Ann. Bull. RILP, No. 9, 171-192 (1975)

PERCEPTION OF SYNTACTIC STRUCTURE IN JAPANESE*

Tazuko Uyeno and S. I. Harada

The study of linguistic performance in the context of current developments in linguistic theory has raised a number of important problems and produced quite a few penetrating insights into the ways in which human beings actually use their own language. Sentence comprehension has been one of those areas in psycholinguistics in which much intensive research has been done in recent years. Most of the past studies have, however, been concerned only with Indo-European languages, in particular, the English language. The conclusions drawn from such studies are thus often quite tentative and controversial if considered in a broader perspective. We can, for instance, conceive of a general theory of performance which would specify language-independent principles of linguistic performance, but in order to formulate such a general theory we need much more information about performance from individual languages than has hitherto been available. The purpose of our present study, to be reported herein, was to test the validity of the major hypotheses on sentence comprehension that are found in the psycholinguistic literature against performance data obtained from speakers of the Japanese language, a language typologically dissimilar to Indo-European languages.

1. Background

Within the framework of transformationally oriented psycholinguistics three models have been proposed to account for syntactic perception, the initial step in the hearer's understanding of sentences. The earliest model, which we shall call the "analysis-by-analysis" (AA) model (after Fodor, Bever and Garrett, 1974), hypothesizes that the hearer makes direct use of his internalized generative grammar. According to this model, the hearer, after detecting the surface structure of the utterance he hears, goes through the grammatical derivation of that sentence backwards to arrive at its deep structure (or possibly its semantic representation). The second model, usually referred to as the "analysis-by-synthesis" (AS) model, considers syntactic perception to be essentially a matching process. The hearer is supposed to perform a preliminary analysis on the input utterance and yield a trial pair of surface and deep (or, possibly, semantic) structures. He independently activates his internalized generative grammar and generates a sentence that he supposes to correctly represent the input utterance. The AS model of syntactic perception contains a matching device that compares the results of preliminary analysis and internal generation, and if they show a satisfactory level of resemblance, the structural description of the sentence internally generated will be accepted by the hearer as the perceptual representation of the input utterance.

* This research was supported by the Grant in Aid for Scientific Research No. 920194, Ministry of Education.

The third model, which we shall call the "perceptual strategy" (PS) model claims, on the contrary, that the hearer makes no direct use of the internalized generative grammar in sentence perception.* According to this model the hearer instead appeals to a device exclusively intended for syntactic perception, called "perceptual strategies". Perceptual strategies are processes through which the diagnostic properties of surface structure are directly associated with pieces of semantic or functional information, without any reference to the intermediate stages of grammatical derivation.

In our study we assumed the PS model from the outset, on grounds that recent psycholinguistic studies have yielded results incompatible with the former two models. An extensive discussion can be found in Fodor, Bever, and Garrett (1974), and we will present just a brief description of the status quo. The main problem with the AA and the AS models is the assumption that the hearer follows every step in the grammatical derivation, either forwards or backwards. They, therefore, predict that transformationally more complicated sentences are in general harder to perceive than transformationally less complicated ones. In the earlier period of psycholinguistic research, this prediction seemed to be borne out. Miller, McKean, and Slobin, for instance, conducted a series of verification experiments and gathered data on reaction time.** They found that simple affirmative active declarative sentences were verified in shorter time than simple negative active declaratives or simple affirmative passive declaratives, which in turn were verified in shorter time than negative passive declaratives. Assuming the version of grammatical theory proposed in Chomsky's Syntactic Structures (1957), they claimed that these experimental results were in perfect accordance with the AA model, since in that theory, both negative and passive sentences were derived transformationally from the corresponding affirmative and active versions, respectively. Later developments in psycholinguistics and theoretical linguistics have made their conclusion untenable, however. On one hand, later versions of grammatical theory attributed the differences between active and passive sentences or between affirmative and negative sentences not merely to differences in transformational derivation but also (or rather) to differences in underlying structure. On the other hand, later psycholinguistic experiments have revealed cases in which those sentences that are transformationally more complicated in later versions of grammatical theory are verified in a shorter time than are transformationally simpler ones.

The PS model, which we have chosen as our theoretical basis, seems to be free from these difficulties. This model, divorced from generative grammar, is immune to the revision of grammatical theory and allows us to investigate the problem of perception in its own right. It needs to be said here, however, that we do not mean to imply that generative grammar has no psychological reality. The psychological reality of generative grammar is beyond any doubt, but it is simply irrelevant at the level of perception.

---

* The notion "perceptual strategy" was first proposed by Bever (1970), and is now extensively discussed in Fodor, Bever, and Garrett (1974).

** Reported in Miller (1962).

2. Experimental Design

In the fall of 1974, we conducted two series of experiments on syntactic perception. These experiments consisted basically of repetition tasks, but subsidiary tasks such as addition or recall of a random sequence of numbers were inserted as distractors.

Figure 1 represents the model of the performance of our Ss. Our assumption was that the S would first segment the stimulus sentence into a string of words and, after a short lag, establish functional and/or semantic relations between the segmented words. It thus seemed quite conceivable that without distractors the S would be forced to respond when the establishment of functional-semantic relations had not been completed, which would cause him to refer just to the string of segmented words in repeating the stimulus sequence. Moreover, it was very likely that the memory of the initial segmentation into words would strongly affect the reaction if the S were to respond immediately after the stimulus had been presented to him. On the other hand, too much delay in reaction was also undesirable since the S might go on to further steps in sentence comprehension beyond mere perceptual analysis in the meantime. The distractors were therefore inserted for two purposes: (1) to minimize the effect of the memory of word segmentation, and (2) to prevent the S from going beyond perceptual analysis.

In the remainder of this section we shall discuss the constructions tested in our experiments and present the actual stimuli employed. Of the ten constructions originally included in our experiments, Cleft Formation and Predicate Raising are excluded in the present discussion, since the responses to the stimuli of these constructions showed too broad a variety for us to arrive at a solid interpretation.

