日本語版62項目の Token Test の妥当性に関する 量的および質的研究

平 口 真 理 鳥 居 方 策
榎 戸 秀 昭 相野田 紀 子*

Clinical Validation of an Auditory Language Comprehension Test for Aphasia: a Quantitative and Qualitative Study on the Token Test

Mari Hiraguchi, Hosaku Torii, Hideaki Enokido and Noriko Ainoda*

A full length version of the Token Test (Japanese edition) and WAIS were administered to 32 aphasics, 23 non-aphasic brain-damaged patients and 20 controls. The diagnostic aphasia test (SLTA or Rohken's) was also administered to aphasic patients. The results are as follows :

1) The Token Test could discriminate satisfactorily between aphasic and non-aphasic brain-damaged patients. On the cut-off score of 48 (lowest score shown by the control group) in the pass-fail scoring, 89.1% of all the patients were correctly classified while 6.2% of the aphasic patients and 17.4% of the non-aphasic patients were classified incorrectly.

2) In the control group the influence of age and IQ on the Token Test scores was investigated. No significant correlation was found. On the other hand, in the aphasic and the non-aphasic groups some significant correlations were found : in the non-aphasic group a moderate influence of VIQ (at the time of examination) on the Token Test scores was found ; in the aphasic group a slight influence of age (negative correlation), and moderate influence of VIQ and PIQ (at the time of examination) on the Token Test scores were found.

3) In Parts (I-V), semantic errors were examined concerning shape, color, and size. In the aphasic group the highest error rate was shown with shape, the second with color and the least error rate with size. In the other two groups, those error rates were less than those concerning size made by the aphasic group. Syntactic errors were committed more frequently than semantic errors in Part V. The aphasic group had higher error rates than the other groups. But, the patterns of difficulty the aphasic group felt were similar to those the other two groups felt trying to respond correctly to each item of the Test.

4) In the aphasic group significant correlations were found between the Token Test scores and the scores of both auditory comprehension ability and speaking ability assessed by the differential diagnostic aphasia test.

The Token Test seemed to have about the same discriminating power in an Italian, an English, a German and a Dutch speaking group of neurological patients. And, it was

金沢医科大学神経精神医学教室

* 金沢医科大学耳鼻咽喉科学教室

demonstrated that there was a close relationship between the severity of aphasia and the Token Test score. The implications of these results were also discussed.

I. はじめに

De Renzi と Vignolo³⁾ が鋭敏な口頭言語理解障害の検査として考案した Token Test は、発表以来20年にも満たないわずかな期間に、多くの国語に訳され、世界各国でその臨床的価値が認められてきた。しかし、psychometric test として完成するには、厳密な意味での妥当性や信頼性の研究、そして標準化の作業が残されている。

とくに日本では、Token Test の原版に関する研究はほとんど行われないうで、実用的な簡便さのみが追求された結果、短縮版の Token Test が安易に使用され、広く普及してしまった。しかし、Token Test が鋭敏で手軽な道具であり、今後さらに利用が広がるとすれば、その基礎的な研究は不可欠であり、日本でも Token Test 研究はもっと多くなされなければならないと考える。

われわれは数年来、原版をもとにした Token Test を実用するかたわら、本テストに関する基礎的な研究を手がけてきたので、その結果を報告し若干の検討を加えてみたい。

II. Token Test に関する知見の概要

イタリアの De Renzi と Vignolo³⁾ が、成人失語症患者の口頭言語理解の軽度な障害を発見するテストとして、Token Test を考案したのは1962年である。彼らは、従来の多数の口頭言語の理解力テストが、年齢、性、教育年数、知的機能などに左右されることを指摘し、これらの要因から独立して、純粋に言語理解の能力を測定するテストとして、この Token Test を新しくデザインしたのである。このテストを構成する口頭命令には、冗長な表現はまったく含まれず、言語学的に必要な最少限の要素からなり、かつ、日常の言葉と日常のいいまわしのみが用いられている。また、テスト材料として、具体的な物品（例えば鉛筆や歯ブラシなど）ではなく、token（ゲームに使う代用貨幣）を用いることにした。その結果、形をいい表す言葉（名詞の「丸」と「長方形」）が、他の要素である色と大きさを表す言葉（形容詞）と同等に扱われることになり、患者がおかす言語学的な誤りの質的検討をも可能にした。

De Renzi らの原版³⁾ は、具体的な大きさの記載はないが、大小の丸と大小の長方形の4系列に、それぞれ赤、青、緑、黄、白の5色で色分けした token 計20個を用いる。テストは5つの part からなり、part を追うごとに難しくなる。Part 1～4までは、動詞と目的語

からなる構文が口頭命令として用いられる。Part 1は、大きい丸と大きい長方形の2系列を用いて、「赤の丸を取りなさい」という形式の口頭命令を与える。Part 2は、さらに小さい丸と小さい長方形を加えた4系列を提示し、「小さい白の四角を取りなさい」などの項目を与える。Part 3は、大きい2系列を提示し、「赤の丸と緑の四角を取りなさい」という形式である。Part 4は、再びすべての系列を提示して、「大きい白の丸と小さい緑の四角を取りなさい」などである。最後の part 5では、大きい2系列を用いて、前置詞、接続詞、副詞などの文法要素を取り入れた、より複雑な構文を口頭命令として用いる。Part 1～4はそれぞれ10項目ずつ、part 5は21項目で計61項目である（後に part 5には項目番号11が加えられ計62項目となる）。

その後、Token Test の妥当性研究が各国で行われた。イタリアでは、Boller と Vignolo¹⁾、ドイツでは Orgass と Poeck¹⁴⁾、アメリカでは Swisher と Sarno²¹⁾、によって報告され、より最近では、オランダの Van Dongen と Van Harskamp²²⁾、ドイツの Hartje ら⁵⁾、および Cohen ら²³⁾ が報告している。これらの報告によれば、採点方法の違いはあるが、失語症を他群から区別する弁別力は84～91%と高率であり、いずれも Token Test の妥当性を支持している。さらに、従来からよく使われている Marie の「3枚の紙」試験よりも鋭敏であることが示された (Orgass と Poeck¹⁴⁾、Van Dongen と Van Harskamp²⁵⁾)。

