

日本語リーディングスパン・テストにおけるターゲット語と刺激文の検討

大塚一徳・宮谷真人

Target words and sentences for Japanese version of the reading span test

Kazunori Otsuka and Makoto Miyatani

日本におけるワーキングメモリスパン課題の一つである日本語版 RST(苧阪, 1998, 2002; 苧阪・苧阪, 1994)は, ワーキングメモリ研究の有力な研究道具として広く普及している。本研究は, 日本語版 RST(苧阪, 2002)のターゲット語と刺激文について検討し, 日本語版 RST を実施するうえでの基礎的データを提供することが第1の目的である。さらに, 日本語版 RST を内的整合性と再検査法によって検討し, 信頼性について4つの得点化法に基づいて検討することが第2の目的である。まず, 日本語版 RST の刺激文とターゲット語について詳細に検討し, その特徴について記述した。次に, 148名の大学生の実験参加者に日本語版 RST を実施し, 4つの得点化法ごとに分布の正規性及び日本語版 RST の内的整合性について検討した。次に51名の実験参加者に2回目の日本語版 RST を実施し, 各得点化法の信頼性について検討した。その結果, 日本語版 RST のターゲット語と刺激文についてその属性ごとの特性や正答率に関する基礎的データを提供することができ, 日本語版 RST の内的整合性と信頼性について実証できた。

キーワード: 日本語版リーディングスパン・テスト, ワーキングメモリ, 信頼性

問題

ワーキングメモリは, 処理した情報の保持と, その情報に基づく認知活動を並列的にダイナミックに展開する一時的な記憶システムのことをいう(Just & Carpenter, 1992)。ワーキングメモリは, 言語処理(Miyake, Just, & Carpenter, 1994)や, 演繹推理(Gilhooly, Logie, Wetherick, & Wynn, 1993)など, 課題の遂行と情報の保持が並列的に処理されることが必要な多くの認知活動と深く関わっていることが指摘されている(苧阪・苧阪, 1994)。従来の短期記憶が情報の貯蔵庫としての機能を重視するのに対し, ワーキングメモリは, 認知課題の遂行中に情報がいかに操作されるかといった, 情報の処理機能を重視する記憶システム概念である(Miyake et al., 1994)。

ワーキングメモリ研究において, Daneman & Carpenter (1980)によって開発されたリーディングスパン・テスト(Reading Span Test, 以下 RST と略)は, 有益な研究道具として広く使われている(齊藤・

三宅, 2000)。RST は次々と提示される文を実験参加者自身に口頭で読ませながら、文中の単語を記憶させていくという手法を用いるものである。このテストにおいて読みと並列的に記銘する単語の再生数から算出される得点がリーディングスパンである。RST の実験参加者は、音読をしながら単語を保持するという二重課題を行う必要があり、ワーキングメモリでの並列的な処理が要求される(近藤・森下・苧阪, 1999)。

RST はもともと読みにおけるワーキングメモリ容量を測定するために開発されたもので、基本的には記憶課題でありながら、読み理解テストや言語の理解度を測定する言語的な認知課題と高い相関を示すテストであるところに特徴がある。ただし、RST に含まれると想定される心的過程は数多いため、RST がいったい何を測定しているのかについてはいまだ多くの議論や解釈が存在する(齊藤・三宅, 2000)。

RST の他にも、課題の遂行と情報の保持が要求されるワーキングメモリスパン課題として様々なものが考案されている。例えば、文とそこに含まれるターゲットが聴覚的に提示され、聴取しながらターゲット語を覚えるというリスニングスパン・テスト(Daneman & Carpenter, 1980)、カード内のドットの数をかぞえながらその数を覚えておくという計数スパン(counting span: Case, Kurland, & Goldberg, 1982)、計算課題を行いその答えの部分覚えておくという演算スパン(operation span: Turner & Engle, 1989)、また心的回転(mental rotation)を行い、回転すべき文字の頭の部分の方向を覚えておくという空間スパン(spatial span: Shah & Miyake, 1996)等がある。RST と同様にこのような他のワーキングメモリスパン課題もワーキングメモリ研究の有益な研究道具として広く利用されている(Miyake, 2001; Saito & Miyake, 2004)。

認知研究においてワーキングメモリスパン課題が有力な研究道具として広汎に利用されるようになるにつれ、課題の信頼性、得点化方法、実施方法等の検討も行われるようになってきた(Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005; Klein & Fiss, 1999; Unsworth, Heitz, Schrock, & Engle, 2005)。特に、現在多くのワーキングメモリ研究が RST を利用していることから、RST に焦点をあて、課題の信頼性や得点化方法についての検討も行われている(Friedman & Miyake, 2005; Tirre & Peña, 1992; Waters, & Caplan, 1996, 2003)。

日本では、苧阪・苧阪(1994)によって開発された日本語版 RST が普及している(齊藤・三宅, 2000)。この日本語版 RST は、Daneman & Carpenter(1980)の基本的な手続きを忠実に踏襲しているものの、英語版とのいくつかの違いがある。もっとも大きな違いとしては、日本語の構文の特徴による制約から、ターゲット語を常に英語版のように文の最後の単語とすることは困難であり、文中の任意の位置にある単語にアンダーラインをひいてターゲット語を指定していることが挙げられる。苧阪・苧阪(1994)によって開発された日本語版 RST は、苧阪(1998)版で一部改定が行われ、さらに苧阪(2002)においても一部刺激文等の見直しが行われ、多くの研究で利用されている(中條・中尾, 2005; 西崎・苧阪, 2004; 大塚, 2003; Saito & Miyake, 2004)。