Passivization

This is a cognate of the corresponding well-known rule in English. Its effect is to convert a transitive sentence into a structure in which the original object is made the derived subject and the original subject is made an agent phrase through addition of the particle ni. * The actual stimuli in our experiments were as follows:

(1)a. (Active) Ueyama-san no otooto wa Tanaka-san ni hikooki no anzen-na
title brother airplane safe
soozyuu-hoo o osei-masi-ta.
pilot way teach

'Mr. Ueyama's brother has taught Mr. Tanaka safe pilotage.'

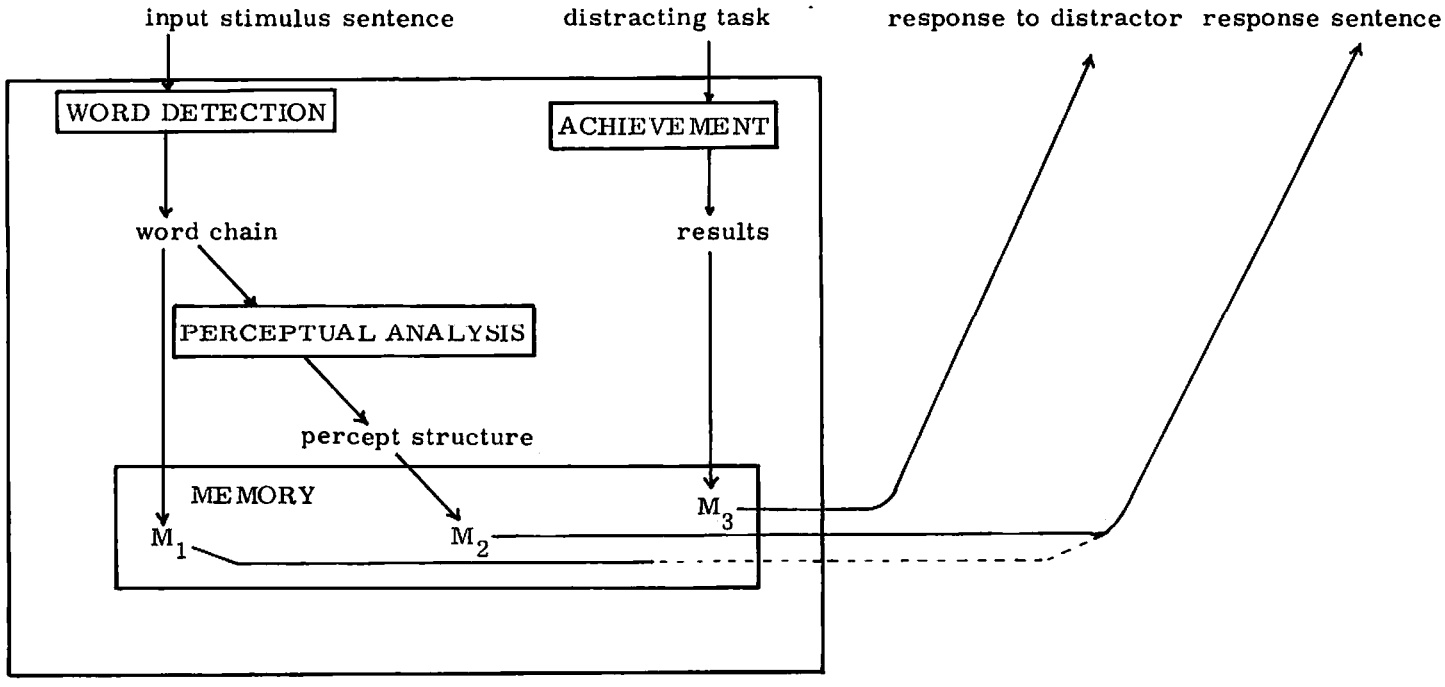
b. (Passive) Tanaka-san wa Ueyama-san no otooto ni hikooki no anzen-na soozyuu-hoo o osie-rare-masi-ta.

'Mr. Tanaka was taught safe pilotage by Mr. Ueyama's brother.'

There was another pair in which the NPs "Ueyama-san no otooto" and "Tanaka-san" were interchanged. The same holds for all the other cases of reversible pairs.

*We follow Kuno's (1973) analysis here. See Howard and Niyekawa-Howard (in press) for an alternative conception of Japanese passives.

Figure 1. Model of performance of the S in our experiments



(2)a. (Active) Mitaka-si no koomuin ga Oota-ku no zyosi-ginkooin o
City public servant Ward woman bank-clerk

kon'yaku-hurikoo de uttae-masi-ta.
engagement not-carrying-out accuse

'A public servant in Mitaka City has accused a woman bank clerk in Ota Ward of insincere engagement.'

b. (Passive) Oota-ku no zyosi-ginkooin ga Mitaka-si no koomuin ni

kon'yaku-hurikoo de uttae-rare-masi-ta.

'A woman bank clerk in Ota Ward has been accused of insincere engagement by a public servant in Mitaka City.'

The following is a non-reversible pair:

(3)(a. (Active) Misuborasi-i hukusoo no otoko ga dooro ni oti-te i-ta
poor outfit man road drop

saihu o hirot-ta.
purse pick up

'A man in a shabby outfit picked up a purse left on the street.'

b. (Passive) Dooro ni oti-te i-ta saihu ga misuborasi-i hukusoo no

otoko ni hirow-are-ta.

'A purse left on the street was picked up by a man in a shabby outfit.'

Subject Raising

This transformation was proposed by Kuno (1972) to account for the relation between the following pair of sentence types.