一方、De Renzi と Vignolo³⁾ の原版を短縮して用いようとする研究も多い。Spellacy と Spreen¹⁹⁾ は、原版を約60%に短縮した39項目版と、それをさらに短縮した16項目版を作成し、いずれも信頼性と妥当性において満足な結果を得たと報告している。彼らはまた、2つの採点システムを提案している（後述）。その後の短縮版の研究は、Van Harskamp と Van Dongen²⁵⁾ が16項目と10項目、De Renzi と Faglioni⁴⁾ が36項目について報告しているが、いずれもその有効性ととみに限界について述べている。

以上の研究では同時に、非失語症者においてテスト得点に影響を及ぼすと考えられる年齢、知能、教育年数、性別などの要因についての分析が行われている。本テストの得点に対する年齢の影響 (Orgass と Poeck¹⁴⁾、Hartje ら⁵⁾)、知能の影響 (Van Dongen と Van Harskamp²⁵⁾、Hartje ら⁵⁾)、教育年数との関係 (Boller と Vignolo¹⁾、Orgass と Poeck¹³⁾、De Renzi と Faglioni⁴⁾) につい

て、それぞれいくつかの指摘があるが、一致した見解は得られていない。性差については、有意な差を報告しているものはない。

失語症のタイプと Token Test の得点との関係を検討した報告もある。Orgass と Poeck¹⁴⁾ は、7種類 (Motor, Predominantly motor, Amnesic, Predominantly amnesic, Sensory, Predominantly sensory, Mixed) の失語型について検討したが、失語の種類による差はみられなかった。Poeck ら¹⁵⁾ は、単に fluent type と non-fluent type の2型に分類して検討したが、やはり差は認められなかった。

Token Test で生ずる誤りの質的側面について、言語学的な立場から分析している研究もあり、Whitaker と Noll²⁷⁾ は子どもを対象にした発達の研究で、誤りの質的分析を行い、誤りの数が有意に多い8項目について言語学的考察を行っている。Poeck ら¹⁶⁾ は、失語症者と非失語脳損傷者を対象に誤りの質的分析を行った結果、両者の誤りの多い項目が類似していること、Whitaker と Noll²⁷⁾ の言語学的仮説は、失語症者の誤りの一部分しか説明できないことを示した。また Noll と Randolph¹³⁾ は、失語症者と正常者を対象に誤りの分析をした結果、失語症者と正常者の困難度のパターンは類似していること、part 5 では意味的な誤りより統語的誤りが多くなることを報告している。

その他、子どもを対象とした妥当性研究も行われている (Lass と Golden⁹⁾, Tallal²⁸⁾, Lass ら⁹⁾)。

Token Test を用いた実験的研究として、Zaidel²⁸⁾ は、交連切癒術や脳半球切除術を受けた患者を対象に、一側の視野で Token Test を実施した結果、左視野での実行が劣ることを示し、失語症者では、右半球損傷で影響の大きい短期聴覚的言語記憶よりも、左半球損傷と関連した言語学的変数により多く影響されると示唆した。Toppin と Brookshire²⁴⁾ は、失語症者を対象に、Token Test 遂行時の視覚的要素の関与について、3つの異なる条件で遅延反応形式でテストをし、失語症者は視覚や視空間の要因を方略として多く用いないことを報告している。

日本における Token Test の利用は、Spreen と Benton²⁰⁾ の短縮版を改訂した日本語版が比較的広く普及している¹⁷⁾。細川 (恵) と細川 (徹)⁹⁾ は、日本語版について失語症者を対象に Token Test が測定している内容を分析し、妥当性を認める一方、失語症の鑑別に限界があることを示唆している。鈴木²¹⁾ は、やはり短縮版を訳改訂したものを用いて、初老期痴呆の巣症状である失語症、ないし潜在的な失語症の診断に用いると報告している。

表 1 失語症群(A-1)、非失語脳損傷群 (NA-1)、対照群(C-1)の人数、性、平均年齢、平均教育年数

	失 語 症 群	非失語脳損傷群	対 照 群
人 数	14	15	18
男	9	12	8
女	5	3	10
平均年齢* (S. D.)	48.4 (12.1)	51.0 (13.8)	45.8 (12.7)
平均教育年数** (S. D.)	10.2 (2.5)	8.3 (1.4)	11.3 (3.4)

* 3群間に有意差なし

** 失語症群×非失語脳損傷群 p<0.05

対 照 群×非失語脳損傷群 p<0.05

失語症群×対 照 群 n. s.

本報告では、full length の Token Test を日本語訳したものについて、日本語での妥当性を検討することを主な目的とし、知能、年齢などの要因が Token Test 得点にどの程度影響するか検討する。さらに、誤りの質的検討を行い、Token Test の臨床的利用について考察したい。

III. 予 備 実 験

A. 方 法

被験者：

昭和51年4月から昭和52年3月までの当科外来患者、および入院患者のうち、失語症患者14名 (失語症群、以下A-1群)、失語のない脳損傷患者15名 (非失語脳損傷群、以下NA-1群) を対象としたが、このほかに対照群として、神経症または神経症傾向のある精神科、および内科入院患者12名と病院職員6名の計18 (以下C-1群) にも同じ検査を行った。これら3群の概要は表1に示した。年齢に関しては3群間に有意差はなく、教育年数はNA-1群が他の2群より有意に低かった。失語症患者の病因はすべて脳血管障害である。失語の有無は医師の診断により決定されたが、その際に失語症鑑別検査の成績も参考に供された。発症から検査までの期間はまちまちであった。利き手は対象者全員が右利きである。

手続き：

Token Test は De Renzi と Vignolo⁹⁾ の原版を著者らが日本語に訳したものをを用いた。用具は20個の token で、その形は円形と長方形の2種類、大きさは大と小の2種類、色は赤、青、黄、白、緑の5種類である。token の実際の大きさは、大きい円形の直径が4cm、小さい円形の直径が2cmであり、長方形は、大きいものが4×2cm、小さいものが3×2cmである。厚紙を貼り合わせて厚さを5mm位にした token に、ポスターカラーを塗りラッカーで仕上げたものを使用した。なお、