日本語版 RST の標準的な実施手続きは、以下の通りである。例えば 2 文条件では次のような 2 文が、それぞれ 1 枚のカードに印字されて順に提示される。この時、実験参加者が様々な記銘方略を使うのを妨げるために、文と文の間にいっさいポーズを入れないように注意が払われる。

“それは、ゆれながら水銀のように光って上に上がりました。”

“二人の子供が、青い湖のそばで遊んでいました。”

この2文に続いて白紙のカードが提示され、それを合図に実験参加者は刺激文中に下線をひいて示してあるターゲット語を再生する。ただし、親近性効果を考慮してリスト最後の単語を最初に再生することはしないように教示される。採点の際には再生の順序は考慮されない。文条件では上記のような2文条件から、5つの文から構成され5つの単語の再生を求められる5文条件まであり、それぞれの条件につき5セットが用意されている。まず、2文条件から開始され、通常の記憶範囲実験と同じように、2文条件をクリアすると3文条件へ、というように次々と昇順で進んでいく。試行はある文条件で3セット以上再生できなくなるまで続き、2セット以下しか再生できない時点で打ち切る場合と、5文条件まで各文条件の再生数に関係なく続ける場合がある。現在では、パーソナル・コンピュータを用いて、刺激文の提示等を行う方法が一般的である。

RSTを用いた先行研究のなかに、記銘すべき単語(以下、ターゲット語と略す)の属性が実験参加者の成績に影響を及ぼすという報告がある。ターゲット語が単独呈示される場合よりも影響は小さいが、RSTにおいて音節数の多いターゲット語を用いると成績が低下する語長効果(word length effect)が報告されている(La Pointe & Engle, 1990)。また、Engle, Nations, & Cantor (1990)は、ターゲット語の出現頻度(frequency)の高低によって、結果に有意な差が認められることを示唆している。一方、RSTの刺激文の長さの違いによって、ターゲット語を保持しなければならない時間が異なることがRSTの成績に影響を及ぼすという刺激文順序効果(sentence order effect: Towse, Hitch, & Hutton, 2000)も報告されている。このように、RSTを構成するターゲット語の属性や刺激文の影響は実験実施にあたって無視できないことから、最近では実施にあたって刺激文とターゲット語の長さや難易度を統制したRSTが利用されている(Friedman & Miyake, 2005)。

日本語版RSTにおいても、ターゲット語の具体性や文中におけるターゲット語の位置が日本語版RSTの成績に影響を及ぼすことが示唆されている(近藤・苧阪, 2000)。しかし、現在多くの研究で利用されている苧阪(2002)による日本語版RSTにおいて、各試行の正答率、刺激文長、ターゲット語長等をもとにした分析は行われていない。日本では、苧阪(2002)のターゲット語や刺激文は多くの研究で利用されていることから、詳細な分析を行い基礎的なデータを提示しておくことは必要であると思われる。特に、RSTの実施に際しては、各文条件をランダムに実施する方法や、各文条件で5セット行うか3セット行うかといったセット数を調整する方法等も推奨されている(Conway et al., 2005; Friedman & Miyake, 2005)。したがって、標準的な方法で実施された苧阪(2002)の日本語版RSTの結果を十分に分析することが必要であると思われる。そこで本研究では、標準的な方法で実施された苧阪(2002)の日本語版RSTの結果について、各ターゲット語の正答率、ターゲット語の属性、刺激文とターゲット語の位置、刺激文の属性等をもとに分析を行う。

Friedman & Miyake (2005)は、Daneman & Carpenter (1980)らが開発したリーディングスパン・テストの得点化方法に関して、ターゲット語の総正再生数(total words)、正再生率(proportion words)、総

正答セット再生数(correct sets words), スパン得点(truncated span)という 4 つの得点化方法を先行研究より抽出し比較検討を行った。総正再生数は、すべての試行を通して正しく再生できたターゲット語数であり、最大 70 である。正再生率は、各セットの正答率を算出し全セット分の正答率の平均を算出したものである。例えば、4 文条件の 1 セット目の正再生数が 2 であった場合の正答率は 0.5、2 セット目の正再生数が 1 であった場合の正答率は 0.2 というようにセットごとに正答率を算出し、この正答率の平均を算出して正再生率とする。正再生率は、最大 1.00 となる。正答セット再生数は、完全に正答できた試行の再生数の合計である。例えば 3 文条件のある試行で正再生数が 3 の場合は 3 と得点化されるが、2 の場合は 0 と得点化され、最大 70 である。スパン得点は、各文条件で 5 セット実施する場合、正再生数が 3 以上の場合その文条件をクリアしたとしてその文条件数を得点とする方法である。正再生数が 2 の場合は、0.5 ポイントを付与する。例えば、3 文条件で正再生数が 4 であり、4 文条件で正再生数が 2 であった場合、スパン得点は 3.5 となる。

