来ないようには出来る。例えば，解答欄の選択肢には，数字等を用いれば，評価されうる。[2] 併し，terminal juncture, intonation の弁別を行なわせるテストでは，音声学のテストでもなければ，content の面が現われてくるし，無理に content の面を出さないようにすれば，不自然なものなる恐れがある。更に，stress は学者より，3 個と見る人と，4 個と見る人がおり，日本の現状では，2 個の強弱しか用いられていない。実際のテストに，4 個の stress 弁別をテストすることは実行困難である。又 intonation の場合でも，個人差が多少あり，emotion の加入をどの程度にするかが問題となる。これらの点を統一すれば，同数のシラブルの 2 個又は 3 個の文又は語群を聞かせて，同異をテストして評価しうる。この際に，無暗に色々な記号を用いると，問題を更に複雑化することになるので sound のテストに数字を用いたならば，他の面も数字に統一すべきであろう。

一方 speaking の側のテストをするには，interview で実際に話させることが理想ではあるが，前にあげた基準を満足するような形式では，しかるべき機械でも発見されない限り，不可能の状態である。併し，pencil-and-paper 方式でこのように話すと“考える”ということからはテスト出来る。これが partial production test である。

sound では発音記号の知識のないものには，問題点とする発音を含む語を

---

(1) 日本語と英語の音の phoneme, allophone の比較については，高本捨三郎“日本人の英語発音の困難点”(英語教育　1960 年 5 月—8 月)

(2) Lado の作成した「日本学生のための聴覚能力のテスト」が良い例である。

選択肢から選択させるのが，一番便利であろう。空所をスペリングで補充するものは応々にしてスペリングをテストする結果になる。音とそれを表わす文字に，規則性がある場合は問題ないとしても，不規則な関係にあるものでは，本当の sound のテストにはならない。それを匡正するものとして，複雑化という欠点はあるが，発音を数字で代表わして行うことが出来る。stress, terminal juncture, intonation のテストでは，要求するような内容をもったものを条件として，それぞれをテストすることが出来る。併し，*Do you…?* といった明らかなものはさけるべきである。

㈢　次に文法構造をテストする問題に移ろう。対象となる文法構造は，発音の際は上述の諸要素と密接に関係するが，これでは言語学上の syntax 及び morphology に関連する問題と考える。各々の言語とも複雑な構造から成って居り，それを比較対照分析することは音構造の分析より困難である。(1) こゝではその問題には触れない。

hearing の側からは，文構造が表現しようとする内容の理解をテストする場合と，文構造内の特定の部分の機能の理解をテストする場合がある。耳で聞かせて，選択肢から正しいものを選ばせることが，最も普通に行われている。又 recognition だけであれば，与えられた文に対して選択肢の中から，その内容に合致するのを選ばせられうる。一方文の代りに，絵が用いられるが，絵の示めす内容がテストを受ける者の風俗，習慣の違いから，内容が正確に把握されない危険がある。例えば，絵の中の人物が表わす動作が，期待するのとは違った内容に理解されることがある。これは cultural items に属するもので，我々が問題としているテストの範囲外に属する。特に，複雑な絵であっては，文構造のテストには望ましくはない。更に，和訳を選択肢

(1) 例えば Kleinjars, Everett, "A Comparison of Japanese and English Object Structures," reprinted in *Selected Articles from Language Learning, No. 2* (Michigan, 1963), pp. 192—197.

としてその内から与えられた問題の英語の内容に合致するものを選ばせることは，訳し方のテストになるが，和訳をたゞその文の概略をつたえるものであれば，文構造のテストにはなるが，この方法がそのテストの中核となるべきではない。

speaking の側からは，sound の場合と同じ様に partial production test 方式で，空所を補充させる完成法，語順配列法，更に英語の問に対して解答を書かせる方法が考えられる。この場合は，期待する文構造で答えない場合が起りうる。同様のことが，日本語で内容を与え，それに相当する文を英語で作らせる場合も起こりうる。