表 2 Token Test の構成

part	項目数	口頭命令に含まれる属性	例
1	10	(色×形)	赤の 丸にさわりなさい
2	10	(大きさ×色×形)	大きい 青の 四角に さわりなさい
3	10	(色×形)と(色×形)	白の 丸と 赤の 四角にさわりなさい
4	10	(大きさ×色×形)と(大きさ×色×形)	小さい 黄の 丸と 大きい 黒の 丸に さわりなさい
5	22	(色×形)+文法事項	赤の 丸を 黒の 四角の 上 に おきなさい

token に厚みをつけたのはつまみやすくするためである。

テストの構成は表 2 に示したように 5 つの part からなっている。Part 1～4 までは、各項目とも動詞（さわる touch）と目的語からなる構文を有し、各 part 10 項目ずつである。Part 5 は、それぞれより複雑な文法的要素を取り入れた構文を含む 22 項目からなっている。どの項目も余分なことは全く含まれず、正しく遂行するためには常に正確な理解が必要となる。

検査は静かな部屋で 1 対 1 で行った。被検査者を検査者に対して机に向かい合うように座らせ、机の上に並べた token について各項目ごとに口頭命令に従うよう指示した。なお、token の並べ方は色についてはランダムであるが、形と大きさについては同じものが同列になるように並べた。口頭命令は原則として 1 回のみとし、被検査者のモチベーションを持続させるために 2 回以上繰り返した場合でも、1 回目を採点の対象とした。

採点は Spellacy ら¹⁹⁾が用いた Pass-fail scoring と weighted scoring の両方で行った。Pass-fail scoring では、口頭命令が正しく実行されれば 1 項目に 1 点を与え、誤れば 0 点とし、最高 62 点となる。Weighted scoring は、口頭命令の言語学的要素 1 つ 1 つに重みづけをして採点するもので、part 1 では、1 項目につき色と形に、それぞれ 1 点計 2 点が与えられる。同様に part 2 では 3 点、part 3 では 4 点、part 4 では 6 点がそれぞれ各項目に対して与えられる。Part 5 は項目によって異なり、1 項目につき 5～3 点が与えられ、part 1～5 までの最高可能点は 290 点となる。

B. 結 果

1. Pass-fail scoring による結果

図 1 は各被験者群ごとに各 part の平均得点率を示したものであり、表 3 はその平均得点を実数で示したものである。A-1 群はすべての part で他の 2 群と有意差があり（表 3）、図 1 からあきらかにその差は後の part ほど大きい。NA-1 群と C-1 群の間には、part 4 および 5 で有意差があった。全 part の合計得点

表 3 Pass-fail scoring (PFS) による各群の part 別と全体の平均得点

	part	失語症群	非失語脳損傷群	対 照 群
PFS	1	6.8 ***	9.3	9.7
	2	5.5 **	8.9	9.5
	3	3.4 ***	8.4	9.4
	4	2.2 ***	7.0 *	8.8
	5	3.9 ***	14.6 *	17.9
	全 体	24.9 ***	48.3 **	55.3

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

PFS %

図 1 Pass-fail scoring (PFS) による各群の part 別と全体の得点率

WS %

図 2 Weighted scoring (WS) による各群の part 別と全体の得点率

では、3 群間に有意差があった。なお、有意差検定は以下に示すものもすべて Token Test によった。

2. Weighted scoring による結果

図 2 と表 4 に対象群ごとの平均得点率と平均得点を示した。Pass-fail scoring の場合と同様に、A-1 群はすべての part で他の群との間に有意差を示し、後の part ほどその差が大きかった。NA-1 群と C-1 群の間には、part 3 以後で有意差があった。合計得点では 3 群間にそれぞれ有意差が認められた。

3. 誤りの検討

被験者が示した誤りのすべてを、token の属性であ

表 4 Weighted scoring (WS) による各群 part 別と全体の平均得点

		失語症群	非失語脳損傷群	対照群
W	1	15.9 *	19.3	19.8
	2	22.7 **	28.8	29.5
	3	27.5 ***	37.7 *	39.3
S	4	36.7 ***	54.5 **	58.3
	5	78.1 ***	125.7 *	132.1
	全 体	190.9 ***	266.0 **	279.0

*p<0.05 **p>0.01 ***p<0.001

表 5 各群における要素別の誤り率

	形	色	大きさ	口頭忘却	文法誤り
失語症群	18.5%	11.9%	7.0%	8.2%	50.0%
非失語脳損傷群	4.3	6.3	4.2	1.5	20.0
対照群	1.8	1.9	1.1	1.1	10.4

る形、色、大きさの誤り、口頭命令の忘却、および文法的誤りの5種類に分け、それぞれについて3群を比較した(表5)。色、形、および大きさの誤りに関しては、それぞれの全提示回数に対する総誤り数の割り合いを百分率で示した。口頭命令の忘却というのは、不注意やモチベーションの低下あるいは記憶力障害などによると思われる命令の一部、あるいは全部の忘却を意味し、これが62項目中に占める割り合いを算定した。文法的誤りは part 5 の22項目についてその割り合いを比較した。A-1群では、token の3属性のうち形の誤りが最も多く、ついで色の誤りが多かった。これに対して非失語群では、形よりも色の誤りが多かった。大きさの誤りは3群とも比較的少なかった。口頭命令の忘却はA-1群に多く、最も簡単な part 1 にも認められ、part 3, 4, 5 と命令の要素数が多くなり複雑になるにつれて増加する傾向があった。NA-1群では part 4 で多くみられ、C-1群では part 5 でわずかに認められた。文法的誤りはA-1群に最も多く、問われた文法事項の50%にも達していた。

4. 色の誤りについて

検査中、青と緑の混同がしばしば観察されたため、この2色の混同が色の誤りのうちのどの位の割り合いを占めるか、また形や大きさの誤りはどの色との組み合わせのときに多いかを検討した。表6に、それぞれの誤りの%を示す。3群とも、色の誤りのうち50%以上が青と緑の混同による誤りであり、3群とも偶然による誤りの確率40%を越えていた。形については、A-1群とC-1群で偶然による誤りの確率40%を越えており、大きさについても、NA-1群とC-1群で40%を越えていた。以上の結果より、青と緑の混同が反応に影響を及ぼしていることはあきらかで、とくにC-1群でその影響が最も大き