(4) a. (nonSR) Kono sinzin-kasyu wa sono beteran-haiyuu ga subarasi-i
this fresh singer that veteran actor wonderful

tarento da to omot-ta.
talent be think

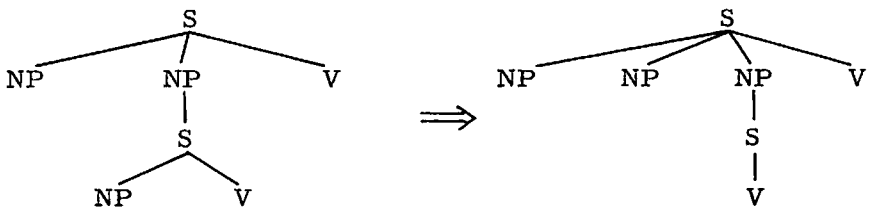
'This new singer thought that that veteran actor was a wonderful talent.'

b. (SR) Kono sinzin-kasyu wa sono beteran-haiyuu o subarasi-i tarento

da to omot-ta.

'This new singer thought that veteran actor to be a wonderful talent.'

A conspicuous difference in surface structure is observed in the choice of the particles ga vs. o, but, as Kuno has argued convincingly, there is in fact a difference in phrase structure. The effect of Subject Raising can be represented diagrammatically in the following way.



That is, the complement subject in underlying structure is raised into the main clause as the direct object. The other sentences used in our experiments are as follow:

(5)a. (nonSR) Sono takusii no untensyu wa Otyanomizu kara not-ta gakusei
taxi driver from ride student

ga hannin da to iihat-ta.
culprit insist

'The taxi driver insisted that the student from Ochanomizu was the culprit.'

b. (SR) Sono takusii no untensyu wa Otyanomizu kara not-ta gakusei o

hannin da to iihat-ta.

Subjectivization

This is a rule proposed by Kuno (1973). This rule extracts a possessive phrase in the subject and makes it a derived subject. The resulting surface structure thus has a double subject.

(6) a. (X-no-Y-ga) Yosioka-san wa kyoo made Sasaki-san no okusan ga
today until wife

eigo no sensei da to sinzi-te i-masi-ta.
English teacher believe

'Until today, Mr. Yoshioka had believed that Mrs. Sasaki was an English teacher.'

b. (X-ga-Y-ga) Yosioka-san was kyoo made Sasaki-san wa okusan ga

eigo no sensei da to sinzi-te i-masi-ta.

Conjunct Movement

Conjunct Movement is a rule that moves one member of a coordinated NP out of it and converts it to a complement of the verb (cf. Kuno 1973).

(7) a. (X-to-Y-to) Hannin to kono haiyuu to wa kaodati ga taihen yoku
culprit features very well

ni-te i-ru to omo-u.
resemble

'I think that the culprit and this actor have highly similar facial features.'

b. (X-ga-Y-to) Hannin wa kono haiyuu to kaodati ga taihen yoku ni-te

i-ru to omo-u.

'I think that the culprit is very similar to this actor in facial features.'

Quantifier Extraction

Quantifier Extraction has been investigated by Okutsu (1969) and Kamio (1973). This rule is similar to Subjectivization in that it extracts a prenominal phrase from a NP, but the element to be extracted is a quantifier phrase (e. g. san-nin 'three people') and the movement is rightwards.

(8) a. (Qx) Hutari no yakuza ga sakuya Sibuya de gakusei o meta-meta
two people last night heavily

ni nagut-ta.
clobber

'Two rogues beat up a student at Shibuya last night.'

b. (xQ) Yakuza ga hutari sakuya Sibuya de gakusei o meta-meta ni nagut-

ta.

Scrambling

This is a rule that reorders the preverbal constituents. In our experiments, we paid special attention to the position of adverbials. In our first experiment, the focus was put on the positions of the subject and an adverbial clause.

(9) a. (S-ADV) Imooto wa Satoo-san ga byooki ni nat-ta no de kaisya o
sister sick because company

mudan de yasumi-masi-ta.
without notice absent.

'My sister absented herself from the company without notice because Mr. Sato was sick.'

b. (ADV-S) Satoo-san ga byooki ni nat-ta no de imooto wa kaisya o mudan de yasumi-masi-ta.

Note that the former sentence (type S-ADV) seems to reflect the underlying order of constituents. Scrambling is the central problem to which our second experiment was addressed. The actual sentences tested are as follows:

(10) (SV) Omawari-san ga arui-te imasu.
policeman walk 'A policeman is walking.'

(11) a. (SOV) Minna wa biiru o nomi-masi-ta.
everyone beer drink 'Everybody drank beer.'

b. (OSV) Biiru o minna wa nomi-masi-ta.

(12) a. (SOiOdV) Katyoo wa onna-no-ko ni syorui o watasi-masi-ta.
dept. chief girl papers hand

'The department chief handed the girl the papers.'

b. (OiSOdV) Onna-no-ko ni katyoo wa syorui o watasi-masi-ta.

(13) a. (SOAV) Sensei wa eiga o kinoo mi-masi-ta.
movie yesterday see

'The professor went to see the movie yesterday.'

b. (AOSV) Kinoo eiga o sensei wa mi-masi-ta.

(14) a. (SOAV) Ozisan wa ie de tegami o kaki-masi-ta.
uncle house letter write

'Uncle wrote a letter at home.'

(15) a. (ASOV) Isoide imooto wa yoohuku o simai-masi-ta.
hurriedly clothes put away

'My sister put the clothes away hurriedly.'

b. (OASV) Yoohuku o isoide imooto wa simai-masi-ta.'

Ga-No Conversion

This rule was proposed by Harada (1971). It converts the case particle ga to no in prenominal clauses. Unlike Subject Raising or Subjectivization, this rule does not affect the phrase structure composition.

(16) a. (ga) Noguti-kun wa Ueda-san ga syookai-si-ta aite to kekkon-suru
title introduce marry

tumori rasi-i.
intend seem

'Mr. Noguchi appears to be planning to marry the girl Mr. Ueda has introduced to him.'

b. (no) Noguti-kun wa Ueda-san no syookai-si-ta aite to kekkon-suru

tumori rasi-i.