(四) 以上，諸要素を個別にテストする方法を考えてきたが，我々が言語を用いるときは，その内容の把握に中心をおく。production の側からは，interview 方式を取ることが考えられるが，その評価には主観性が入ることになり，他のテストの補助的な参考資料として，取扱うことにすべきであろう。recognition の側からは auditory comprehension test とか aural comprehension test とか呼ばれうるものである。[1] 前述のテストでは1っの文が中心となって作成されて居るが，こゝでは数個の文の内容の理解度を見ることである。問題点の選択に困難がともなうので，production の場合と同様，他のテストの補助的な参考資料となるものである。形式として考えられるのは，短いパラグラフ，さらに，相当長いものも考えられる。文の内容及びスタイルは多様であり，又，取上げる問題点もかなり広範囲にあたるため，作成はかなり複雑であり，困難である。問題として提示するパラグラフも単に1っではなく，短いものに長いもの，具体的な内容のもの，それに会話体・論文体と，少なくとも2っ以上考えるべきであろう。又，短いパラグ

(1) Lado の言う auditory or aural conprehension は，諸要素に関して，受験者に，どの要素のテストかを意識させないで行う。それ故に，各要素を個別ではなく，色々とまぜてやることのようである。Lado, *Op cit* pp. 204—208.

ラフと長いパラグラフとでは，内容理解も，個人の記憶力の差により異なってくると思われる。長いものは，単に英語の技能の習得以外ものも評価の際に考慮しなくてはならない。一方,解答欄には概略を書かせるものは,評価の客観性を低くするので，選択肢から選ばせる方法が一般的と考えられる。たゞその中に，地理上や歴史上の知識を問うようなものは避けるべきである。さもないも，comprehensionの度合に関係なく，解答されうることになる。このテストで知ろうとするのは，全体的な内容の把握の度合であるから，その標題とか，内容の大要とかが適当と考える。耳で聞かせず，文章を与えて，内容の把握を検べる reading comprehension も考えられる。auditory comprehension と異なる点は，その方法の違いはさることながら，対象なる文章の内容，文構造及び語いは，reading の場合には，一層，むづかしくなることは勿論である。

四　結　　び

以上，paper-and-pencil 方式による，言語の両面の技能をテストする方法を考えて来た。学習者の外国語の綜合的な技能習得の度合をしらべるためには，個々の技能を個別にテストしたものを主として，内容理解の度合を従としたのは，最初に述べた基準を中心として考察したからである。テストの目的が，純粋に技能習得以外にあれば，比重も変ってこよう。又，normsとか standard deviation 等の educational measurement については，テストと密接に関係するが，本稿ではふれなかった。

外国語教授及び学習に於て，対象言語と母国語の比較対照による問題点の摘出の重要性が唱えられ，その方面の研究の結果がまたれる。その様な問題点を教え又学ぶことが外国語習得であり，又それらをテストすることが，本来のテストの姿であろう。

List of Books and Articles Referred

Brook, Nelson, *Language and Language Learning.* (New York, 1962)

Bugelski, B. R., *The Psychology of Learning.* (Buffalo, 1956)

Carol, John B., "Fundamental Considerations in Testing for English Language Proficiency of Foreign Students," in *Testing: the English Proficiency of Foreign Students.* (Washington, D. C., 1961)

Carol, John B., "Language Development in Children," in *Psycholinguistics: A Book of Readings,* ed by Sol Saporta and Jarvis R. Bastian. (New York, 1961)

Fries, Charles C. and A. Aileen Traver, *English Word Lists.*

Hall, Robert A. Jr., *Sound and Spelling in English.* (Philadelhhia, 1961)

Hocket, Charles F., *A Manual of Phonology.* (Baltimore, 1955)

Hoffman, Banesh, *The Tyranny of Testing.* (New York, 1962)

Lado, Robert, *Linguistics across Cultures: Applied Linguistics for Language Teachers.* (Ann Arbor, 1958)

_____ Aural Perception Test for Japanese Students. (英語教育シリーズ)

_____ *Language Testing: The Construction and Use of Foreign Language Tests.* (London, 1961)

Leopold, W. F., "A Child's Learning of Two Languages," in *Report of the Fifth Annual Round Table Meeting on Linguistics and Language Teaching,* ed by Hug J. Mueller. (Washington, D. C., 1954)

Kleinjans, Everett, "A Comparison of Japanese and English Object Structures," reprinted in *Selected Articles from Language Learning, No. 2.*

Sawyer, J and S. Silver, "Dictation in Language Learning," reprinted in *Selected Articles from Language Learning,* No. 2. (Michigan, 1963)

高本捨三郎："日本人の英語発音の困難点"（英語教育　1960年5月—8月）

Travers, Robert M. W., *Educational Measurement.* (New York, 1955)