表 6 各群における「青」「緑」のときの誤り率

	青と緑の誤り	青あるいは緑のときの形の誤り	青あるいは緑のときの大きさの誤り
失語症群	55.1%	42.5%	4.8%
非失語脳損傷群	60.6	29.2	47.4
対照群	77.8	47.1	60.0
偶然による誤りの確率	40.0	40.0	40.0

表 7 失語症群(A-2)、非失語脳損傷群(NA-2)の診断別人数

診断分類	失語症群 (28人)	非失語脳損傷群 (23人)
左半球損傷		
脳血管障害	26	5
頭部外傷	1	0
右半球損傷		
脳血管障害	1	9
脳腫瘍	0	2
両側性損傷		
脳血管障害	0	2
頭部外傷	0	2
初老期痴呆	0	1
脳炎	0	1
その他	0	1

表 8 失語症群(A-2)、非失語脳損傷群(NA-2)、対照群(C-2)の人数、性、平均年齢、平均教育年数

	失語症群	非失語脳損傷群	対照群
人 数	32	23	20
男	27	13	5
女	5	10	15
平均年齢* (S.D.)	48.8 (11.1)	50.8 (12.0)	49.3 (9.3)
平均教育年数* (S.D.)	10.5 (3.2)	9.3 (2.9)	9.5 (2.6)

* 3群間に有意差なし

いことが示された。

IV. 実 験

A. 方 法

被験者:

昭和52年4月から昭和54年12月までに当科を受診した患者のうち、失語症患者28名(失語症群、以下A-2群)、失語のない脳損傷患者23名(非失語脳損傷群、以下NA-2群)を対象とし、これに神経症またはうつ状態の診断で当科に入院中の患者12名と、病院職員8名の計20名を対照群(以下C-2群)として加えた。

A-2群とNA-2群の診断別内訳は表7に示したとおりである。利き手については、NA-2群の両側性脳血

管障害の1例が両手利きであった以外はすべて右利き者であった。表8は3群の概要を示したものである。A-2群では、4名がテストを完全に遂行できなかったので結果の整理から除いた。そのためA-2群のデータは、最終的には24名についてであったが、再検査データ（6人が2回、1人が3回、回復過程を追って再検査された）も含めたため延べ32人となった。年齢および教育年数に関しては、3群間に有意差はなかった。発症から検査までの期間は、A-2群は26日～4年10カ月の範囲内に、またNA-2群は22日～8年の範囲内にあった。検査時点で片麻痺が認められたのは、A-2群で16例（いずれも右片麻痺）、NA-2群で14例（右4例、左10例）であった。

手続き：

予備実験で得られた結果を考慮して、検査手続きの一部を改訂した。改訂の第1は token についてであり、緑色を黒色に、また長方形を正方形（大きいのは一辺4cm、小さいのは一辺2cm）に変えた。形の変更については、口頭命令の「四角」に対して、token の長方形は「なが四角」であり、そのために反応困難であった失語症患者がいたためである。第2に、token の並べ方を固定し、テスト材料の条件を被検者に対して等しくした（付録参照）。第3の改訂は指示内容に関するもので、大きさ、色、形の各下位要素の出現頻度を等しくした。さらに第4に、part 5の口頭命令の表現の一部を、より日本語らしいまわしに変更した。テストの構成は原版と同様に5つの part、計62項目からなるが、予備検査として part 1の前に、形、色、大きさの弁別検査を加えた（付録参照）。

検査の実施は予備実験と同様である。

採点も予備実験で用いた2種類の採点法で行ったが、weighted scoring では part 5の一部を改訂したため最高可能点は292点になった。

対象となった被験者は、C-2群のうちの病院職員8名を除いて、他の全員に可能な限り WAIS を実施した。さらに失語症患者に対しては、スピーチ・セラピストにより失語症鑑別検査（SLTA、老研版失語症鑑別検査）が実施された。なお、これらの検査は Token Test と同じ日に、あるいはその前後（1週間以内）に行われた。

B. 結果

1. Pass-fail scoring による結果

図3は、各被験者群の平均得点率を part ごとに示したものであり、表9は各群の平均得点ならびに各群間の有意差検定の結果である。A-2群とNA-2群の間では各 part と全体の平均得点のそれぞれに有意差が認めら

図3 Pass-fail scoring (PFS) による各群の part 別と全体の得点率

図4 Weighted scoring (WS) による各群の part 別と全体の得点率

表9 Pass-fail scoring (PFS) による各群の part 別と全体の平均得点

	part	失語症群	非失語脳損傷群	対照群
PFS	1	8.6	10.0	10.0
	2	7.3	9.2	9.9
	3	5.3	9.0	9.7
	4	3.8	8.0	9.1
	5	8.1	14.6	18.4
	全体	33.0	51.5	57.0

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

れた。NA-2群とC-2群の間では part 4, 5および全体の平均得点に有意差があった。

2. Weighted scoring による結果

Pass-fail scoring の場合と同様に、図4に各群の平均得点率を示し、表10に平均得点と各群間の有意差検定の結果を示した。A-2群とNA-2群の間では各 part と全体の平均得点のそれぞれに有意差があり、NA-2群とC-2群の間では part 5と全体の平均得点に有意差が認められた。

両採点法の結果を比較すると、pass-fail scores の方が群間の差を検出するのに優れていると認められるため、以下の分析は pass-fail scores について行った。

表10 Weighted scoring (WS) による各群の part 別と全体の平均得点

	part	失語症群	非失語脳損傷群	失語症群	対照群
W	1	18.3	*	20.0	20.0
	2	26.2	**	29.2	29.9
	3	31.7	***	38.8	39.6
S	4	45.4	***	57.4	58.8
	5	104.5	***	129.0	135.8
	全 体	226.3	***	274.1	284.0