Friedman & Miyake (2005)は、これらの 4 つの得点化方法において、総再生数や再生率といったより連続的な測度が、他の測度に比べ信頼性があり正規性も認められたことを報告している。しかし、日本語版 RST(苧阪, 2002)に関して、このような得点化方法と信頼性についてはこれまでに十分に検討されていない。そこで本研究では Friedman & Miyake (2005)で比較検討された 4 つの得点化方法と信頼性に関して日本語版 RST についても検討を行う。

方法

実験参加者 長崎県立大学経済学部の大学生 148 名 (男性 93 名, 女性 55 名) が実験に参加した。平均年齢は 20.9 歳 (範囲: 18—25 歳) であった。

RST 日本語版 RST(苧阪, 2002)を利用した。

実験装置 刺激呈示に、パーソナル・コンピュータ(NEC MA10T)と 17 インチの液晶モニターを利用した。

手続き RST の測定は、苧阪・苧阪(1994)に準じ個別に行われ、2 文条件から 5 文条件までそれぞれ 5 試行が行われた。後日、148 名の実験参加者からランダムに選ばれた 51 名の実験参加者に対して、再度同じ手続きで RST の測定が行われた。

結果

日本語版 RST(苧阪, 2002)の刺激文とターゲット語 苧阪・苧阪(1994)における日本語版 RST では、刺激文は高等学校の教科書から、漢字仮名交じりで 20 文字から 30 文字の長さの文章が選ばれ、あらかじめ 7 段階の難易度評定を行い平均評定値 2.20—5.12 の範囲から 80 文が選ばれたものが利用されている。その後日本語版 RST は一部改定(苧阪, 1998)を経て、苧阪(2002)の日本語版 RST が現在公開されている。本研究で利用した日本語版 RST(苧阪, 2002)における刺激文とターゲット語に関する代表値を示したものが表 1 から表 6 である。

表 1

日本語版 RST における刺激文の文字数

文条件	<i>M</i>	<i>Max</i>	<i>Min</i>	<i>SD</i>
2	23.5	26.0	20.0	1.9
3	23.5	28.0	18.0	2.7
4	24.6	30.0	20.0	2.7
5	23.6	29.0	20.0	2.3
all	23.8	30.0	18.0	2.4

表 2

日本語版 RST における刺激文のモーラ数

文条件	<i>M</i>	<i>Max</i>	<i>Min</i>	<i>SD</i>
2	28.8	33.0	26.0	2.1
3	30.2	42.0	22.0	4.1
4	30.9	38.0	22.0	4.4
5	29.3	36.0	24.0	3.3
all	29.9	42.0	22.0	3.7

表 3

日本語版 RST におけるターゲット語の文字数

文条件	<i>M</i>	<i>Max</i>	<i>Min</i>	<i>SD</i>
2	2.4	4.0	1.0	1.0
3	2.5	6.0	1.0	1.2
4	2.6	5.0	1.0	0.9
5	2.8	6.0	1.0	1.1
all	2.6	6.0	1.0	1.1

表 4

日本語版 RST におけるターゲット語のモーラ数

文条件	<i>M</i>	<i>Max</i>	<i>Min</i>	<i>SD</i>
2	3.8	8.0	2.0	1.6
3	3.7	6.0	2.0	1.0
4	3.7	6.0	2.0	1.3
5	4.1	6.0	2.0	1.2
all	3.8	8.0	2.0	1.2

表 5

日本語版 RST における品詞ごとのターゲット語数

文条件	名詞	副詞	動詞
2	9.0	1.0	0.0
3	13.0	2.0	0.0
4	17.0	2.0	1.0
5	17.0	5.0	3.0
all	56.0	10.0	4.0

表 6

日本語版 RST における表記形態ごとのターゲット語数

文条件	漢字	平仮名	混
2	9.0	1.0	0.0
3	11.0	4.0	0.0
4	11.0	4.0	5.0
5	13.0	2.0	10.0
all	44.0	11.0	15.0

荻阪・荻阪(1994)の日本語版 RST と同様に荻阪(2002)の日本語版 RST においても、セット内の短文はできるだけ相互に意味的関連性をもたないようになっていた。また、ターゲット語もセット内では意味的に関連しないものが用いられていた。荻阪・荻阪(1994)の日本語版 RST では、漢字仮名交じりで 20 文字から 30 文字の長さが用いられていたが、荻阪(2002)版日本語版 RST では 18 文字から 30 文字の長さが用いられている(表 1)。刺激文とターゲット語のモーラ数について分析した結果を示したものが、表 2 と表 4 である。また、ターゲット語の品詞について分析したものが、表 5 である。荻阪・荻阪(1994)版日本語版 RST と同様に、ターゲット語には名詞、副詞、動詞が含まれて