Stative Object Marking

Stative predicates in Japanese take ga-marked objects sometimes obligatorily and sometimes optionally. Derived stative predicates like potentials or desideratives optionally take ga-marked objects. See Kuno (1973) for further details.

(17) a. (Do) Kyuuryoo-bi na no de, Nakamura wa hisasiburi ni koibito to
pay day love

eiga o mi-ta-i to omot-ta.
want

'Nakaura felt like going to the movie with his girlfriend, because it was his payday.'

b. (Dga) Kyuuryoo-bi na no de, Nakamura wa hisasiburi ni koibito to

eiga ga mi-ta-i to omot-ta.

(18) a. (X-ga-Y-o) Karada no yowa-i Saburoo ga kono byooki o kokuhuku-
body weak disease overcome

deki-ru ka doo ka taihen kigakari-desu.
whether anxious

'I'm very anxious about whether Saburo, who is physically weak, can overcome this disease.'

b. (X-ga-Y-ga) Karada no yowa-i Saburoo ga kono byooki ga kokuhuku-

deki-ru ka doo ka taihen kigakari-desu.

There is an independent rule that converts a ga-phrase into a ni-phrase if it is followed by another ga-phrase. The following two sentences are therefore included in the stimuli tested.

(19) a. (X-ni-Y-ga) Karada no yowa-i Saburoo ni kono byooki ga kokuhuku-

deki-ru ka doo ka taihen kigakari-desu.

b. (X-ni-Y-o) *Karada no yowa-i Saburoo ni kono byooki o kokuhuku-

deki-ru ka doo ka taihen kigakari-desu.

Notice that the latter sentence is ungrammatical.

3. Description of the Experiments

3.1. Experiment I

Material

Ten types of grammatical constructions in Japanese, each involving a transformation widely accepted in the literature of the transformational analysis of Japanese, were chosen and for each construction a set of stimuli was constructed. The ten constructions tested are discussed in the last section.

Reversibility of a pair of NPs was taken into consideration, and whenever possible we constructed two sets of stimuli, one consisting of reversible sentences and the other of nonreversible sentences. Moreover, for each set, two kinds of lexical composition were selected. The minimal contrastive set consisted of a sentence derived without application of the relevant transformation and one or more sentences derived through the application of the transformation. Thus, for example, we constructed 12 stimuli for Subject Raising: two pairs of nonreversible sentences and four pairs of reversible ones, each pair consisting of the same set of lexical items. The total number of stimuli was 84. Each stimulus was composed of 7 lexical morphemes and 5-8 grammatical morphemes.

The entire set of stimuli was broken up into two subsets of 42 sentences and these were randomized and recorded on separate tapes by one of the authors. After each stimulus sentence was recorded a subsidiary task, which in this experiment was an addition of two two-digit numbers (e.g. 27 + 59) not exceeding 100 in total.

Subjects

The Ss were 46 adults, 25 male and 21 female. Of these, 20 male and 17 female Ss were students at the University of Tokyo. The others were staff members of the Research Institute of Logopedics and Phoniatrics, at the University of Tokyo. The average age was 25 for male and 22 for female Ss.

Procedure

The test was administered to all Ss individually. The two tape-recorded sets of stimuli were tested in separate sessions at an interval of one week. The stimuli were given to each S binaurally through headphones. Ss were first given instructions and then went through five pre-test practice sentences. Ss were instructed to give the answer to the calculation task first and then repeat the given sentence. The Ss' responses were tape-recorded and were later transcribed for analysis.

In order to check the effect of session order, 10 male and 10 female

Ss were given the two stimulus sets in the order opposite to that used with the other Ss.

Scoring

The transcribed responses were analyzed along several dimensions and scored either as "correct" or "incorrect" on each dimension. For example, if the sentence

Hanako ga Taroo ni nagur-are-ta.
hit passive past

'Hanako was hit by Taro.'

was observed as a response to a stimulus sentence

Taroo ga Hanako ni nagur-are-masi-ta.

'Taro was hit by Hanako. [polite style]'

then the response was scored as "syntactically correct", "semantically incorrect", and so on. Of the dimensions taken into consideration we regarded the following two as the most important:

(1) preservation of the core syntactic structure

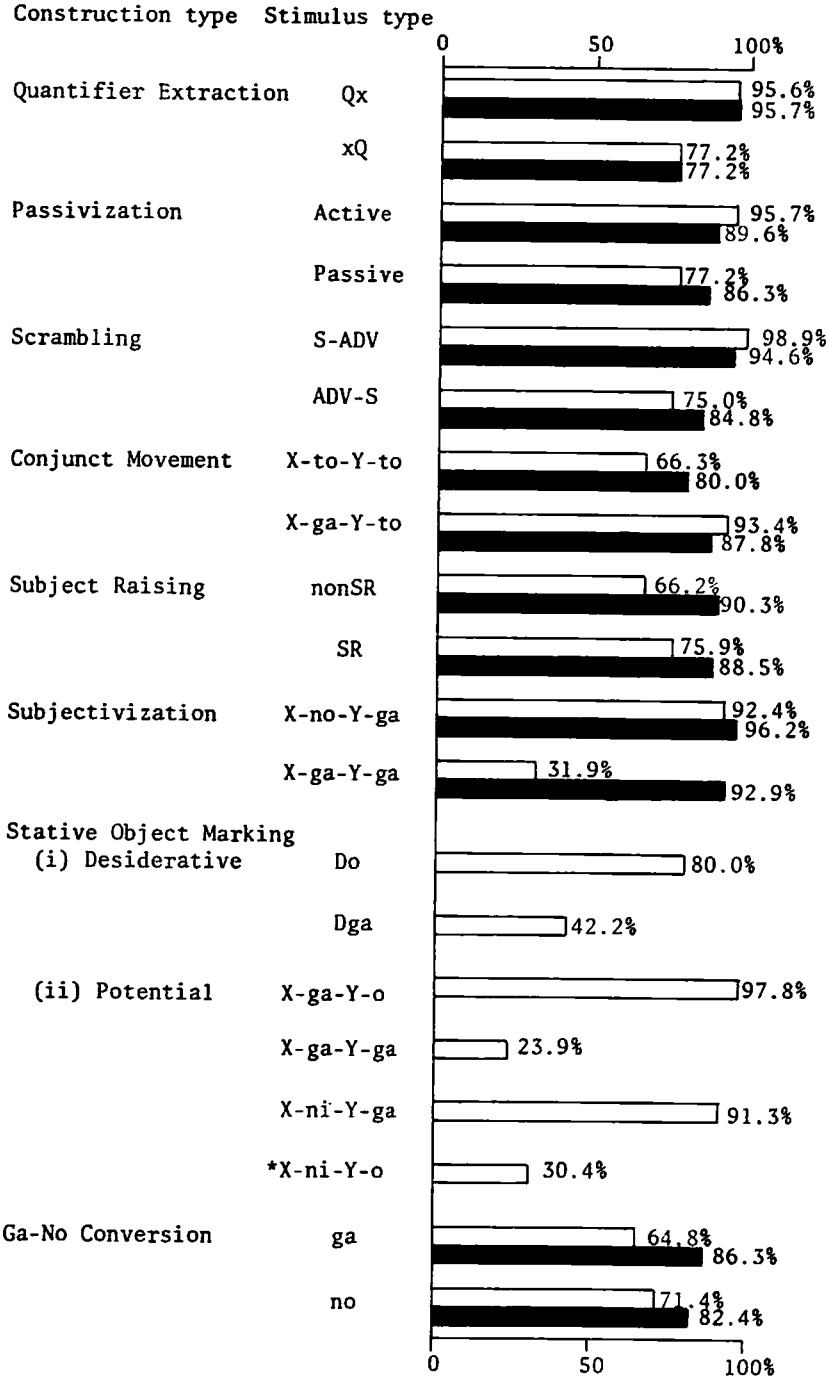
(2) preservation of the basic semantic relations

By the term "core syntactic structure" we mean the syntactic relations among the major constituents of a sentence, such as the subject, object, agent phrase, main predicate, etc. Adverbials or prenominal modifiers were ignored in scoring unless, as in the case of Quantifier Extraction, such elements were the focus of the syntactic contrast being tested. Similarly, we focused our attention on the semantic relations among these major constituents of a sentence in the semantic scoring. Thus, for instance, when the time adverb "kinoo" (yesterday) was incorrectly reproduced as "ototoi" (day before yesterday), we did not count this as a semantic error. In the case of the major constituents, however, errors in the choice of particular lexical items were disregarded only when the lexical item in the response was considered synonymous with the original one.

Proper names caused a surprisingly large number of reproduction errors, and we were forced to set up a still looser criterion for semantic scoring. When the stimulus contained one proper name, errors in the reproduction were not counted as semantic errors. When the stimulus contained more than one proper name, the errors were counted only if the confusion was obvious. Thus, interchange of two proper names contained in the stimulus was counted as a semantic error, while replacement of one of the proper names with an entirely different one was not so counted.

The two dimensional scoring was devised so as to separate two independent aspects of syntactic perception: perceptual processing and the resulting percept structure. Our assumption here is that semantic errors are indicative of the complexity of the perceptual processes involved, and

Figure 2. Percentages of Syntactically and Semantically Correct Responses in Experiment I



syntactic errors are indicative of the remoteness of the percept structure from the reproduced sentence. See Section 4.2 for further discussion.

Results

Figure 2 summarizes the results of Experiment I. The blank bars indicate the percentages of syntactically correct responses in the total responses for the specified stimulus types, and the black bars indicate the percentages of semantically correct responses.

3.2 Experiment II

Material

The main material in this experiment was 12 sets of synonymous sentences differing in word order. The stimuli in this experiment were shorter than those in Experiment I and consisted of 4-7 lexical morphemes plus 3-6 grammatical morphemes. In addition to these, 6 pairs of stimuli from Experiment I were included. The total number of stimuli was 38.

The test sentences were broken up into two sets and were recorded on separate tapes. Each sentence was followed by a distractor, which in this experiment was a sequence of 9 randomly chosen digits.

Subjects

The Ss were 63 female students at a business school in Tokyo.

Procedure

The Ss were divided into two groups, and each group went through a session on the same occasion. The order of tested stimuli sets was different for the two groups.

The Ss were instructed to write down as many digits as they could recall and then reproduce the stimulus sentence in written form. The stimuli were recorded at an interval of 25 seconds, and the Ss were also instructed to stop writing to attend to the next stimulus when they heard the call sign immediately before the new stimulus.

Results

Table I shows the reaction types and the frequencies of each response type observed in the data from Experiment II. The leftmost numbers refer to the stimulus sentences listed in Section 2 above.

4. Discussion

4.1. Confirmation of the PS model

The results of Experiment I, summarized in Figure 2 above, provide some evidence for the PS model. Notice that of the seven constructions for which semantic scoring is made, there are only two in which the transformationally more complicated stimuli provoked significantly more semantic errors than transformationally simpler ones: Quantifier Extraction and Scrambling. In the case of Quantifier Extraction, the difference in the percentage of semantic errors between type Qx stimuli and type xQ stimuli is significant, $\chi^2 = 13.38$, $p < .001$. In the case of Scrambling, $\chi^2 = 4.75$,