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

図 5 Pass-fail scoring (PFS) による各群の得点分布

3. 3被験者群の得点分布

図 5 に、個々の被験者の総得点の分布を示した。平均値では 3 群間に有意差があったが、個々の分布をみると群間にある程度の overlap が認められた。そこで、C-2 群の最低点である 48 点を cut-off score として採用する^{1,22,25)}と、A-2 群の 93.8% が正しく弁別され、NA-2 群の 17.4% が失語群として誤って弁別されることになる。

4. Token Test の得点と年齢および知能との関係

総得点と年齢、WAIS の言語性 IQ (VIQ) および動作性 IQ (PIQ) とのそれぞれの相関を、Pearson の相関係数を求めて検討し、表 11 にその結果を示した。まず、Token Test の得点と年齢との間には、A-2 群で有意な負の相関があったが、他の 2 群では有意な相関はなかった。つぎに Token Test の得点と IQ との関係は、A-2 群では VIQ, PIQ とともに有意な正相関があり、NA-2 群では VIQ と有意な正相関があった。C-2 群ではどちら

表11 各群の pass-fail scoring (PFS) による得点と年齢、VIQ, PIQ との相関 (Pearson 相関)

	失語症群	非失語脳損傷群	対照群
年齢×T. T. PFS	-0.368* (n=32)	-0.014 (n=23)	-0.070 (n=20)
VIQ×T. T. PFS	0.734* (n=11)	0.726** (n=14)	0.125 (n=9)
PIQ×T. T. PFS	0.636* (n=13)	0.053 (n=12)	0.044 (n=9)

* $p < 0.05$ ** $p < 0.01$

表12 各群における要素別の誤り率

	形	色	大きさ	口命忘却	文法誤り
失語症群	17.1%	12.7%	9.5%	3.5%	41.9%
非失語脳損傷群	3.2	2.3	4.2	0.4	24.3
対照群	1.1	0.9	0.8	0.3	14.1

表13 各群における総誤り得点とそれに占める要素別誤りの割合

	総誤り得点	形の誤り	色の誤り	大きさの誤り	忘却および文法的誤り
失語症群 (N=32)	2102	27.1%	19.3%	4.3%	49.3%
非失語脳損傷群 (N=23)	411	18.5	12.9	7.1	61.6
対照群 (N=20)	160	13.8	10.6	3.1	72.5

の IQ とも有意な相関はなかった。

5. 誤りの検討

緑を黒に変えたことが、被験者の誤りにどう影響したかを分析した。予備実験と同様に、各属性の提示回数に対する誤り数の割合を百分率で示し、これを 3 群で比較した。結果は表 12 に示したとおりである。青と緑の混乱がなくなったことにより、色の誤り率が 3 群とも減少することが予想されていた。実際には、NA-2 群と C-2 群は、NA-1 群、C-1 群に比べ約半数に減少しているが、A-2 群は A-1 群に比べわずかに増加している (表 5)。

つぎに、weighted scores の減点分として表れる総誤り得点に対して、各要素の誤りがどの位の割合を占めているかを、3 群で比較した。表 13 はその結果である。3 群とも忘却と文法的誤りの占める割合が最も高かった。token の 3 属性については、3 群とも形の誤りが最も多く、ついで色、大きさの順であった。

各被験者群がおかす文法的誤りについて検討するため、part 5 で各項目ごとに各群の誤る割合を算定し、それぞれの群で全 22 項目を文法的誤りの多い順に並べてみた (表 14)。項目 3, 5, 6 は 3 群とも上位に入っており、とくに項目 3 はどの群も誤る割合が 70% 前後の

表14 各群の part 5 における文法的誤りの多い順位と誤る割合

順位	失語症群		非失語脳損傷群		対照群	
	項目 No.	誤る割合	項目 No.	誤る割合	項目 No.	誤る割合
1	5	77%	6	82%	3	80%
2	3	73	3	68	5	70
3	12	67	9	55	6, 9	35
4	6, 20	63	5, 14	45		
5					2	20
6	2	60	22	32	12, 22	15
7	10	53	2	27		
8	14, 21	50	7, 12, 20, 21	18	15, 16	10
9						
10	22	47			7, 14, 20, 21	5
11	9, 16	40				
12			4, 10, 18	14		
13	7, 15	30			1, 4, 8, 10, 11, 13, 17, 18, 19	0
14						
15	4, 8	27	8, 11, 16, 17	9		
16						
17	17	23				
18	18	20				
19	11	17	1, 13, 15, 19	5		
20	13	13				
21	19	10				
22	1	3				

高率である。A-2 群で上位にあり、誤る割合が60%以上の項目 12, 20は、NA-2 群、C-2 群では中位で誤る割合も 20% 以下である。項目 9は、NA-2 群、C-2 群で上位にあり、A-2 群で中位にあるが、誤る割合は群間で大きな違いはない。一方、項目 1, 13, 19は、3 群とも誤りの少ない項目である。また、ほぼすべての項目で A-2 群は他群に比べて誤る割合が高い。

6. 失語の重症度と Token Test の得点

A-2 群を対象として、Token Test 得点と失語の重症度との関係を検討した。失語の重症度については、まず、わが国の患者向けに修正した Schuell の重症度尺度¹⁸⁾ (表15) に従って「聞いて理解する能力」と「話す能力」をそれぞれ7段階に分け、つぎにこれら2つの能力に関する評点から図6の区分に従って「重度」、「中等度」および「軽度」の3段階に評価した。図7はそれぞれの重症度に評価された失語症者の Token Test 得点の分布を示したものである。この図から、それぞれの間にはある程度の overlap もみられるが、Token Test 得点は失語の重症度をかなり反映しているとわかる。さらに詳しく分析するため、失語症鑑別検査による評価と

表15 失語症重症度尺度

I. 聞いて理解する能力

0. 障害なし
1. ラジオの放送番組（天気予報、ニュース）や、少人数での話し合いに、ほとんど苦勞なくついていけるが、細かい点を聞き落とすことがある。
2. 普通の日常会話にはほとんど支障なくついていけるが、複雑な内容の理解は困難である。
3. たいていの日常会話についていけるが、ときどき肝心なことがらを聞きのがす。
4. 簡単な会話にはついていけるが、相手に繰り返していってもらわないと、わからないことが多い。
5. ジェスチャーなどの手がかりを混えて何回か聞いてもらえば、ごく短い文や単語による話しかけがどうか理解できる。
6. 実用となる理解力はない。