いた。ターゲット語の表記形態について分析したものが、表 6 である。苧阪・苧阪(1994)版日本語版 RST では、平仮名表記のものが 10 語、漢字表記は 45 語、漢字仮名交じり表記が 15 語であったが、苧阪(2002)版日本語版 RST では平仮名表記が 11 語、漢字表記が 44 語、漢字仮名交じり表記が 15 語となっていた。日本語版 RST において、ターゲット語の文内での出現位置は文章ごとにランダムとなっている。刺激文におけるターゲット語の相対的出現位置を、ターゲット語の開始位置/刺激文の文字数という式によって求め、25%ごとにまとめたものが表 7 である。同様の分析を刺激文とターゲット語のモーラ数をもとにまとめたものが表 8 である。

表 7

日本語版 RST におけるターゲット語の出現位置[†]

文条件	<i>M</i>	<i>Max</i>	<i>Min</i>	<i>SD</i>	0~25%	26%~50%	51%~75%	76%~100%
2	8.9	21.0	1.0	6.2	3.0	4.0	2.0	1.0
3	12.7	21.0	1.0	6.3	2.0	4.0	6.0	3.0
4	10.8	22.0	1.0	5.8	3.0	10.0	5.0	2.0
5	12.5	24.0	1.0	6.7	5.0	6.0	8.0	6.0
all	11.5	24.0	1.0	6.3	13.0	24.0	21.0	12.0

[†] 刺激文におけるターゲット語の相対的出現位置を、ターゲット語の開始位置/刺激文の文字数という式によって求め、25%ごとにまとめたもの

表 8

日本語版 RST におけるターゲット語モーラの出現位置[†]

文条件	<i>M</i>	<i>Max</i>	<i>Min</i>	<i>SD</i>	0~25%	26%~50%	51%~75%	76%~100%
2	8.9	21.0	1.0	6.2	3.0	4.0	2.0	1.0
3	12.7	21.0	1.0	6.3	2.0	4.0	6.0	3.0
4	10.8	22.0	1.0	5.8	3.0	10.0	5.0	2.0
5	12.5	24.0	1.0	6.7	5.0	6.0	8.0	6.0
all	11.5	24.0	1.0	6.3	13.0	24.0	21.0	12.0

[†] 刺激文におけるターゲット語の相対的出現位置を、ターゲット語モーラの開始位置/刺激文のモーラ数という式によって求め、25%ごとにまとめたもの

各ターゲット語の正答率 各ターゲット語の正答率をまとめたものが、表 9 である。さらに、各文条件におけるターゲット語の各系列位置の再生率をまとめたものが、表 10 である。5 文条件における各ターゲット語の再生率について 1 要因分散分析を行った結果、主効果が有意であった($F(4, 16)=12.81, p<.001$)。多重比較(Bonferroni 法、以下特に記述のない場合はこの方法)の結果、5 試行目>4 試行目($p<.05$)、5 試行目>3 試行目($p<.01$)であった。4 文条件における各ターゲット語の再生率について 1 要因分散分析を行った結果、主効果が有意であった($F(3, 12)=14.26, p<.001$)。多重比較の結果、4 試行目>3 試行目($p<.05$)、4 試行目>2 試行目($p<.01$)、1 試行目>3 試行目($p<.05$)であった。3 文条件における各ターゲット語の平均値について 1 要因分散分析を行った結果、主効果が有意であった($F(2, 8)=4.87, p<.05$)。多重比較(LSD 検定)の結果、3 試行目>2 試行目($p<.05$)、3 試行目>1 試行目

($p<.05$)であった。2 文条件における各ターゲット語の平均値について t 検定を行った結果、有意差はなかった($t(4)=-1.40, ns$)。

表 9
各ターゲット語の正答率(%、 $N=148$)

文	セット	リスト	正答率	文	セット	リスト	正答率
2	1	1	97.0	4	3	3	34.0
		2	94.0			4	88.0
	2	1	100.0		4	1	74.0
		2	100.0			2	51.0
	3	1	68.0		4	3	49.0
		2	96.0			4	95.0
	4	1	82.0		5	1	70.0
		2	97.0			2	68.0
	5	1	97.0		3	3	56.0
		2	98.0			4	97.0
3	1	1	82.0	5	1	1	80.0
		2	61.0			2	62.0
		3	83.0			3	36.0
	2	1	39.0		2	4	69.0
		2	43.0			5	92.0
		3	95.0			1	38.0
	3	1	57.0		2	2	57.0
		2	89.0			3	57.0
		3	99.0			4	45.0
	4	1	84.0		3	5	92.0
		2	55.0			1	50.0
		3	100.0			2	39.0
	5	1	59.0		3	3	49.0
		2	90.0			4	43.0
		3	96.0			5	97.0
4	1	1	52.0	4	1	59.0	
		2	74.0		2	24.0	
		3	15.0		3	49.0	
		4	86.0		4	43.0	
		5	91.0		5	99.0	
	2	1	49.0	5	1	62.0	
		2	49.0		2	70.0	
		3	55.0		3	37.0	
		4	97.0		4	49.0	
	3	1	81.0	4	5	97.0	
		2	47.0		5	97.0	

表 10
各文条件におけるターゲット語の系列位置別再生率(%、 $N=148$)

系列位置	2		3		4		5	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1	88.8	13.6	64.2	18.9	73.6	14.5	57.8	15.5
2	97.0	2.2	67.6	21.0	57.8	12.3	50.4	18.6
3	-	-	94.6	6.8	41.8	17.4	45.6	8.9
4	-	-	-	-	92.6	5.2	49.8	11.0
5	-	-	-	-	-	-	95.4	3.2