Table I. Word Order of Subject, Object, and Adverb

Stimulus type	D(S)	Response type	D(R)	Frequency	
SV [10]	0	SV	0	63	100%
SOV [11 a]	0	SOV	0	63	100%
OSV [11b]	1	OSV	1	53	84. 1%
		SOV	0	10	15. 9%
SOiOdV [12a]	0	SOiOdV	0	63	100%
OiSOdV [12b]	2	OiSOdV	1	59	93. 7%
		SOiOdV	0	4	6. 3%
SOAV [13a]	1	SOAV	1	26	41. 3%
		OSAV	2	1	1. 6%
		SAOV	0	36	57. 1%
SOAV [14a]	1	SOAV	1	47	74. 6%
		SAOV	0	16	25. 4%
ASOV [15a]	1	ASOV	1	41	65. 1%
		SAOV	0	22	34. 9%
AOSV [13b]	2	AOSV	2	26	41. 3%
		ASOV	1	29	46. 0%
		SOAV	1	1	1. 6%
		SAOV	0	7	11. 1%
OSAV [14b]	2	OSAV	2	41	65. 1%
		ASOV	1	2	3. 2%
		SOAV	1	16	25. 4%
		SAOV	0	4	6. 3%
OASV [15b]	3	OASV	3	20	31. 7%
		OSAV	2	16	25. 4%
		AOSV	2	1	1. 6%
		SOAV	1	6	9. 5%
		SAOV	0	20	31. 7%

The figures in square brackets immediately following the stimulus types refer to the sentences quoted in Section 2. D(S) and D(R) indicate the "distance" of the stimulus and the response sentences from the canonical form, respectively. For the definition of "distance", see Section 4. 3 below.

$p < .05$. In all the other cases there are no significant differences between the derivationally related stimuli so far as semantic correctness is concerned. This fact is inconsistent with the AA and the AS models, since they both predict that the transformationally more complicated members go through more steps in syntactic perception and are hence liable to produce more semantic errors than transformationally simpler ones. We can safely conclude, therefore, that syntactic perception itself is quite independent from grammatical derivation.

4.2. Dominant vs. Recessive Syntactic Types

An examination of Figure 2 would immediately reveal the existence of significant differences between related stimuli in the percentage of syntactically correct responses. In seven out of eight constructions, one of the members is "dominant" over the other(s). Table 2 indicates the dominant and recessive members for these constructions.

Table 2

CONSTRUCTION	DOMINANT MEMBER		RECESSIVE MEMBER		CHI-SQUARE
Quantifier Extraction	Qx	95.6	xQ	77.2	11.85 ($p < .001$)
Passivization	Active	95.7	Passive	77.2	11.85 ($p < .001$)
Scrambling	S-ADV	98.9	ADV-S	75.0	*
Conjunct Movement	X-ga-Y-to	93.4	X-to-Y-to	66.3	41.51 ($p < .001$)
Subject Raising	SR	75.9	nonSR	66.2	5.01 ($p < .05$)
Subjectivization	X-no-Y-ga	92.9	X-ga-Y-ga	31.9	*
Stative Object marking (a)	Do	80.0	Dga	42.2	13.51 ($p < .001$)
(b)	X-ga-Y-o	97.8	X-ga-Y-ga	23.9	*
	X-ni-Y-ga	91.3			42.77 ($p < .001$)

The only instance in which no significant difference is found between the pair of stimuli is that of Ga-No Conversion, $\chi^2 = 1.82$.

The problem we must now face is how to interpret these facts. One possible account attributes the difference between the dominant and the recessive syntactic structures simply to the existence of preferred constructions at the production level. If this is correct, then the facts noted immediately above cannot be interpreted as indicative of anything about the nature of syntactic perception. While we admit that factors of speech production might be involved here, however, we do not agree that the observed

* For these constructions the value of chi-square is not specified because there was a response type with fewer than five occurrences.

facts are totally unrelated to syntactic perception. The responses obtained in our experiments were surely outputs of production processes but in producing them the Ss must appeal to the internal representation shaped through syntactic perception. That is, the task our Ss were required to perform was essentially a reproduction task. In reproducing the original sentence, the S could refer potentially to two sources of information: the word chain obtained through the word detection process, and the percept structure resulting from perceptual analysis of the word chain. Because of the time lag between the acceptance of the stimulus and the reproduction, the memory of the word chain itself is likely to decay rapidly unless further perceptual processes consolidate it. This entails that those Ss who have lost the memory of the word chain at the stage at which reproduction is called for can only avail themselves of the percept structure. Suppose now that the percept structure for a pair of transformationally related stimuli has a greater affinity to one of the pair. In such cases, the stimulus closer to the percept structure would provoke more syntactically correct responses than the one remoter from it, since in order to correctly reproduce the latter kind of stimulus the Ss must refer to the information contained in the word chain. The former kind of stimulus, on the contrary, is easily reproduced becaused the Ss may just refer to the percept structure.

We would thus like to propose the following hypothesis: The percept structure of a sentence takes a certain designated form, which we shall call the "canonical form". Given this hypothesis, we can interpret the above-mentioned facts about the contrast between dominant and recessive syntactic structures in the following way: the dominant syntactic structures are closer to the canonical form than the recessive ones. *

4. 3. Canonical Form of a Sentence

In the preceding section we proposed the hypothesis that the percept structure takes a certain designated form called the "canonical form". This means that the hearer has a kind of "template" in his mind, and the role of perceptual strategies is basically to fill in the "slots" on this template with the elements in the input utterance. The question that we must address ourselves to at this point is that of the empirical content of this template. The results of our experiments, expecially those of Experiment II, strongly suggest that the canonical form of a sentence not only consists of "slots" to be filled in but those slots are arranged in some sequence. A closer examination reveals that this is in fact the case, and the following can be proposed as a first approximation of the canonical form:

(1) Subject-Adverbial-Indirect Object-Direct Object-Verb

---

* It might even be possible to strengthen this hypothesis and propose a model of performance in which sentence perception and sentence production are tightly integrated. For instance, we may conceive of a "motor theory of syntactic perception" and, correspondingly, a "perception reference theory of sentence production". These conjectures are, however, highly speculative at the present moment, and we will not pursue this matter here.

In the following discussion, we will demonstrate that this is the case.