II. 話す能力

0. 障害なし
1. 日常場面では大体楽に話すことができるが複雑な内容を表現しようとすると、迂遠になったりまとまりを欠くことがある。
2. 普通の日常会話ができるが、構文が少々おかしかったり、喚語困難などのため発語が途切れたりすることがある。
3. 簡単な会話はできるが文による表現は困難で断片的な句や単語によって意思を伝達しようとする。
4. 単語のられつによってある程度実用的に意思を伝達することができる。
5. 使用する単語も極めて限られており、誤りの多い話し方ではあるが、自分の要求や望みのある程度は相手に伝えることができる。
6. 実用となる話し言葉はない。

(笹沼澄子他、文献¹⁸⁾より引用)

図6 失語症重症度の重度、中等度、軽度の区分

Token Test 得点の相関について検討した。表16は、聴く過程（単語、短文、口頭命令の理解）ならびに話す過程（呼称、動作説明、叙述）の正答率と、Token Test の得点との関係について、それぞれ算出した Spearman の順位相関係数を示したものである。この表が示すように、両者とも有意な正の相関があり、とくに聞く過程と

図 7 Pass-fail scoring (PFS) による失語症群 (A-2) の重症度別得点分布

表16 失語症鑑別検査による評価と Token Test 得点の相関 (Spearman 順位相関)

聴く (単語, 短文, 口頭命令の理解) × T.T. PFS : 0.678***
話す (呼称, 動作説明, 叙述) × T.T. PFS : 0.466*

* $p < 0.05$ *** $p < 0.001$

の相関が高かった。

V. 考 察

予備実験の結果として, De Renzi と Vignolo の原版を忠実に訳した Token Test をそのまま日本人に適用するには, 青と緑の混乱があることや要素の提示頻度に偏りがあることなどから, 改訂して使う必要が示された。改訂した62項目の Token Test で Pass-fail scoring のときの cut-off score を48とすると, 失語症群の93.8%が正しく分類され, 非失語脳損傷群の17.4%が誤って分類された。両者を総合すると正しく分類されたのは89.1%である。この値は, 以前の研究と類似しており^{21,24)}, 日本語版62項目の Token Test の妥当性を裏づけている。

対照群では, Token Test 得点と年齢, IQ の間に有意な相関はなく, ここで用いた対照群の年齢範囲 (34~65歳) と IQ 範囲 (78~133) に限られるが, Token Test 得点は年齢や IQ とはほぼ無関係であることが示された。しかし脳損傷患者群では, そのいずれかと有意な相

関がみられた。失語症群では, Token Test 得点は年齢と負の相関があり, 年齢が高いほど Token Test 得点が低下することを示し, 失語症群では加齢による影響の大きいことが示唆された。さらに WAIS の VIQ, PIQ とともに正の相関があり, テスト時点での測定 IQ が高いと Token Test 得点も高くなるといえる。非失語脳損傷群でも VIQ と正相関があり, テスト時点での言語性知能のレベルが Token Test 遂行に影響するといえる。Token Test 得点と PIQ との相関が失語症群で有意であり, 非失語脳損傷群では有意でなかったことは, 失語症群では Token Test 遂行に対する非言語的要素の影響が, 比較的強大であることを示唆している。したがって, Lesser¹⁰⁾ のいうように, 失語症者の Token Test 遂行時には, 言語能力に加えて, 非言語的な要素である視覚情報の処理能力や, 反応動作機能の要因の関与を軽視してはならないのであろう。

誤りの分析では, 青と緑の混乱を避けるため緑を黒に変更した結果, 非失語脳損傷群で色の誤りに減少をもたらした。すなわち, 原版 (予備実験) では非失語脳損傷群の色の誤りは, 青と緑の混乱の影響が大きかったといえる。これに対し, 失語症群では変更後の色の誤りの生起率は低下せず, かえってわずかに増加したほどであり, 失語症群における青と緑の混乱は, 色の誤りのわずかな部分しか占めていなかったといえよう。あるいは, 失語症群の色の誤りは, 青と緑の混同というレベルに達していない, それ以前の色を表すことばの了解の問題であるといえるかもしれない。緑を黒に変更した結果, 3群とも token の3属性の誤りは, 形が最も多く, ついで色, 最も少ないのは大きさとなった。Kreindler⁷⁾ は, 形, 色, 大きさの3属性の語順と失語症者の了解の関係について調べた結果, 語順の最後のものの了解がはじめの2つより容易であること, そして語順に関係なく形の了解が最も難しく, つぎに色, 大きさの順であると報告している。ただし, この著者らの母国語であるルーマニア語ではこの3属性の普通の語順は「形・大きさ・色」あるいは「形・色・大きさ」である。これに対し, 日本語や英語では普通の語順は「大きさ・色・形」あるいは「色・大きさ・形」であり, ルーマニア語の語順とは逆である。しかし, 本研究における了解の困難度は, 形>色>大きさの順であり, ルーマニア語の場合と同じであることは, 語順よりも, むしろそれぞれの属性の特質としての使用頻度や抽象的概念の程度などが, 了解の困難性と関係していると考えられる。

Part 5 では, 大きさ, 色, 形に関する誤りのほかに, 動詞や前置詞などの了解障害に基づく誤りが数多く出現する。われわれはこのような誤りを欧米の文献にならっ