ターゲット語長 ターゲット語長に関して、語長ごとの正答率の平均値を求めたものが表 11 である。正答率の平均値について 1 要因分散分析を行った結果、主効果が有意であった($F(5, 735)=44.70, p<.001$)。多重比較の結果、語長 1<語長 2, 語長 1<語長 4, 語長 1>語長 5, 語長 2>語長 3, 語長 2<語長 4, 語長 2>語長 5, 語長 3<語長 4, 語長 3>語長 5, 語長 4>語長 5, 語長 4>語長 6, 語長 5>語長 6 の平均値間に 5%水準で有意差があった。

ターゲット語のモーラ長 ターゲット語のモーラ長に関して、モーラ長ごとの正答率の平均値を示したものが表 12 である。モーラ長 8 のターゲット語を有する刺激文は 1 文のみであるため、モーラ長 2~6 までの正答率の平均値について 1 要因分散分析を行った結果、主効果が有意であった($F(4, 588)=27.38, p<.001$)。多重比較の結果、モーラ長 2>モーラ長 3, モーラ長 2>モーラ長 4, モーラ長 2>モーラ長 5, モーラ長 2>モーラ長 6, モーラ長 3>モーラ長 4, モーラ長 3<モーラ長 5, モーラ長 5>モーラ長 6 の平均値間に 5%水準で有意差があった。

表 11

ターゲット語長別の平均正再生率(%、 $N=148$)

語長	語数	<i>M</i>	<i>SD</i>
1	6	68.0	20.9
2	30	72.9	10.3
3	23	64.7	13.8
4	7	82.0	13.5
5	2	48.3	35.7
6	2	65.9	32.0

表 12

ターゲット語のモーラ数別の平均正再生率(%、 $N=148$)

モーラ数	語数	<i>M</i>	<i>SD</i>
2	9	77.0	16.3
3	20	65.0	13.9
4	24	69.8	10.9
5	10	71.6	13.6
6	6	66.9	18.9
8	1	100.0	0.0

ターゲット語の開始位置 ターゲット語が刺激文のどの位置から開始されるかについて、刺激文に対するターゲット語の相対的位置を求め、刺激文に対するターゲット語の相対的 4 分位位置ごとに分類し正答率を求めたものが表 13 である。各正答率の平均値について 1 要因分散分析を行った結果、主効果が有意であった($F(3, 441)=237.05, p<.001$)。多重比較の結果、5%水準ですべての平均値間に有意差がみられた。

ターゲット語モーラの開始位置 ターゲット語が刺激文のどの位置から開始されるかについて、刺激文に対するターゲット語の相対的位置を刺激文のモーラ数をもとに求め、刺激文に対するターゲット語の相対的 4 分位位置ごとに分類し正答率を求めたものが表 14 である。各正答率の平均値について 1 要因分散分析を行った結果、主効果が有意であった($F(3, 441)=89.04, p<.001$)。多重比較の結果、5%水準で有意差があったのは、 $Q1>Q2, Q1>Q4, Q2<Q3, Q2>Q4, Q3>Q4$ の平均値間であった。

ターゲット語の品詞 ターゲット語の品詞ごとの正答率の平均値を示したものが表 15 である。各正答率の平均値について 1 要因分散分析を行った結果、主効果が有意であった($F(2, 294)=79.62, p<.001$)。多重比較の結果、5%水準で有意差があったのは、名詞>動詞、副詞>動詞の平均値間であった。

表 13

ターゲット語の刺激文中における各開始位置と平均再生率(%、N=148)

開始位置*	語数	M	SD
Q1	13	77.0	12.9
Q2	24	65.4	13.2
Q3	21	80.3	10.1
Q4	12	52.4	18.9

*Q1 はターゲット語が刺激文頭より 25%以内で始まる。Q2 はターゲット語が刺激文頭より 25%～50%以内で始まる。Q3 はターゲット語が刺激文頭より 50%～75%以内で始まる。Q4 はターゲット語が刺激文頭より 75%～で始まる。

表 14

ターゲット語の刺激文モーラ数全体における各開始位置と平均再生率(%、N=148)

開始位置*	語数	M	SD
Q1	20	73.6	12.0
Q2	26	68.6	13.2
Q3	13	75.7	11.5
Q4	11	58.7	17.8

*Q1 はターゲット語が刺激文頭より刺激文モーラ数 25%以内で始まる。Q2 はターゲット語が刺激文頭より刺激文モーラ数 25%～50%以内で始まる。Q3 はターゲット語が刺激文頭より刺激文モーラ数 50%～75%以内で始まる。Q4 はターゲット語が刺激文頭より刺激文モーラ数 75%～で始まる。

表 15

ターゲット語の品詞と平均再生率(%、N=148)

品詞	語数	M	SD
名詞	56	71.3	10.5
副詞	10	68.9	16.0
動詞	4	50.3	27.7

表 16

ターゲット語の各表記形態と平均再生率(%、N=148)