Let us first define the notion "distance" in the following way. The "distance" of a sentence K from another sentence P is the number of permutation operations of adjacent elements that are necessary and sufficient in converting K to P (or vice versa). Thus, the distance of a sentence of the type SOAV from form (1), for example, is 1, since it takes only one permutation (i. e. that of A and O) to derive the type SOAV from form (1). Likewise, sentence type OSAV has a distance 2 from (1), sentence type OASV has a distance 3 from (1), and so on. *

Now, on the basis of Table 1 we can derive Table 3, which exhibits the relations between the distance of the stimuli from (1) and that of the corresponding responses. The figures in each row indicate the percentages of the responses whose distance from (1) is as indicated above the column.

Table 3

		Distance from (1) of responses			
		0	1	2	3
Distance from (1) of stimuli	3	31.7	9.5	27.0	31.7
	2	8.7	38.1	53.2	0
	1	39.2	60.3	0.5	0
	0	100	0	0	0

This table shows remarkable tendencies. First, it is obvious that the closer the stimulus is to form (1), the higher is the percentage of correct responses (cf. the figures that appear diagonally from the upper righthand corner to the lower lefthand corner). In other words the stimuli closer to form (1) are easier to reproduce correctly. Moreover, it can also be seen that the distance of a response from (1) is always smaller (with only one exception) than that of the corresponding stimulus.

We can therefore conclude that form (1) is indeed a kind of standard structure, which we identify with the canonical form proposed in the preceding section. Granted that form (1) represents the canonical form, the correlations between the distance of the stimuli and that of the responses as observed in Table 3 can be explained in the following fashion. The hearer analyzes the input utterance in terms of the canonical form, and the input sentence is perceived in this form unless he is required to pay attention to the exact form of the input. In a repetition task such as the one involved in our experiments, the Ss memorize the original sentence as a combination of the canonized sentence form and the operations needed to convert this canonical form into the original input. The latter are liable to diminish as time goes on, and hence the responses tend to appear in forms closer to the canonical

* We are grateful to Hideo Hayashibe for suggesting the notion of "distance" as well as for other insightful comments.

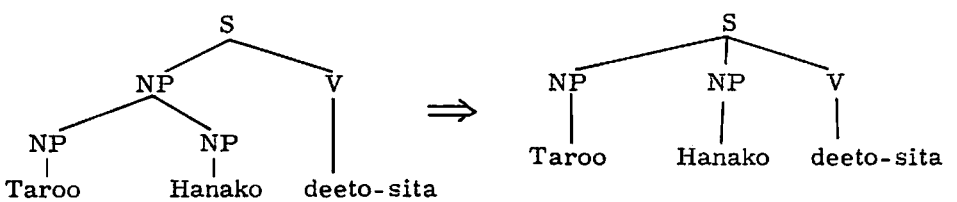
form. Furthermore, when the number of the operations needed to convert the canonical form into the input form goes beyond 2, the Ss tend to forget the entire operations altogether. This seems to be what is going on behind the abrupt increase of the responses in the canonical form (i. e. distance zero) for the stimuli with distance 3.

The above remarks are obviously just the beginning of an understanding of the notion of "canonical form", but it seems to us that the basic lines of our theory are sufficiently clear.

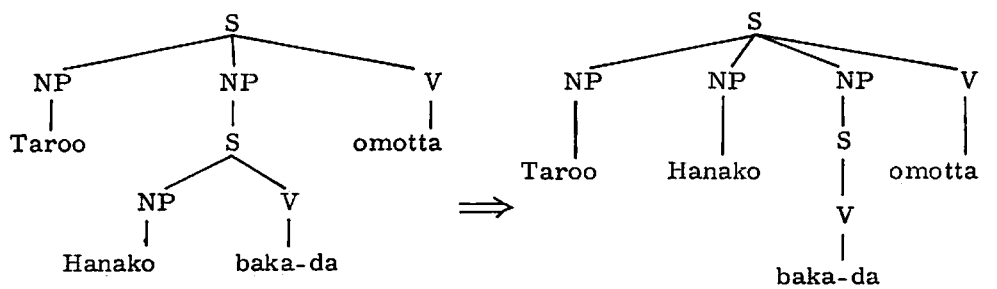
4. 4. Perceptual Optimality of Multiple-Branching Constructions

Another look at the results of Experiment I reveals the fact that constructions with fewer hierarchical layers are easier to perceive than those with more elaborate hierarchical structures. This is indicated in two cases: Conjunct Movement and Subject Raising. The effects of these rules are schematically shown below:

Conjunct Movement



Subject Raising

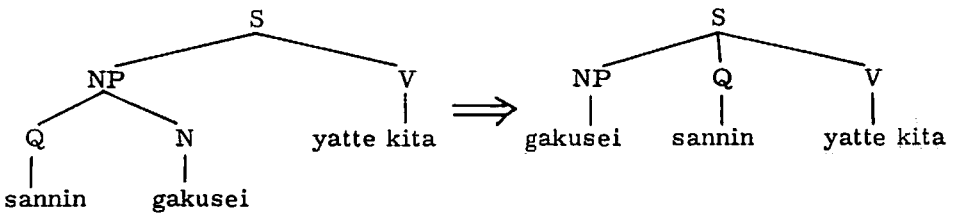


Notice that in each case the application of the transformation results in a decrease in the number of branching nodes. In other words, the derived structures are "flatter" than the underlying structures. That such structures are perceptually easier than those with more hierarchical strata was first explicitly stated by Miller and Chomsky (1963) in the form of the maxim that multiple-branching constructions are optimal in acceptability.

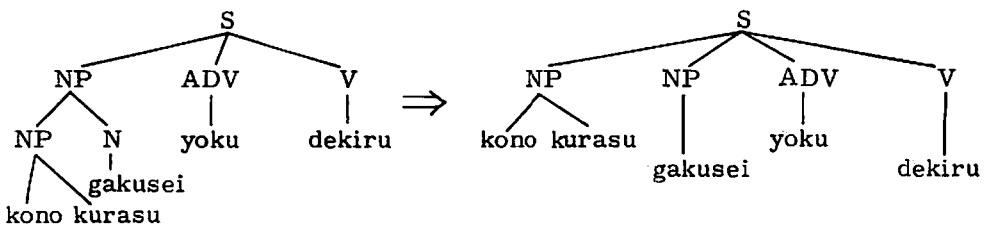
There are, however, cases in which the opposite is the case. These are the cases of Quantifier Extraction and Subjectivization. In these constructions the transformationally derived versions are much worse than those without transformational application in reproduction accuracy.