て「文法的誤り（または統語的誤り）」と呼ぶことにしたが、これらについて若干検討を加えておきたい。各被験者群が part 5 でおかす誤りは、他の part に比してあきらかに多数であったが、これは文法的な誤りが新たに加わるからである。項目 3, 5, 6 はどの群でも最も誤りやすかった。これらの項目では、同じ token を選択させられるのだが、要求される動作は少しずつ異なっており、相互の違いを正確に区別することが必要である。とくに項目 5 は part 3 の口頭命令と全く同じであるが、part 3 では正しく反応した被験者が part 5 の中で項目 3, 4 に引き続いて命ぜられると、意外にこの項目で誤ることが多いのである。この理由は新しい口頭命令として受け取られるためなのか、あるいは part 1～4 までは「さわりなさい」という一定した動作の要求に対し、半自動的に反応すればよかったのに、この part では項目ごとに動詞が変化しており、その都度異なった動作が要求されることによる混乱のためであろう。PoECK¹⁶⁾は、失語症群と非失語脳損傷群のそれぞれ 100 名について、誤りの頻度により項目をランクづけているが、やはり 3, 5, 6 はいずれも上位を占めている。Whitaker と Noll²⁷⁾、Tallal²⁸⁾ は子どもについて文法的誤りの多い項目を検討しているが、項目 3, 4, 5 が最も誤りが多い群に入っており、項目 6 はこれに入っていない。項目 3, 4, 5 における動作の区別は子どもにとっては非常に難しいが、項目 4 は語順のとおりに行えばよく、大人では割り合いやさしいと考えられる。これに対し、大人で項目 6 が困難というのは、それまで 2 つの token がいわれた場合には 2 つとも操作する必要があったが、この項目にきて一方を選択するという新しい事態が要求されるためであろうと考えられる。いずれにしても、項目 3～6 は連続して同じ token を用いることが、要求された動作の区別をより困難にしていると考えられる。失語症群では part 5 の半数に近い 9 項目で被験者の 50% 以上が誤るのに対し、非失語脳損傷群と対照群では 50% 以上が誤るのはそれぞれ 3 項目と 2 項目に過ぎない。この割り合いは、Noll と Randolph¹⁸⁾ が成人の軽度失語症患者 25 名と対照群 25 名について行った結果とほぼ一致している。しかし、われわれのランクは Noll らのランクの順位とあまり一致しておらず、むしろ PoECK らのものとよく似ている。

Token Test は失語症患者のいかなる欠陥をうつしだして、非失語症者からの識別を可能にするのかということはまだよくわかっていないが、これが臨床的に有用な検査であることは衆目の認めるところとなっている。しかしこのテストが実際にどれほど有効に利用されているかということは問題である。そこで、われわれの実験に

おける失語症患者間の得点の差は何によるものかを考え、より有効な臨床的使用について考えてみたい。先に、失語症群ではこのテストの成績が IQ や年齢と関連することを述べたが、それに加えて失語症の重症度をもかなり反映していることが示された。さらに、Token Test と同じ時期に実施した失語症鑑別診断検査との関係を調べると、聴く過程および話す過程の正答率と Token Test 得点に有意な相関があった。すなわち、聴く過程が良いものほど Token Test 得点は高いことはもちろん、話す過程の良さとも相関していることから、このテストは失語症のタイプにかかわらず識別の道具として使用できるものと考えられる。Morley¹²⁾は、聴覚言語理解を測定するとされる 3 つの検査を比較し、それぞれの特徴を指摘している。それによると、Boston 失語症診断検査 (BDAE) は最も広い弁別範囲を持つという。Porch 理解能力指標 (PICA) は重症の聴覚言語理解障害の患者で、言語治療の有効性について判定するのに適しているが、高いレベルの弁別力は低い。それに対し、Token Test は重症の患者には難しすぎ、高いレベルの患者のテストとして範囲が広く、有効であるとされている。われわれの経験でも、重症度の高い患者には 62 項目の Token Test は、1 時間以上も要することがあり、患者も検者も非常に疲れることになる。これに対し、他の失語症鑑別検査ではほとんど障害の目だたなくなった患者では、テスト時間は 20 分程度ですみ障害の検出にも優れていることがわかる。時間を短縮する目的で Token Test の short form が開発されたり、重症の患者にも使用しようとして変法が工夫されたり (Martion¹¹⁾) しているが、De Renzi らの当初の目的である、軽微な聴覚言語理解の障害を検出するテストとしての価値を尊重するなら、62 項目の Token Test を高いレベルの患者のテストとして使用するのが最も妥当と思われる。日本では、Token Test の他に聴覚言語理解を測定するディープテストがあまり開発されていないが、BDAE や PICA のような測定範囲の異なるテストの開発も、これからの重要な課題であると思われる。

VI. 要 約

1. まず、予備実験として、失語症患者 14 名、非失語脳損傷患者 15 名、および対照群 18 名 (神経症または神経症傾向のある入院患者 12 名と病院職員 6 名) について、De Renzi と Vignolo の原版を忠実に訳した Toke Test を施行した。その結果、pass-fail scores でも weighted scores でも 3 群間に有意な得点差が認められた。しかし、「青」と「緑」の混乱を招いたり、要素の提示頻度に偏りがあることから、日本人に合った改訂が必要と考

えられた。

2. つぎに、本実験として、「緑」を「黒」に、「長方形」を「正方形」に変え、要素の提示頻度を等しくするなどの改訂を加えた Token Test を失語症患者32名、非失語脳損傷患者23名、および対照群20名（神経症またはうつ状態の診断で入院中の患者12名と病院職員8名）に施行した。改訂した Token Test の妥当性、Token Test 得点と年齢や IQ との関係、Token Test で生じた誤りの分析、Token Test 得点と失語症鑑別検査の成績との関係などについて検討しつぎの結果を得た。

1) Pass-fail scores, weighted scores とともに対象の3群間に有意差があった。両採点法を比較すると、pass-fail scores の方が群間の差を検出するのに優れていた。

2) Pass-fail scores の cut-off score を対照群の最低点であった48とすると、失語症群の93.8%が正しく分類され、非失語脳損傷群の17.4%が失語症に分類された。全体では89.1%が正しく分類された。

3) 対照群では、Token Test 得点と年齢、IQ の間に有意な相関はなかった。失語症群では、Token Test 得点は年齢と負の相関があり、WAIS の VIQ ならびに PIQ とは正の相関があった。非失語脳損傷群では、Token Test 得点と年齢との相関はなく、WAIS の VIQ と正の相関があった。

4) Token の3属性についての誤りを分析したところ、失語症群では形>色>大きさの順に誤り率が高かった。これに対し、非失語脳損傷群と対照群は、これら3属性の誤り率は失語症群の大きさの誤り率よりも低く、3つの間にあきらかな差はなかった。文法的誤り（統語的誤り）については、失語症群が各項目で最も高い誤りの頻度を示したが、3群の誤りのパターンには大きな違いはなかった。