表記形態	語数	M	SD
漢字	44	72.4	10.3
ひらがな	11	76.8	13.7
混合	15	56.9	18.5

ターゲット語の表記 ターゲット語の表記ごとの正答率の平均値を示したものが表 16 である。各正答率の平均値について 1 要因分散分析を行った結果、主効果が有意であった($F(2, 294)=142.87, p<.001$)。多重比較の結果、5%水準ですべての平均値間に有意差がみられた。

刺激文字数 刺激文の文字数をもっとも少ない文字数より 25%ごとに区切り、正答率を求めたものが表 17 である。各正答率の平均値について 1 要因分散分析を行った結果、主効果が有意であった($F(3, 441)=158.06, p<.001$)。多重比較の結果、5%水準で有意差があったのは、18～21<22～24、18～21<28～30、22～24>25～27、22～24<28～30、25～27<28～30 の平均値間であった。

表 17

各刺激文長と平均再生率(N=148)

文長(語)	刺激文数	M	SD
18～21	11	61.9	17.2
22～24	34	71.2	11.8
25～27	18	65.0	13.7
28～30	7	87.6	13.6

表 18

各刺激文のモーラ長の度数と平均再生率(N=148)

文長(語)	語数	M	SD
22～26	11	63.0	14.9
27～31	37	74.2	11.2
32～36	19	65.0	14.9
37～42	3	70.3	23.1

刺激文モーラ数 刺激文のモーラ数をもっとも少ないモーラ数より 25%ごとに区切り、正答率を求めたものが表 18 である。各正答率の平均値について 1 要因分散分析を行った結果、主効果が有意であった($F(3, 441)=18.50, p<.001$)。多重比較の結果、5%水準で有意差があったのは、22~26<27~31, 22~26<37~42, 27~31>32~36, の平均値間であった。

得点化方法 日本語版 RST における各得点化方法による成績をまとめたものが、表 19 である。尖度と歪度による正規性の検定を行った結果、総正再生数と正再生率に関して有意ではなかったが、総正答セット再生数とスパン得点については有意であり正規性が認められなかった。また、Shapiro-Wilks test による正規性の検定を行った結果、総正答セット再生数(Shapiro-Wilks $W=.930, p<.001$)、スパン得点(Shapiro-Wilks $W=.805, p<.001$)の Shapiro-Wilks test 統計量は有意であった。一方、総正再生数(Shapiro-Wilks $W=.986, p=.154$)、正再生率(Shapiro-Wilks $W=.991, p=.446$)の Shapiro-Wilks test 統計量は有意ではなかった。すなわち、総正答セット再生数、スパン得点は正規性が認められず、総正再生数、正再生率には正規分布にしたがうことが示された。

表 19

日本語リーディングスパンにおける各得点化法の統計値

得点化法	<i>M</i>	<i>SD</i>	範囲	歪度 [†]	尖度 [‡]
総正再生数	48.85	7.53	30-67	0.15	-0.45
正再生率	0.74	0.10	.49-.97	0.06	-0.41
総正答セット再生数	22.70	10.82	6-58	1.00*	0.96*
スパン得点	2.62	0.68	2-5	1.47*	2.14*

[†]標準誤差=0.20, [‡]標準誤差=0.40, * $p<.05$

日本語版 RST における各得点化方法間の相関係数を示したものが、表 20 である。また、日本語版 RST における各得点化法の信頼性について、各文条件の正答数をもとに求めた α 係数と再テストの得点との相関を示したものが表 21 である。

表 20

日本語リーディングスパンにおける
得点法間の相関係数[†]

得点化法	1	2	3	4
1.総正再生数	-			
2.正再生率	.99	-		
3.総正答セット再生数	.88	.88	-	
4.スパン得点	.50	.55	.56	-

[†]すべて 1%水準で有意

表 21

日本語リーディングスパンにおける
各得点化法の信頼性

得点化法	α [†]	Test-Retest [‡]
総正再生数	.66	.67
正再生率	.71	.69
総正答セット再生数	.58	.63
スパン得点	—	.48

[†] $N=148,$ [‡] $N=51$

考察

本研究の目的の一つは、ワーキングメモリスパン課題として日本で普及している芋阪(2002)の日本語版 RST の刺激文とターゲット語を検討することであった。分析の前提として、各ターゲット語の正答率について系列位置に関する 1 要因分散分析を行った結果、初頭効果、親近性効果といった系列位置効果が 2 文条件をのぞく文条件でみられた。

日本語版 RST のターゲット語を検討した結果、語長、モーラ長、品詞、表記形態といった属性ごとに分類可能であった。また、各属性に関する正答率の 1 要因分散分析の結果、すべての属性で主効果が有意であった。ただし、特定の属性以外は統制されておらず複数の要因が交絡している。したがって、ターゲット語の各属性が正答率にどのような影響を及ぼすかについては、厳密に他の要因を統制した実験が必要である。近藤・芋阪(2000)は、ターゲット語の単語の具体性という属性に焦点をあて、その効果を検討した。彼らの研究では、ターゲット語の表記形態は平仮名表記の場合のスパン得点と漢字表記の場合のスパン得点の間に有意差はなかったが、高具体性のターゲット語を用いた場合のスパン得点は、具体性の低いターゲット語を用いた場合のスパン得点よりも有意に高かった。彼らは、この結果から日本語版 RST の成績の個人差の要因として、ターゲット語を保持する時の実験参加者の用いる手がかりの多様さがあることを指摘している。彼らが指摘しているように、ターゲット語の属性は、ターゲット語を保持する際に実験参加者が利用する手がかりに影響を及ぼす要因と考えられる。したがって、日本語版 RST を実施するには、できるだけターゲット語の属性を統制することが必要であると思われる。Conway et al. (2005)は、刺激文中の単語をターゲット語とする場合、実験参加者による刺激文の要約化といった読解能力の個人差の影響を受ける可能性があることから、刺激文を読んだ後全く関連のない単語や任意の文字列をターゲット語とする方法が望ましいことを指摘している。日本語版 RST においても、このようなターゲット語の統制が必要であろう。