To account for the contrast between Conjunct Movement and Subject Raising on one hand and Quantifier Extraction and Subjectivization on the other, we must pay attention to the functional difference between these two sets of rules. Notice that the latter rules both destroy the structure of a NP.

Quantifier Extraction



Subjectivization



Thus, although these transformations flatten the phrase structure and must therefore produce perceptually optimal structures, they at the same time break a NP into two pieces, each functioning as an immediate constituent of the derived sentence.

This consideration strongly suggests that NPs of the form "modifier + head" are perceptual islands, and the destruction of NPs of this form requires the hearer to go through an extra step in shaping the percept structure.

4. 5. Degrees of Importance of Syntactic Cues in Perception

Let us finally consider the problem of the relative degree of relevance to perception of syntactic cues such as word order, phrase structure hierarchy, and particle choice. An examination of the results of Experiment I, as they are reorganized in Table 2, suggests that word order is the primary factor in syntactic perception, phrase structure the secondary, and particle choice the least important in syntactic perception. To show this, we must reanalyze Table 2 as follows:

Table 4

CONSTRUCTION	% of correct responses		Word Order	Hierarch- ical Depth	Particle Choice
	Dominant	Recessive			
Group I					
Quantifier Extraction	95.6	77.2	Different	Different	Different
Passivization	95.7	77.2	Different	Same	Different
Scrambling	98.9	75.0	Diffeeent	Same	Same
Group II					
Conjunct Movement	93.4	66.3	Same	Different	Different
Subject Raising	75.9	66.2	Same	Different	Different
Subjectivization	92.9	45.7	Same	Different	Different
Group III					
Stative Object Marking					
(i) Desiderative	80.0	42.2	Same	Same	Different
(ii) Potential (a)	97.8	23.9	Same	Same	Different
(b)	91.3	23.9	Same	Same	Different
Ga-No Conversion	71.4	64.8	Same	Same	Different

The constructions listed above can be grouped into three categories. The first group consists of those stimulus pairs in which the paired members differ in word order. The second group consists of those pairs whose members do not differ in word order but in hierarchical depth (as well as in particle choice). The third group are those whose members differ only in the choice of the particle. The average percentages of correct responses for these groups are indicated in Table 5.

Table 5

	Dominant	Recessive	Mean
Group I	96.7	76.4	86.6
Group II	84.7	55.5	70.0
Group III	79.3	54.2	67.7

Note that there is a significant difference in the percentage of correct responses between Group I on one hand and Groups II and III on the other, $\chi^2 = 15.87$, $p < .001$. The difference between Group II and Group III is also significant, $\chi^2 = 4.92$, $p < .05$. This entails that word order difference is the primary cue for discriminating synonymous sentences with syntactic differences. Hierarchical depth is less effective, and the choice of particles has the least effect on syntactic perception. This result is of some interest because some recent attempts to account for sentence acceptability in perceptual terms assume that particles are the primary cues for syntactic perception (see, in particular, Shibatani, 1975). The result of our study

summarized in Table 5 thus provides counter-evidence to such an assumption.

It is also worthwhile to note that the above-mentioned facts are corroborated by the findings of recent studies in child language development. For instance, Hayashibe (1975) has found that young children (about 3-6 years old) come to use word order strategy considerably before they come to appeal to the strategy in terms of particles.

Acknowledgments

We are deeply indebted to all the people who have helped us at various stages in our study, but we feel that special thanks should go to Hideo Hayashibe, Akio Kamio, Keiko Matsuki, Hajime Narita, and Hiroshi Yamada. We do not mean to imply, however, that any of them is responsible for errors that may remain in the present paper.

References

Bever, Thomas G. (1970) "Cognitive basis for linguistic structures", in John R. Hayes (ed.), Cognition and the Development of Language (New York: Wiley), pp. 279-362.

Chomsky, Noam (1957) Syntactic Structures. The Hague: Mouton.

Fodor, J. A., T. G. Bever, and M. F. Garrett (1974) The Psychology of Language. New York: McGraw-Hill.

Harada, S. I. (1971) "Ga-No Conversion and idiolectal variations in Japanese", Gengo Kenkyu 60:25-38.

Hayashibe, Hideo (1975) "The del elopment of sentence comprehension in Japanese children", Descriptive and Applied Linguistics (Tokyo: International Christian University), Vol. 8.

Howard, Irwin, and A. Niyekawa-Howard (in press) "Passivization", in M. Shibatani (ed.), Japanese Generative Grammar (New York: Seminar Press).

Kamio, Akio (1973) "Observations on Japanese quantifiers", Descriptive and Applied Linguistics, Vol. 6, pp. 69-92.

Kuno, Susumu (1972) "Evidence for Subject Raising in Japanese" Papers in Japanese Linguistics 1:1 24-51.

--------- (1973) The Structure of the Japanese Language. Cambridge, Mass.: MIT Press.

Miller, George A. (1962) "Some psychological studies of grammar", American Psychologist 17:748-762.

Miller, George A., and Noam Chomsky (1963) "Finitary models of language users", in R. D. Luce, R. R. Bush, and E. Galanter (eds.), Handbook of Mathematical Psychology, Vol. II (New York: Wiley), pp. 419-491.

Okutsu, Keiichiro (1969), "Suuryoo-teki hyoogen no bunpoo" [The grammar of quantifying expressions], Nihongo Kyooiku (Tokyo: International Christian University), 14:42-60.

Shibatani, Masayoshi (1975) "Perceptual strategies and the phenomena of particle conversion in Japanese", to appear in the 1975 Chicago Linguistic Society parasession volume.