5) 失語症群では、Token Test 得点と失語症鑑別検査の聴く過程、および話す過程の正答率との間に正の相関があった。

以上の結果に基づいて、日本語版62項目の Token Test の妥当性と Token Test 得点を規定する要因について考察を加えた。

本研究に対して多くの有力なご助言を賜った金沢大学文学部心理学教室の田中富士夫教授に深甚の謝意を表します。

文 献

- 1) Boller, F. and Vignolo, L. A. : Latent sensory aphasia in hemispheredamaged patients : An experimental study with the Token Test. *Brain*, 89 : 815~830, 1966.
- 2) Cohen, R., Kelter, S., Engel, D., List, G. and Strohnner, H. : Zur validitat des Token-Tests. *Der Nervenarzt*, 47 : 357~361, 1976.
- 3) De Renzi, E. and Vignolo, L.A. : The Token Test : A sensitive test to detect receptive disturbances in aphasics. *Brain*, 85 : 665~678, 1962.
- 4) De Renzi, E. and Foglioni, P. : Normative data and screening power of a shortened version of the Token Test. *Cortex*, 14 : 41~49, 1978.
- 5) Hartje, W., Kerschensteiner, M., Poeck, K. and Orgass, B. : A cross-validation study on the Token Test. *Neuropsychologia*, 11 : 119~121, 1973.
- 6) 細川恵子, 細川 徹 : Token Test による失語症の評価とその妥当性. 聴覚言語障害, 6 : 77~86, 1977.
- 7) Kreindler, A., Gheorghita, N. and Voinescu, I. : Analysis of verbal reception of a complex order with three elements in aphasics. *Brain*, 94 : 375~386, 1971.
- 8) Lass, N.J. and Golden, S.S. : A comparative study of childrens performance on three tests for receptive language abilities. *J. Auditory Reserch*, 15 : 177~182, 1975.
- 9) Lass, N. J., De Paolo, A. M., Simcoe, J.C. and Samuel, S.M. : A normative study of children's performance on the short form of the Token Test. *J. Communication Dis.*, 8 : 193~198, 1975.
- 10) Lesser, R. : Verbal and non-verbal memory components in the Token Test. *Neuropsychologia*, 14 : 79~85, 1976.
- 11) Martino, A.A., Pizzamiglio, L. and Razzano, C. : A new version of the "Token Test" for aphasics : A concrete objects form. *J. Communication Dis.*, 9 : 1~5, 1976.
- 12) Morley, G.K., Lundgren, S. and Haxby, J. : Comparison and clinical applicability of auditory comprehension socres on the Behavioral Neurology Deficit Evaluation, Benton Diagnostic Aphasia Examination, Porch Index of Communicative Ability and Token Test. *J. Clin. Neuropsychol.*, 1 : 249~258, 1979.
- 13) Noll, J.D. and Randolph, S.R. : Auditory semantic, syntactic, and retention errors made by aphasic subjects on the Token Test. *J. Communication Dis.*, 11 : 543~553, 1978.
- 14) Orgass, B. and Poeck, K. : Clinical validation of a new test for aphasia : An experimental study

- on the Token Test. *Cortex*, 2 : 222~243, 1966.
- 15) Poeck, K., Kerschensteiner, M. and Hartje, W. : A quantitative study on language understanding in fluent and nonfluent aphasia. *Cortex*, 8 : 299~304, 1972.
 - 16) Poeck, K., Orgass, B., Kerschensteiner, M. and Hartje, W. : A qualitative study on Token Test performance in aphasic and non-aphasic brain damaged patients. *Neuropsychologia*, 12 : 49~54, 1974.
 - 17) 笹沼澄子 : Token Test の手引 (笹沼澄子編) : 言語障害. 129~134, 医歯薬出版, 1975.
 - 18) 笹沼澄子, 伊藤元信, 綿森淑子, 福迫陽子, 物井寿子 : 失語症分類と重症度. 失語症の言語治療, 33~38, 医学書院, 1978.
 - 19) Spellacy, F. and Spreen, O. : A short form of the Token Test. *Cortex*, 5 : 390~397, 1969.
 - 20) Spreen, O. and Benton, A.L. : Neurosensory Center Comprehensive Examination for Aphasia, edition A. Neuropsychology Laboratory, University of Victoria, 1969.
 - 21) 鈴木康之 : 初老期痴呆における Token Test. 名市大医学会雑誌, 27 : 447~464, 1977.
 - 22) Swisher, L.P. and Sarno, M.T. : Token Test scores of three matched groups : Left brain-damaged with aphasia ; Right brain-damaged without aphasia ; Non-brain-damaged. *Cortex*, 5 : 264~273, 1969.
 - 23) Tallal, P. : Perceptual and linguistic factors in the language impairment of developmental dysphasics : An experimental investigation with the Token Test. *Cortex*, 11 : 196~205, 1975.
 - 24) Toppin, C. J. and Brookshire, R. H. : Effects of response delay and token relocation on Token Test performance of aphasic subjects. *J. Communication Dis.*, 11 : 65~78, 1978.
 - 25) Van Dongen, H.R. and Van Harskamp, F. : The Token Test : A preliminary evaluation of a method to detect aphasia. *Psychiat. Neurol. Neurochir.*, 75 : 129~134, 1972.
 - 26) Van Harskamp, F. and Van Dongen, H.R. : Construction and validation of different short forms of the Token Test. *Neuropsychologia*, 15 : 467~470, 1977.
 - 27) Whitaker, H. A. and Noll, J. D. : Some linguistic parameters of the Token Test. *Neuropsychologia*, 10 : 395~404, 1972.
 - 28) Zaidel, E. : Unilateral auditory language comprehension on the Token Test following cerebral commissurotomy and hemispherectomy. *Neuropsychologia*, 15 : 1~18, 1977.

別刷請求先 : 〒920-02 河北郡内灘町字大学 1-1
 金沢医科大学神経精神医学教室
 平口真理