日本語版 RST では、刺激文中におけるターゲット語の出現位置もランダムに配置されている。RST では、ターゲット語の位置と刺激文の長さは、ワーキングメモリにおける処理と保持との関連で重要な影響を及ぼすことが明らかになってきた(Towse et al., 2000)。本研究における分析においても、複数の要因が交絡しているものの、刺激文中におけるターゲット語の出現位置ごとの正答率に関する 1 要因分散分析の結果、主効果がみられた。このような RST における刺激文とターゲット語の位置関係に関する分析を考慮すると、ターゲット語の刺激文中の位置も統制することが望ましいであろう。例えば、Saito & Miyake (2004)では、日本語版 RST においても英語版 RST と同様に刺激文の最後部にターゲット語を配置する方法で実施されている。

Friedman & Miyake (2004)では、刺激文の長さや難易度が RST における処理の量と密接に関連することから、刺激文の長さや難易度を統制した RST が利用されている。本研究においても、複数の要因が交絡しているものの日本語版 RST の刺激文長及び刺激文のモーラ長ごとの正答率に関する 1 要因分散分析の結果、主効果がみられた。日本語版 RST においては、開発時に刺激文の難易度と長さはある程度統制されているものの、十分でない可能性も考えられる。したがって、日本語版 RST に

においても、刺激文の難易度をより一定にしかつ刺激文の長さの範囲もより狭くする必要性があるであろう。

Friedman & Miyake (2005)は、RST を実施する際に文条件を 2 文条件から 5 文条件へと漸次増加させる方法ではなく、ランダムに実施する方法が実験参加者に対して負荷をかけず望ましいとして推奨している。本研究における日本語版 RST の刺激文やターゲット語の検討によって、日本語版 RST を実施するうえでターゲット語や刺激文の属性を統制したり、各文条件をランダムに行う際の各セットを検討したり、刺激文を取捨選択するうえでの基礎的データを提供することができたと考えられる。

荻阪(2002)の日本語版 RST の得点化方法と信頼性について検討することも、本研究の目的の一つであった。得点化方法として、Friedman & Miyake (2005)と同様に 4 つの得点化方法について検討した。その結果、尖度、歪度及び正規性の検定において、総正再生数と正再生率という測度では正規性は保証されたが、総正答セット再生数とスパン得点については正規性が保証されなかった。この結果は、Friedman & Miyake(2005)による分析結果と同じ結果であった。日本語版 RST においても、総正再生数もしくは正再生率という得点化方法を利用することが望まれる。

再検査法による再検査信頼性係数(Test-Retest reliability)を算出した結果、総正再生数(.67)、正再生率(.69)は、総正答セット再生数(.63)、スパン得点(.48)よりも信頼性があった。内的整合性の指標である α 係数も、総正再生数(.66)、正再生率(.71)は、総正答セット再生数(.58)よりも信頼性があった。この結果は、Friedman & Miyake (2005)による 3 セット版の RST の再検査信頼性係数(総正再生数:.72、正再生率:.72、総正答セット再生数:.59、スパン得点:.52)とほぼ同様であった。しかし、Friedman & Miyake (2005)による 5 セット版の RST の再検査信頼性係数(総正再生数:.80、正再生率:.79、総正答セット再生数:.74、スパン得点:.68)に比べ低い信頼性であったと考えられる。Klein & Fiss (1999)の研究では、演算スパンテストにおける再検査信頼性係数が.67～.81 の範囲であった。また、Tirre & Peña (1992)の研究では、RST における総再生数による再検査信頼性が.73 であった。このような先行研究で報告されている英語版の RST や演算スパンテストの再検査信頼性係数と比べ、本研究で得られた日本語版 RST の再検査信頼性係数も、Friedman & Miyake (2005)による 5 セット版を除けばほぼ同等と考えられ、信頼性があると判断できる。Friedman & Miyake (2005)による 5 セット版では、刺激文の長さやターゲット語の難易度が統制されている。一方、本研究で用いた日本語版 RST はそのような統制は行っておらず刺激文やターゲット語の様々な属性が影響していたと思われる。この点が、Friedman & Miyake (2005)による 5 セット版の再検査信頼性係数と本研究における日本語版 RST の再検査信頼性係数の差異となっていると推測できる。

Waters & Caplan (1996)の研究においてもスパン得点による再検査信頼性は.48 ($N=44$)であり、スパン得点の再検査信頼性は他の得点化法に比べて低いと考えられる。スパン得点によってスパンが 4.0 以上の実験参加者(高スパン群)、3.0 及び 3.5 の実験参加者(中スパン群)、2.0 及び 2.5 の実験参加者(低スパン群)に分類する方法はもっとも一般的である(Caplan & Waters, 1999)。本研究においても、この分類方法によって 1 回目の RST のスパン得点と 2 回目のスパン得点をもとに分類した結果、25 名(49%)は分類される群に変化はなかったが、41%の実験参加者が 1 回目と 2 回目の間で属するスパン

群に変更があった。19名(37%)の実験参加者は2回目が高いスパン群に分類され7名の実験参加者(14%)は2回目が低いスパン群に分類された。Waters & Caplan (1996)の研究(N=44)においても、41%の実験参加者は、1回目と2回目のスパン群に変更がおきたことが報告されている。スパン得点のみで実験参加者の分類を行う際には注意が必要である。

ワーキングメモリに関する認知心理学的研究法のなかでも、GP分析法は、ワーキングメモリスパン課題の成績をワーキングメモリ容量の指標として利用し、ワーキングメモリの処理資源容量の大きい人と小さい人といったワーキングメモリ容量の個人差が、ある特定の認知課題でどのような差を示すかを分析し、ワーキングメモリ容量の限界時における認知活動の制約について検討を進める手法である(三宅・齊藤, 2001)。このような手法を利用する場合は、スパン得点だけでなく他の得点化方法による得点や上位25%下位25%を抽出する方法、さらに複数のワーキングメモリスパン課題を用いることが推奨されている(Conway et al., 2005)。本研究結果からもこのような手法をとることが推奨される。

本研究における実験参加者はすべて、社会科学部系の学部生であり、成人を対象とした日本語版RSTの実験参加者の信頼性を検討するには偏った実験参加者と考えられる。今後、成人の各年齢層から抽出した実験参加者を対象とした検討が必要であろう。日本語版RSTに影響を及ぼす要因について十分に統制し、また他の日本語版のワーキングメモリスパン課題との関連等も検討し、より信頼性の高い日本語版RSTを利用していくことが望まれる。

引用文献

- Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, **22**, 77-126.
- Case, R., Kurland, M., & Goldberg, J. (1982). Operational efficiency and growth of short term memory span. *Journal of Experimental Child Psychology*, **33**, 386-404.
- 中條和光・中尾美月 (2005). 作業記憶容量と補償的文章読解方略の使用との関係—眼球運動を指標として— *読書科学*, **49**, 41-52.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review*, **12**, 769-786.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, **19**, 450-466.
- Engle, R. W., Nations, J. K., & Cantor, J. (1990). Is "working memory" capacity just another name for word knowledge? *Journal of Educational Psychology*, **82**, 799-804.
- Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, **51**, 136-158.
- Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test.

- Behavior Research Methods*, **37**, 581-590.
- Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic reasoning tasks. *Memory & Cognition*, **21**, 115-124.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, **99**, 122-149.
- Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments, & Computers*, **31**, 429-432.
- 近藤洋史・森下正修・苅阪直行 (1999). 読みのワーキングメモリとリーディングスパンテスト 心理学評論, **42**, 506-523.
- 近藤洋史・苅阪直行 (2000). 読みのワーキングメモリに及ぼす単語の具体性の効果—日本語リーディングスパンテストによる検討— 心理学研究, **71**, 51-56.
- La Pointe, L. B. & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **16**, 1118-1133.
- Miyake, A. (2001). Individual differences in working memory: Introduction to the special section. *Journal of Experimental Psychology: General*, **130**, 163-168.
- Miyake, A., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language*, **33**, 175-202.
- 三宅 晶・齊藤 智 (2001). 作動記憶研究の現状と展開 心理学研究, **72**, 336-350.
- 西崎友規子・苅阪満里子 (2004). 文章理解とワーキングメモリの個人差 心理学研究, **75**, 220-228.
- 苅阪満里子 (1998). ワーキングメモリと言語理解の脳内メカニズム 心理学評論, **41**, 174-193.
- 苅阪満里子 (2002). 脳のメモ帳 ワーキングメモリ 新曜社
- 苅阪満里子・苅阪直行 (1994). 読みとワーキングメモリ容量—日本語版リーディングスパンテストによる測定— 心理学研究, **65**, 339-345.
- 大塚一徳 (2003). 問題解決におけるリーディングスパン個人差の影響 心理学研究, **74**, 460-465.
- 齊藤 智・三宅 晶 (2000). リーディングスパン・テストをめぐる6つの仮説の比較検討 心理学評論, **43**, 387-410.
- Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing-storage relationship in reading span performance. *Journal of Memory and Language*, **50**, 425-443.
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An Individual differences approach. *Journal of Experimental Psychology: General*, **125**, 4-27.
- Tirre, W. C., & Peña, C. M. (1992). Investigation of functional working memory in the reading span test. *Journal of Educational Psychology*, **84**, 462-472.
- Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition*, **28**, 341-348.

- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, **28**, 127-154.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, **37**, 498-505.
- Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: critique of Just and Carpenter(1992). *Psychological Review*, **130**, 761-772.
- Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers*, **35**, 550-564.